ROADVIEW

Robust Automated Driving in Extreme Weather

Project No. 101069576

Deliverable 8.7

Final KPI specifications

WP8 – Integration and demonstration

Authors	Mucahit Furkan Özdemir (FORD), Osman Tahir Ekşi (FORD)
Lead participant	FORD
Delivery date	29 August 2025 (v1.0) 10 December 2025 (v2.0)
Dissemination level	Public
Type	Report

Version 2.0

Co-funded by the European Union

Innovate UK

Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. Project grant No. 101069576.

UK participants in this project are co-funded by Innovate UK under contract No.10045139.
Swiss participants in this project are co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract No. 22.00123.

Revision history

Author(s)	Description	Date
Mucahit Furkan Özdemir (FORD)	Draft deliverable	22/08/2025
Mikko Tarkiainen (VTT)	Internal review of the deliverable	25/08/2025
Isabelle Morvan (CRF)	Internal review of the deliverable	25/08/2025
Mucahit Furkan Özdemir (FORD)	Final version (v1.0)	28/08/2025
Johannes Ripperger (accelCH)	Final checks and formatting	29/08/2025
Osman Tahir Ekşi (FORD), Mikko Tarkiainen (VTT) Isabelle Morvan (CRF)	Update after EC review (v2.0)	10/12/2025

Contents

List of Figures

List of Tables

Partner short names

HH	Hogskolan Halmstad
THI	Technische Hochschule Ingolstadt
FGI	Maanmittauslaitos – Finnish Geospatial Research Institute
VTT	VTT Technical Research Centre of Finland Ltd
FORD	Ford Otomotiv Sanayi A. S.
CRF	Canon Research Centre France
AVL	AVL Software and Functions GmbH
KO	Konrad GmbH
accelCH	accelopment Schweiz AG

Abbreviations

A	Availability
AET	Average Execution Time
AFPT	Average Fusion Processing Time
AIT	Average Inference Time
ASL	Average System Latency
ASR	Adaptation Success Rate
ATE	Absolute Trajectory Error
ATE_{pos}	Absolute Trajectory Error Positioning
ATE_{ratio}	Absolute Trajectory Error Ratio
CAV	Connected and Automated Vehicle
CCAM	Connected, Cooperative and Automated Mobility
CE_{model}	Coverage Error of Model
CWHIr	Run-Based Criticality-Weighted Human Intervention
CWHIt	Time-Based Criticality-Weighted Human Intervention
D	Deliverable
DA	Decision Accuracy
DMV	Department of Motor Vehicles
EA-NDT	Environment-Aware Normal Distributions Transform
EU	European Union
EVDL	External-to-Vehicle Delivery Latency
$F1_{macro}$	Macro-averaged F1-score
FDR	False Detection Range
FN	False Negatives

FP	False Positives
FPR	False Positive Rate
FPS	Frame per Second
FRr	Run-Based Failure Rate
FRt	Time-Based Failure Rate
GNSS	Global Navigation Satellite System
HD	High Definition
HIRr	Run-Based Human Intervention Rate
HIRt	Time-Based Human Intervention Rate
ID	Identification
IoU	Intersection over Union
IRd	Distance-Based Incident Rate
IRt	Time-Based Incident Rate
ISO	International Organization for Standardization
KPI	Key Performance Indicator
LiDAR	Light Detection and Ranging
mAP	mean Average Precision
MAPEdis	Mean Absolute Percentage Error of Distance
MAPEvis	Mean Absolute Percentage Error of Visibility Estimator
MCAR	Manoeuvre Coordination Acceptance Rate
mIoU	mean Intersection over Union
MRD	Maximum Relative Deviation
MRM	Minimal Risk Manoeuvre
NMRd	Distance-Based Near-Miss Rate
NMRt	Time-Based Near-Miss Rate
OCA	Overall Classification Accuracy
ODD	Operational Design Domain
OTDR	On-Time Demonstration Rate
PWT	Percentage Within Tolerance
R	Reliability
RMSE	Root Mean Square Error
$RMSE_{ratio}$	Root Mean Square Error Ratio
SMART	Specific, Measurable, Attainable, Relevant, Time-Bound
SME	Small and Medium-sized Enterprise
SSR	Scenario Success Rate
T	Task
TLERd	Distance-Based Traction Loss Event Rate

TLERt	Time-Based Traction Loss Event Rate
TN	True Negatives
TP	True Positives
TPR	True Positive Rate
WNA	Weather-Normalized Availability
WP	Work Package
V2X	Vehicles-to-Everything
VD	Velocity Difference
VR	Velocity Rate

Executive summary

The ROADVIEW project, funded by the European Union (EU), aims to enhance the safety and reliability of Connected and Automated Vehicles (CAVs) under adverse weather and varying road conditions. As a critical component of this initiative, Work Package (WP) 8 (Integration and Demonstration) has been integrating and demonstrating the ROADVIEW system in real-world and controlled scenarios. FORD, as the leader, together with AVL, VTT, and THI, has been responsible for specifying and assessing Key Performance Indicators (KPIs) to evaluate the system's performance in Task (T) 8.7.

This deliverable (D8.7), Final KPI Specifications, presents the complete and refined set of KPIs that have been defined, validated, and applied to measure the effectiveness and resilience of the ROADVIEW system. Building on the initial definitions provided in D8.6, Initial KPI Specifications, the KPIs have been reviewed, expanded, and where necessary, redefined based on partners' feedback, preliminary test results, and lessons learned during the integration and demonstration activities.

The final KPIs address key aspects such as system uptime/downtime, operational limits under varying intensities of rain, snow, and fog, execution times of ROADVIEW components, frequency and nature of human interventions, and the number of subsystem failures. These metrics have been calibrated to ensure that they are measurable, relevant, and directly aligned with the project's performance objectives.

By consolidating all feedback and incorporating insights from WP7 performance metrics, this deliverable provides a robust assessment framework that can be consistently applied across all ROADVIEW use cases and demonstrations. The final specifications will serve both as a benchmark for the current project outcomes and as a reference point for future research and development activities in the field of weather-resilient automated driving.

Objectives

The primary objective of this deliverable is to present the final and validated set of KPIs for evaluating the performance of the ROADVIEW system and its subsystems. They are essential for:

- **Measuring System Resilience**: Confirming and quantifying the ability of the ROADVIEW system to function effectively under various adverse weather conditions, including rain, snow, and fog.
- **Evaluating Operational Efficiency**: Monitoring the execution times of individual system components and the overall system uptime/downtime.
- **Identifying Failure Points**: Tracking and categorising the number of failed cases where the system or its subsystems stop working, providing context on root causes and frequency.
- **Assessing Human Interaction**: Recording and analysing the number and types of human interventions required during test runs, highlighting areas for optimisation and automation.
- **Facilitating Continuous Improvement**: Establishing a robust and measurable performance that enables systematic comparison across all project demonstrations and supports long-term development of weather-resilient automated driving systems.

This deliverable consolidates the KPI framework into its final form, ensuring that it is measurable, relevant, and applicable across all ROADVIEW use cases. By systematically applying and analysing these KPIs, the project team can accurately evaluate system performance, validate resilience against challenging weather scenarios, and provide actionable insights for future enhancements in CAV technologies.

Methodology and implementation

The final KPI specifications were developed by refining the initial set defined in D8.6 through iterative partner feedback, data analysis, and validation in WP8 demonstration activities. KPI definitions were aligned with WP7 performance metrics and tailored to the specific operational scenarios of each use case. The final KPI framework was implemented directly into demonstration planning, with systematic data logging, real-time monitoring, and post-test analysis ensuring reliable and traceable performance assessment.

Outcomes

The final KPI framework provides a well-defined and comprehensive basis for assessing the ROADVIEW system's performance across all demonstrations. The defined KPIs will enable the project to assess in D8.8 whether the system can maintain high operational uptime and functional resilience under varying intensities of rain, snow, and fog. They will also support evaluating whether execution times of core components remain within acceptable limits and whether human interventions remain minimal, thereby providing a framework for assessing system reliability. The refined KPIs, designed specifically for the evaluated sub-systems and aligned with real-world test scenarios, provide a solid reference framework that could be used in similar assessments and future developments related to weather-resilient functions of automated driving systems.

Next steps

With the final KPI specifications now defined and validated in D8.7, the next step within WP8 is the preparation of D8.8, the Report on Performance Assessment According to KPIs. This will involve applying the finalised KPI framework to analyse performance data collected from the integrated ROADVIEW system during demonstrations in varying adverse weather conditions. The performance assessment will leverage both the refined KPIs and the performance metrics developed in WP7, ensuring a comprehensive evaluation methodology.

Data collection will continue throughout the final demonstration activities, with a focus on ensuring completeness, accuracy, and consistency across all use cases. The analysis results will be reviewed collaboratively by all partners, and the findings will be consolidated into the D8.8 deliverable. This will mark the transition from KPI definition to full-scale performance evaluation, enabling the project to validate its achievements in enhancing CAV safety and reliability under challenging environmental conditions.

1 Methodological Basis for Final KPI Definition

Key Performance Indicators (KPIs) are measurable values that evaluate the extent to which an objective is achieved, as defined in International Organization for Standardization (ISO) 22400-2:2014 [1]. In the context of the ROADVIEW project, KPIs play a central role in evaluating system performance under adverse weather conditions, ensuring traceability, and guiding continuous improvement.

This deliverable (D8.7) presents the finalised set of KPIs, refined based on the initial definitions in D8.6 [2] and validated through real-world test results and partner feedback. All KPIs included in this report have been prepared in accordance with the KPI definition standards outlined in D8.6, including the SMART criteria (Specific, Measurable, Attainable, Relevant, Time-bound) and the structured development process illustrated in Figure 1.

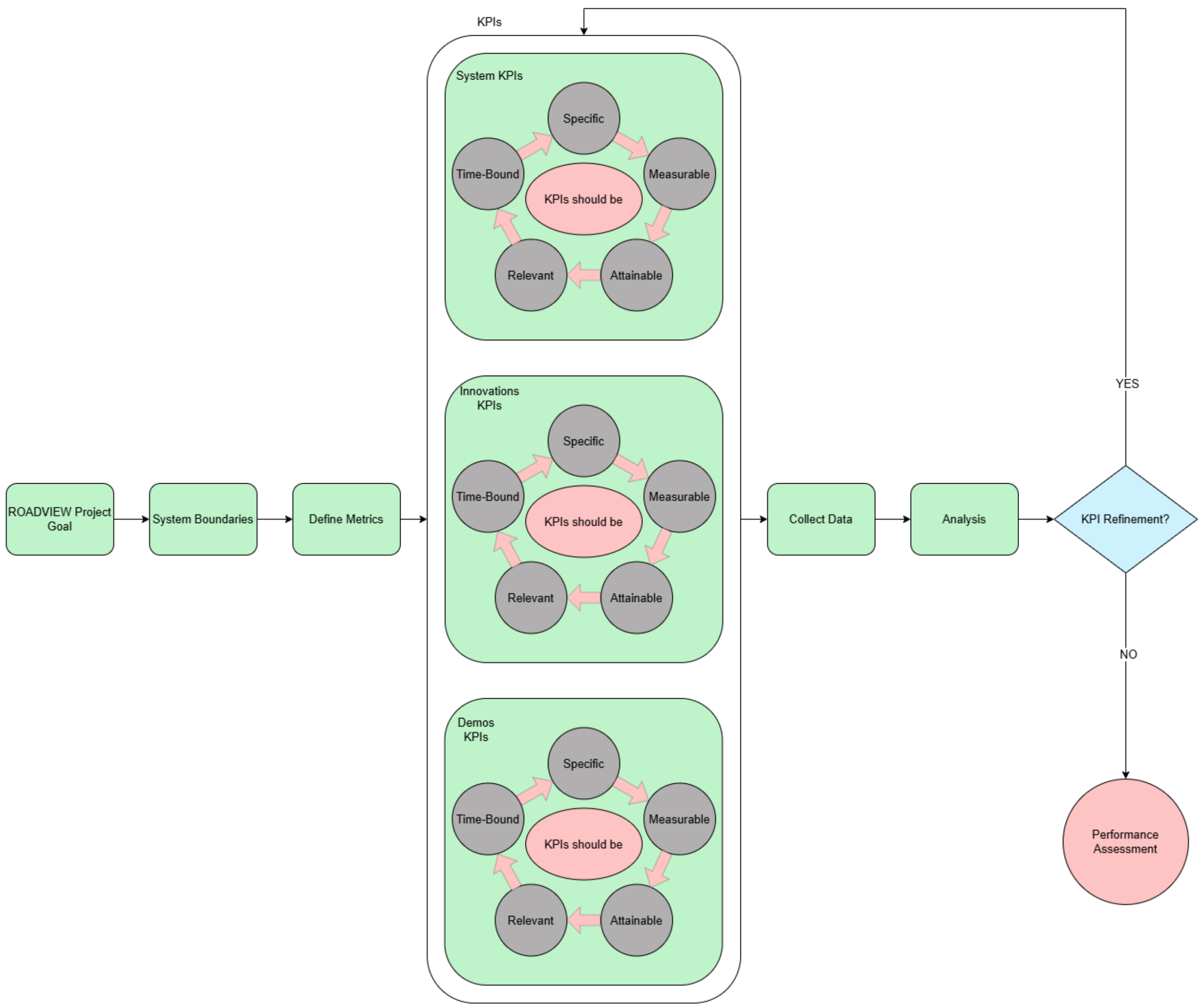

Figure 1. The full cycle of defining KPIs [2].

2 ROADVIEW KPIs

To design and develop the KPIs, the project goals and the system boundaries should be defined clearly, then high level to low-level KPIs should be specified.

2.1 Project Goal

The ROADVIEW project aims to achieve several key goals to enhance the safety, reliability, and acceptance of CAVs under various weather and traffic conditions. These goals include:

- **Enhancing Safety and Reliability in Adverse Weather Conditions**: Develop and integrate a sophisticated in-vehicle system-of-systems capable of advanced environment and traffic recognition, prediction, and decision-making to ensure CAVs can operate safely and reliably in harsh weather conditions such as rain, snow and fog.
- **Improving Perception and Decision-Making Systems**: Create an embedded in-vehicle perception and decision-making system based on enhanced sensing, localization, and improved object/person classification, including vulnerable road users, to accurately determine the appropriate course of action in real-world environments.
- **Developing Cost-Effective Multisensory Solutions**: Innovate a cost-effective multisensory setup with advanced sensor noise modeling and filtering to enhance the robustness and accuracy of CAV systems in various scenarios and weather conditions.
- **Advancing Collaborative Perception and Simulation-Assisted Testing**: Implement collaborative perception techniques and simulation-assisted testing methods to rigorously test and validate the performance of the ROADVIEW system under different scenarios, ensuring it reaches Technology Readiness Level 7 by the end of the project.
- **Contributing to the Connected, Cooperative and Automated Mobility (CCAM) Partnership**: Support the co-programmed European Partnership CCAM by developing more powerful, fail-safe, resilient, and weather-aware technologies, thereby advancing the overall development and adoption of CAVs.
- **Leveraging Consortium Expertise and Testing Infrastructure**: Utilize the unique portfolio of testing sites and infrastructure provided by the consortium members, including hardware-testing facilities, rain and wind tunnels, and test tracks north of the Arctic Circle, to thoroughly evaluate and demonstrate the ROADVIEW system's capabilities.

2.2 System Boundaries

The five general system boundaries of the projects are:

- The ROADVIEW system must be designed to operate effectively in adverse weather conditions such as rain, snow and fog.
- Development and integration of advanced environment and traffic recognition, prediction, and decision-making capabilities.
- Ensuring the ROADVIEW system maintains high uptime and minimal downtime during operation.
- Minimizing the need for human interventions during system operation.
- Rigorous testing and validation of the system using facilities like rain and wind tunnels, and test tracks under varied environmental conditions.

2.3 System KPIs

To define the KPIs we use the cascade method, introduced by the WEDISTRICT Project [3], starting from high-level KPIs to low-level ones. As the high-level KPIs, we need to define the ROADVIEW system's KPIs first.

1. **System Resilience (ID: S-KPI1):**

- **KPI:** Percentage of operational time in adverse weather conditions (rain, snow, fog).
- **Measurement:** The proportion of time the ROADVIEW system remains fully functional during specified weather conditions.
- **Proposed Metric:**
  Weather-Normalized Availability (WNA):

$$\mathrm{WNA} = \left(\frac{\text{Operational time (adverse weather)}}{\text{Total time (adverse weather)}}\right) \times 100\% \quad (1)$$

  Unit: %

2. **Failure Rate (ID: S-KPI2):**
   - **KPI:** Number of failed cases where the system or subsystems stop working.
   - **Measurement:** Total count of system failures per test run or operational hour.
   - **Proposed Metrics:**
     - Run-based:
       Run-Based Failure Rate (FRr):

$$\mathrm{FRr} = \frac{\text{Total number of failure cases}}{\text{Total number of runs}} \quad (2)$$

       Unit: Failure cases per run

     - Time-based:
       Time-Based Failure Rate (FRt):

$$\mathrm{FRt} = \frac{\text{Total number of failure cases}}{\text{Total operational time}} \quad (3)$$

       Unit: Cases per hour

3. **Uptime/Downtime (ID: S-KPI3):**
   - **KPI:** System uptime versus downtime in harsh weather.
   - **Measurement:** Hours of continuous operation without failure compared to hours of downtime.
   - **Proposed Metric:**
     Availability (A):

$$\mathrm{A} = \left(\frac{\text{Uptime}}{\text{Uptime + Downtime}}\right) x\, 100\% \quad (4)$$

     Unit: %

4. **Human Intervention (ID: S-KPI4):**
   - **KPI:** Frequency and types of human interventions.
   - **Measurement:** Number of times human intervention is required during system operation, categorised by intervention type.
   - **Proposed Metrics:**
     - Human intervention rate, whatever is the type of intervention:
       - Run-based:
         Run-Based Human Intervention Rate (HIRr):

$$\mathrm{HIRr} = \frac{\text{Total number of interventions}}{\text{Total number of runs}} \quad (5)$$

         Unit: Interventions per run

       - Time-based:
         Time-Based Human Intervention Rate (HIRt):

$$\mathrm{HIRt} = \frac{\text{Number of total interventions}}{\text{Total operational time}} \tag{6}$$

Unit: Interventions per hour

- Criticality weight-based human intervention rate:
  Intervention Types:
  C3 Safety-Critical, weight is 3.0
  C2 Performance-Critical, weight is 2.0
  C1 Comfort-Critical, weight is 1.0
  (The classification of interventions into safety-critical, performance-critical, and other/comfort-critical categories is based on Department of Motor Vehicles (DMV) disengagement reports from the United States [4]. The weight factors (3.0, 2.0, 1.0) are defined within the KPI framework of this study. However, in the ROADVIEW context, comfort-critical (C1) interventions are not included in the analysis, since comfort-related events are not within the objectives of this project.)

  Run-based:
  Run-Based Criticality-Weighted Human Intervention (CWHIr):

$$\mathrm{CWHIr} = \frac{\sum_{i=1}^{N} w_{c(i)}}{\text{Total number of runs}} \tag{7}$$

  wc(i): i-th intervention's criticality weight
  Unit: Interventions per run

  Time-based:
  Time-Based Criticality-Weighted Human Intervention (CWHIt):

$$\mathrm{CWHIt} = \frac{\sum_{i=1}^{N} w_{c(i)}}{\text{Total operational time}} \tag{8}$$

  $w_{c(i)}$: i-th intervention's criticality weight
  Unit: Interventions per hour

5. **Execution Time (ID: S-KPI5):**
   - **KPI:** Average execution time of ROADVIEW components
   - **Measurement:** Time taken for critical system components to perform their tasks, measured in milliseconds or seconds.
   - **Proposed Metric:**
     Average Execution Time (AET):

$$\mathrm{AET} = \frac{\sum_{i=1}^{N} T_i}{\mathrm{N}} \tag{9}$$

     T_i: i-th component's execution time (ms)
     N: Total number of system component runs
     Unit: Millisecond

2.4 Innovations KPIs

Within the ROADVIEW project, twelve innovations have been identified across different focus areas such as perception, decision-making, and V2X communication. The following section presents the KPIs defined for each innovation, ensuring that their performance can be measured, compared, and validated consistently, based on definitions provided in previous ROADVIEW deliverables and supported by relevant literature.

2.4.1 Infrastructure-Based Perception System including Weather Conditions (V2X)

1. **Accuracy of Weather Condition Detection (ID: I-KPI1):**
   - **Baseline:** Current accuracy level of weather detection.
   - **Target:** Improvement percentage in accuracy with the new system.
   - **Proposed Metric:**
     Overall Classification Accuracy (OCA) [5]:

$$\text{OCA} = \left(\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}\right) \times 100\%, \qquad OCA > \text{baseline} \tag{10}$$

     TP: True positives
     TN: True negatives
     FP: False positives
     FN: False negatives
     Unit: %

   - **Proposed Alternative Metric:**
     F1-Score [5] [6]:

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \qquad F1 > \text{baseline} \tag{11}$$

$$\text{Pr e cision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

     TP: True Positive
     FP: False Positives
     FN: False Negative

2. **System Latency (ID: I-KPI2):**
   - **Baseline:** Current system response time.
   - **Target:** Less than 750 ms between the roadside sensor data acquisition and the reception of perceived objects by the onboard system.
   - **Proposed Metric:**
     External-to-Vehicle Delivery Latency (EVDL):

$$\text{EVDL} = \text{T}_{\text{reception}} - \text{T}_{\text{data\_acquisition}}, \qquad EVDL < 750 \tag{14}$$

     $T_{reception}$: Timestamp when the data received by ego vehicle
     $T_{data\_acquisition}$: Timestamp when the data sent by the external unit (e.g. roadside sensor)
     Unit: Time (ms)

2.4.2 Weather-Aware Infrastructure-Based Manoeuvre Cooperation (V2X)

1. **Manoeuvre Coordination Success Rate (ID: I-KPI3):**

- **Baseline:** Currently, there is no implementation of Minimal Risk Manoeuvre (MRM) request triggered from the V2X system.
- **Target:** 90% of acceptance by the onboard system to follow an MRM request proposed by the infrastructure-based system in case of harsh weather.
- **Proposed Metric:**
  Manoeuvre Coordination Acceptance Rate (MCAR):

$$\text{MCAR} = \left(\frac{\text{Number of accepted MRM}}{\text{Total number of MRM}}\right) \times 100\%, \qquad MCAR > 90\% \tag{15}$$

  Unit: %

2.4.3 Environment-Aware High-Definition Mapping

1. **Real-Time Global Positioning Accuracy (ID: I-KPI4):**
   - **Baseline:** Current accuracy of vehicle online Global Navigation Satellite System (GNSS) positioning.
   - **Target:** Reduced Absolute Trajectory Error (ATE) with Environment-Aware Normal Distributions Transform (EA-NDT) High Definition (HD) mapping.
   - **Proposed Metric:** ATE_{ratio}, which is a percentage of positioning ATE of the EA-NDT HD map-based positioning from ATE of baseline positioning:

$$ATE_{ratio} = \frac{ATE_{pos}(HD\ map)}{ATE_{pos}(GNSS)}\ x\ 100\%, \qquad ATE_{ratio} < baseline \tag{16}$$

$$ATE_{pos} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\boldsymbol{t}_{err,i}\right\|_2^2} \tag{17}$$

x

$t_{err,i}$: The position difference of compared method with respect to the reference positioning at i-th measurement from N.
N: Total number of measurements.

2. **Size of the HD Map representation (ID: I-KPI5):**
   - **Baseline:** Size of NDT HD Map
   - **Target:** Reduced EA-NDT HD map size compared to the baseline NDT HD map with at least equal positioning accuracy compared to online GNSS positioning baseline.
   - **Proposed Metric:** $Size_{ratio}$, which is a percentage of EA-NDT HD map size from the baseline NDT HD map:

$$Size_{ratio} = \frac{Size(HD\ map)}{Size(NDT)}\ x\ 100\%, \qquad Size_{ratio} < baseline \tag{18}$$

where size of HD map and compared NDT map are measured from the memory they consume.

2.4.4 Optical Measurement of Road Grip and Surface Conditions

1. **Grip Prediction Accuracy (ID: I-KPI6):**
   - **Baseline:** Root Mean Square Error (RMSE) of a single grip measurement from an optical road-weather sensor extended over the whole measurement area in road scenes with varying grip values.

- **Target:** Lower grip prediction RMSE when the grip map is predicted pixelwise over the whole road area.
- **Proposed Metric:** $RMSE_{ratio}$, which is a percentage of grip prediction RMSE from the RMSE of baseline grip measurements:

$$RMSE_{ratio} = \frac{RMSE(grip\ prediction)}{RMSE(grip\ baseline)}\ x\ 100\%, \qquad RMSE_{ratio} < baseline \tag{19}$$

2. **Fraction of Ground Truth Grip Measurements within 95% Prediction Confidence Interval (ID: I-KPI7):**
   - **Baseline:** Fraction of ground truth grip measurements within a 95% confidence interval of a normal distribution estimated from 100 last grip measurements from an optical road-weather sensor (with the lower limit of standard deviation set to 0.05 in grip scale) in road scenes with varying grip values.
   - **Target:** A fraction of ground truth grip measurements within the predicted 95% confidence interval is closer to the correct fraction of 95% when grip confidence map is predicted pixelwise over the whole road area.
   - **Proposed Metric:** The absolute ratio between two estimated differences:
     1) The difference between the percentage of ground truth grip measurements within the 95% confidence interval predicted by the model and the correct 95% value, and
     2) The difference between the percentage of ground truth grip measurements within the 95% confidence interval estimated from the baseline normal distribution and the correct 95% value.

     Coverage Error of Model (CE_{model}):

$$CE_{\text{model}} = |p_{\text{model}} - 0.95| \times 100\%, \qquad CE_{\text{model}} < baseline \tag{20}$$

CE_{model}: The deviation of the faction of ground truth measurements falling within the 95% confidence interval predicted by the model from the theoretical 95% value
p_{model}: Fraction of ground truth grip measurements within the 95% confidence interval predicted by the model
Unit: %

2.4.5 Fast and Accurate Multi-Sensor Fusion

1. **Fusion Processing Time (ID: I-KPI8):**
   - **Baseline:** Current processing time for sensor fusion (including Camera, Light Detection and Ranging (LiDAR), and Radar).
   - **Target:** Reduction in processing time.
   - **Proposed Metric:**
     Average Fusion Processing Time (AFPT):

$$\text{AFPT} = \frac{\sum_{i=1}^{N}\left(\text{T}_{\text{fusion\_end,i}} - \text{T}_{\text{fusion\_start,i}}\right)}{\text{N}}, \qquad AFPT < baseline \tag{21}$$

$T_{fusion\_start,\,i}$: i-th timestamp when fusion process starts
$T_{fusion\_end,\,i}$: i-th timestamp when fusion process ends
N: Total number of measurements
Unit: Milliseconds

2. **Accuracy of Fused Sensor Data in 3D object detection and semantic segmentation tasks (ID: I-KPI9):**
   - **Baseline:** Current accuracy of multi-sensor data.
   - **Target:** Enhanced accuracy with joint multi-task learning (including both 3D object detection and semantic segmentation).
   - **Proposed Metrics:**
     3D Object Detection:

mean Average Precision (mAP) [6] [7]:

$$\text{mAP} = \frac{1}{C}\sum_{i=1}^{C} AP_i\,, \qquad mAP > \text{baseline}, \qquad \text{IoU} \geq 0.5{,}0.7 \tag{22}$$

C: Total number of classes
AP_c: Average Precision for class i
Intersection Over Union (IoU) Threshold: Intersection over Union threshold used for evaluation

Semantic Segmentation:
mean Intersection over Union (mIoU) [6] [8]:

$$\text{mIoU} = \frac{1}{\text{C}}\sum_{i=1}^{\text{C}} \left(\frac{\text{TP}_\text{i}}{\text{TP}_\text{i} + \text{FP}_\text{i} + \text{FN}_\text{i}}\right), \qquad mIoU > baseline \tag{23}$$

TP_i: True positive pixel number for class i
FP_i: False positive pixel number for class i
FN_i: False negative pixel number for class i
C: Total number of classes

2.4.6 Fast and Memory Efficient Outlier Detection

1. **Detection Speed (ID: I-KPI10):**
   - **Baseline:** Current speed of conventional statistical outlier detection in 3D point cloud data.
   - **Target:** Faster detection times.
   - **Proposed Metric:**
     Average Inference Time (AIT):

$$\text{AIT} = \frac{\sum_{i=1}^{\text{N}} IT_i}{\text{N}}, \qquad AIT < baseline \tag{24}$$

     N: Number of inferences
     IT_i: i-th inference time
     Unit: Milliseconds

     Supporting Metric:
     Frame per second (FPS) [6] [9]:

$$\text{FPS} = \frac{1000}{\text{AIT}}, \qquad FPS > baseline \tag{25}$$

     AIT: Average inference time
     Unit: frame/second

2. **Accuracy of Outlier Detector (ID: I-KPI11):**
   - **Baseline:** Current accuracy (mIoU) of conventional statistical outlier detection in 3D point cloud data.
   - **Target:** Enhanced accuracy in outlier detection.
   - **Proposed Metric:**
     As previously formulated in Equation 23, mIoU can be applied in this context.

2.4.7 Fast Free Space Detector Using LiDAR Clouds

1. **Free Space Detection Accuracy (ID: I-KPI12):**
   - **Baseline:** Current accuracy of free space detection (mIoU) in LiDAR point clouds.
   - **Target:** Improvement in detection accuracy.

- **Proposed Metric:**
  This innovation KPI follows the formulation, mIoU, given in Equation 23.

2. **Processing Speed (ID: I-KPI13):**
   - **Baseline:** Current speed of free space detection in LiDAR point clouds.
   - **Target:** Increased processing speed.
   - **Proposed Metric:**
     The AIT and FPS formulations defined in Equations 24 and 25 are also applicable to this KPI.

2.4.8 Semantic-Aware Pseudo Point Cloud-Based 3D Object Detector

1. **3D Object Detection Accuracy (including vehicles, pedestrians, and cyclists) (ID: I-KPI14):**
   - **Baseline:** Current accuracy (mAP) of 3D object detection in point clouds.
   - **Target:** Enhanced detection accuracy.
   - **Proposed Metric:**
     Equation 22, which define mAP respectively, can likewise be employed for this KPI.

2. **False Detection Rate (ID: I-KPI15):**
   - **Baseline:** Current false detection rate in point clouds.
   - **Target:** Reduction in false detection rate.
   - **Proposed Metric:**
     False Detection Rate (FDR) [5] [6]:

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}}, \qquad FDR < baseline \tag{26}$$

FP: False positives
TP: True positives

2.4.9 Real-Time Multimodal Weather Type Detector

1. **Weather Type Detection Accuracy (ID: I-KPI16):**
   - **Baseline:** Current accuracy of weather type detection in fused multimodal sensor data.
   - **Target:** Improvement in weather type detection accuracy.
   - **Proposed Metric:**
     For this KPI, the same definition of OCA provided in Equation 10 can be reused.

   - **Proposed Alternative Metric:**
     Macro-averaged F1-score ($F1_{macro}$) [5]:

$$\mathrm{F1}_{\mathrm{macro}} = \frac{1}{\mathrm{C}} \sum_{i=1}^{\mathrm{C}} \left(\frac{2 \times \mathrm{Precision}_i \times \mathrm{Recall}_i}{\mathrm{Precision}_i + \mathrm{Recall}_i} \right), \qquad \mathrm{F1}_{\mathrm{macro}} > baseline \tag{27}$$

C: Total number of classes
For the Precision and Recall calculations, the formulations defined in Equations 12 and 13 can be used.

2. **Detection Response Time (ID: I-KPI17):**
   - **Baseline:** Current response time for weather detection.
   - **Target:** Faster response times in multimodal sensor setup.
   - **Proposed Metric:**
     The AIT and FPS Equations 24 and 25 remain valid and can be applied in the evaluation of this KPI as well.

2.4.10 Harsh Weather Visibility Estimator

1. **Visibility Estimation Accuracy (ID: I-KPI18):**
    - **Baseline:** Current accuracy of the visibility estimator. The automotive industry is not using visibility range estimation functionality in actual environment perception systems.
    - **Target:** Visibility estimation capability installed, and range measurement is measured against radar detection range. The visibility estimator should provide output which gives optical sensor device measurement +-10 % tolerance in turbulent dry snowy or mid-dense fog.
    - **Proposed Metrics:**
      Mean Absolute Percentage Error of Visibility Estimator (MAPEvis) [10]:

$$\mathrm{MAPEvis} = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left|\frac{\mathrm{V}_{\mathrm{pred,i}} - \mathrm{V}_{\mathrm{gt,i}}}{\mathrm{V}_{\mathrm{gt,i}}}\right| x\ 100\%, \qquad MAPEvis < baseline \tag{28}$$

      Vpred, i: Predicted visibility range
      Vgt: Ground truth visibility range (Reference sensor measured range)
      N: Total number of measurements
      Unit: %

      Percentage Within Tolerance (PWT) [11]:

$$\mathrm{PWT} = \left(\frac{\text{Number of predictions in tolerance}}{\text{Total number of predictions}}\right) \times 100\%, \qquad PWT \leq 10\% \tag{29}$$

      Unit: %
2. **Visibility Estimation System Reliability (ID: I-KPI19):**
    - **Baseline:** Current reliability of the visibility estimator. No visibility range reference system at the automotive level is available. There are only infrastructure side expensive systems which do not meet automotive requirements.
    - **Target:** The reliability needs to be 85% or higher. The true-positive rate of seeing reference targets is more than 90% while the false-positive rate is less than 20%.
    - **Proposed Metrics:**
      Reliability (R):

$$\mathrm{R} = \left(\frac{\text{Number of correct measurements}}{\text{Total number of measurements}}\right) \times 100\%, \qquad R \geq 85\% \tag{30}$$

      Unit: %

      True Positive Rate (TPR) [12]:

$$\mathrm{TPR} = \left(\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}\right) \times 100\%, \qquad TPR \geq 90\% \tag{31}$$

      TP: True positives
      FN: False negatives
      Unit: %

      False Positive Rate (FPR) [12]:

$$\mathrm{FPR} = \left(\frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}\right) \times 100\%, \qquad FPR \leq 20\% \tag{32}$$

      FP: False positives
      TN: True negatives
      Unit: %

2.4.11 Weather Conditional Velocity Controller

1. **Velocity Adjustment compared to the experienced human driver (ID: I-KPI20):**
    - **Baseline:** Velocity adjustments done by experienced human drivers according to road slipperiness.
    - **Target:** Velocity limitations done by the velocity controller in slippery conditions should be lower than the human driver.
    - **Proposed Metrics:**
      Velocity Difference (VD):

$$\mathrm{VD} = \mathrm{V}_{\mathrm{human}} - \mathrm{V}_{\mathrm{system}}, \qquad VD > 0 \tag{33}$$

      V_{human}: Velocity of human driver (m/s or km/h)
      V_{system}: Velocity of the system (m/s or km/h)
      Unit: m/s or km/h

    - **Proposed Alternative Metric:**
      Velocity Rate (VR):

$$\mathrm{VR} = \frac{\mathrm{V}_{\mathrm{system}}}{\mathrm{V}_{\mathrm{human}}}, \qquad VR < 1 \tag{34}$$

2. **Robust Traction Control (ID: I-KPI21):**
    - **Baseline:** Number of traction loss events in automated driving mode (when accelerating or braking) without weather conditional velocity controller.
    - **Target:** Improvement in the number of traction loss events with weather conditional velocity control.
    - **Proposed Metric:**

Distance-based:
Distance-Based Traction Loss Event Rate (TLER):

$$\mathrm{TLERd} = \frac{\mathrm{N_{tle}}}{\mathrm{D}}, \quad \mathrm{TLER} < \mathrm{baseline} \tag{35}$$

N_{tle}: Number of traction loss events
D: Distance traveled
Unit: events/km

Time-based:
Time-Based Traction Loss Event Rate:

$$\mathrm{TLERt} = \frac{\mathrm{N_{tle}}}{\mathrm{T}}, \quad \mathrm{TLER} < \mathrm{baseline} \tag{36}$$

N_{tle}: Number of traction loss events
T: Time traveled
Unit: events/h

2.4.12 Development of a Solid State Lidar Target Simulator with a Tier 1 Company

1. **Distance accuracy of simulated targets across LiDAR channels (ID: I-KPI22):**
   - **Baseline:** Current deviation (%) between simulated wall distances and real wall measurements.
   - **Target:** Deviation less than 10% compared to real sensor measurements against real walls at the same distances.
   - **Proposed Metric:**
     Mean Absolute Percentage Error of Distance (MAPEdis) [10]:

$$\mathrm{MAPEdis} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\left|r_i^{\mathrm{sim}} - r_i^{\mathrm{real}}\right|}{r_i^{\mathrm{real}}}\right) \times 100\%, \qquad MAPEdis < baseline \tag{37}$$

N: Total number of test measurements (number of fixed wall positions x number of channels)
r_i^{sim}: Distance measured by LiDAR to the simulated wall in test i
r_i^{real}: Distance measured by LiDAR to the real wall in the same position
Unit: %

2. **Channel-to-channel consistency of simulated distances (ID: I-KPI23):**
   - **Baseline:** Current deviation (%) between channel measurements at the same simulated distance.
   - **Target:** Deviation is less than 5% between channels at a fixed simulated distance.
   - **Proposed Metric:**
     Maximum Relative Deviation (MRD):

$$\Delta_{\mathrm{chan}} = \frac{\max(r_c^{\mathrm{sim}}) - \min(r_c^{\mathrm{sim}})}{\overline{r^{\mathrm{sim}}}} \times 100\%, \qquad \Delta_{\mathrm{chan}} < 5\% \tag{38}$$

r_c^{sim}: Distance measured by LiDAR channel c at the simulated distance

$\max(r_c^{sim})$: Maximum distance measured among all channels

$\min(r_c^{sim})$: Minimum distance measured among all channels

$\overline{r^{sim}}$: Mean of all channel measurements at the simulated distance

Unit: %

2.5 Demonstration KPIs

As the last step to define the KPIs, the demonstration success rate needs to be considered.

1. **Scenario Performance (ID: D-KPI1):**
    - **KPI:** Success rate in diverse demonstration scenarios.
    - **Measurement:** Percentage of scenarios where the system performs as expected without failures.
    - **Proposed Metric:**
      Scenario Success Rate (SSR):

$$\mathrm{SSR} = \left(\frac{\mathrm{N_{success}}}{\mathrm{N_{total}}}\right) \times 100\%, \qquad SSR \approx 100\% \tag{39}$$

$N_{success}$: Number of successful scenarios
N_{total}: Total number of scenarios
Unit: %

2. **Environmental Adaptability (ID: D-KPI2):**
    - **KPI:** System adaptability to changing weather conditions during demonstrations.
    - **Measurement:** Number of successful adaptations to weather changes (e.g., transitioning from rain to snow) during a single demonstration.
    - **Proposed Metric:**
      Adaptation Success Rate (ASR):

$$\mathrm{ASR} = \left(\frac{\mathrm{N_{adapt\_success}}}{\mathrm{N_{adapt\_total}}}\right) \times 100\% \tag{40}$$

$N_{adapt\_success}$: Number of cases where the system continued to operate as expected after the environmental condition change
N_{adapt_total}: Total number of cases where an environmental condition change occurred
Unit: %

3. **Real-Time System Performance (ID: D-KPI3):**
    - **KPI:** Real-time processing and decision-making accuracy during demonstrations.
    - **Measurement:** Latency and accuracy of real-time system responses in demonstration environments.
    - **Proposed Metrics:**
      Average System Latency (ASL):

$$\mathrm{ASL} = \frac{\sum_{i=1}^{N}\left(\mathrm{T_{out,i}} - \mathrm{T_{in,i}}\right)}{\mathrm{N}} \tag{41}$$

T_{in}: Timestamp when sensor data or event enters the processing pipeline
T_{out}: Timestamp when the final system decision or output is produced
N: Total number of processed instances
Unit: milliseconds (ms)

Decision Accuracy (DA):

$$\mathrm{DA} = \left(\frac{\mathrm{N_{success}}}{\mathrm{N_{total}}}\right) \times 100\% \tag{42}$$

$N_{success}$: Number of system decisions during the demonstration that match the expected ground truth outcome or satisfy the predefined success criteria for the scenario.
N_{total}: Total number of decision points made by the system during the demonstration, including both successful and unsuccessful outcomes.
Unit: %

4. **Safety Metrics (ID: D-KPI4):**
   - **KPI:** Safety performance in live demonstrations.
   - **Measurement:** Number of safety incidents or near-misses recorded during demonstrations, aiming for zero occurrences.
   - **Proposed Metrics:**
     Distance-based:
     Distance-Based Incident Rate (IRd):

$$\mathrm{IRd} = \frac{N_{incidents}}{D} \tag{43}$$

$N_{incidents}$: Number of recorded safety incidents during demonstrations
D: Distance traveled
Unit: incidents/km

Time-based:
Time-Based Incident Rate (IRt):

$$\mathrm{IRt} = \frac{N_{incidents}}{T} \tag{44}$$

$N_{incidents}$: Number of recorded safety incidents during demonstrations
T: Time traveled
Unit: incidents/h

Distance-based:
Distance-Based Near-Miss Rate (NMRd):

$$\mathrm{NMRd} = \frac{N_{near\_miss}}{D} \tag{45}$$

N_{near_miss}: Number of events where a collision or unsafe situation was narrowly avoided
D: Distance traveled
Unit: events/km

Time-based:
Time-Based Near-Miss Rate (NMRt):

$$\mathrm{NMRt} = \frac{N_{near\_miss}}{T} \tag{46}$$

N_{near_miss}: Number of events where a collision or unsafe situation was narrowly avoided
T: Total demonstration time
Unit: events/h

5. **Timeliness of Demonstrations (ID: D-KPI5):**
   - **KPI:** Adherence to scheduled demonstration timelines.
   - **Measurement:** Percentage of demonstrations conducted on or before the planned dates.
   - **Proposed Metric:**
     On-Time Demonstration Rate (OTDR):

$$\mathrm{OTDR} = \left(\frac{N_{on\_time}}{N_{total}}\right) \times 100\% \tag{47}$$

N_{on_time}: Number of demonstrations completed on or before the scheduled date
N_{total}: Total number of planned demonstrations
Unit: %

2.6 Performance KPIs

The performance metrics initially defined in Deliverable D7.2 have been reviewed and, where relevant, integrated into the final KPI framework presented in this deliverable. Together with the additional indicators developed in WP8, these metrics form the foundation for assessing the ROADVIEW system across all system-level evaluations and demonstrations. By combining the outputs of D7.2 with the refinements introduced in the previous deliverable D8.6, the project ensures that the KPI definitions remain consistent, measurable, and directly applicable to the evaluation of automated driving performance under adverse weather conditions.

Tables 1–3 present the complete set of KPIs defined in this deliverable, structured into three categories: **system-level KPIs** (Table 1), **innovation-level KPIs** (Table 2), and **demonstration-level KPIs** (Table 3). In the subsequent performance assessment (D8.8), these KPIs will be systematically applied to analyse test results and validate system capabilities. The KPI identification structure has been maintained as previously defined: IDs starting with "S" denote system/project-related KPIs, IDs starting with "I" refer to innovation-related KPIs, and IDs starting with "D" correspond to demonstration-specific KPIs. This standardised approach ensures traceability and consistency across all reporting and analysis activities.

Table 1. List of identified system-level KPIs.

No	Main Responsible	Title of Innovation	Number of KPIs	KPI ID
1	ALL	System Resilience	1	S-KPI1
2	ALL	Failure Rate	1	S-KPI2
3	ALL	Uptime/Downtime	1	S-KPI3
4	ALL	Human Intervention	1	S-KPI4
5	ALL	Execution Time	1	S-KPI5

Table 2. List of identified innovation-level KPIs.

No	Main Responsible	Title of Innovation	Number of KPIs	KPI ID
1	CRF	Infrastructure-Based Perception System Including Weather Conditions for Collaborative Perception Solution Using V2X	2	I-KPI1, I-KPI2
2	CRF	Weather-Aware Infrastructure-Based Manoeuvre Cooperation for Vehicle Control and Decision-Making System Using V2X	1	I-KPI3
3	FGI	Environment-Aware High-Definition Mapping Method for Real-Time Positioning of an Autonomous Vehicle on the Map	2	I-KPI4, I-KPI5
4	FGI	Optical Measurement of Road Grip and Road Surface Conditions in front of the Vehicle	2	I-KPI6, I-KPI7
5	HH	Fast and Accurate Multi-Sensor Fusion for Joint Multitask Learning	2	I-KPI8, I-KPI9
6	HH	A Fast and Memory Efficient Outlier Detector for 3D LiDAR Point Clouds	2	I-KPI10, I-KPI11
7	HH	A Fast Free Space Detector Using LiDAR Point Clouds	2	I-KPI12, I-KPI13

No	Main Responsible	Title of Innovation	Number of KPIs	KPI ID
8	HH	Semantics-Aware Pseudo Point Cloud-Based 3D Object Detector	2	I-KPI14, I-KPI15
9	HH	A Real-Time Multimodal Weather Type Detector for Autonomous Vehicles	2	I-KPI16, I-KPI17
10	VTT	Harsh Weather Visibility Estimator for Safe Automated Driving	2	I-KPI18, I-KPI19
11	VTT	Weather Conditional Velocity Controller for Safe Automated Driving in Harsh Weather Conditions	2	I-KPI20, I-KPI21
12	KO	Development of a Solid State Lidar Target Simulator with a Tier 1 Company	2	I-KPI22, I-KPI23

Table 3. List of identified demonstration-level KPIs.

No	Main Responsible	Title of Innovation	Number of KPIs	KPI ID
1	FORD, AVL, VTT, and THI	Scenario Performance	1	D-KPI1
2	FORD, AVL, VTT, and THI	Environmental Adaptability	1	D-KPI2
3	FORD, AVL, VTT, and THI	Real-Time System Performance	1	D-KPI3
4	FORD, AVL, VTT, and THI	Safety Metrics	1	D-KPI4
5	FORD, AVL, VTT, and THI	Timeliness of Demonstrations	1	D-KPI5

2.7 Impact Assessment of Final KPIs

This section explains how the final KPI framework supports the overall objectives and expected impacts of ROADVIEW, with a particular focus on operation in harsh weather and adverse road conditions. In addition to verifying performance capabilities, the KPI framework serves to identify the operational limits of the system. While open-road demonstrations focus on safe operation within the ODD, data from controlled environments (such as weather chambers and X-in-the-loop simulations) and analysis of safety-critical interventions (S-KPI4) will be used to determine the boundaries where the system can no longer operate effectively.

The KPIs have been defined in three categories:

- System-level KPIs (S-KPIs) describing the behaviour of the integrated ROADVIEW system.
- Innovation-level KPIs (I-KPIs) capturing the performance of specific technological innovations.
- Demonstration-level KPIs (D-KPIs) describing how the system performs in the planned use cases and demonstrations.

While all of these KPIs contribute to the assessment of the system, this section concentrates on those that are directly related to harsh weather, since they are central to the project's ambition of achieving robust automated driving in rain, snow, fog and on low-friction roads.

2.7.1 System-Level KPIs and Project Impacts

At system level, ROADVIEW aims to maintain safe and reliable operation under adverse weather, to extend the operational design domain (ODD) to harsher conditions without unacceptable degradation of performance, and to ensure that the integrated system remains sufficiently operational and efficient for realistic deployment.

Furthermore, the assessment addresses the system's operational limits. Since testing failure modes on public roads is safety-critical, the exact boundaries of the system points where weather intensity exceeds the system's capabilities are primarily explored through controlled tests in weather chambers and X-in-the-loop simulations. In the real-world demonstrations, these limits are characterised by the frequency of human interventions (S-KPI4) and the degradation of availability (S-KPI1) as weather conditions worsen.

2.7.1.1 System Resilience and Availability (S-KPI1, S-KPI3)

Objective: These KPIs quantify for how much of the time the ROADVIEW system is fully operational, and in particular how often it remains available during rain, snow and fog.

Impact: High Weather-Normalised Availability (WNA) and high overall Availability (A) in adverse conditions indicate that the system does not simply fall back or hand over whenever the weather becomes challenging, but can keep operating within its intended ODD. This directly supports the goal of robust, all-weather operation.

2.7.1.2 Failure Rate (S-KPI2)

Objective: The run-based and time-based failure rates capture how often the integrated system or its subsystems stop working during operation.

Impact: A low failure rate, and in particular the absence of systematic failure patterns correlated with harsh weather, is essential to demonstrate reliability and safety in those conditions.

2.7.1.3 Human Intervention (S-KPI4)

Objective: These KPIs measure how often a human has to intervene, and with which criticality. The criticality-weighted variants (CWHIr, CWHIt) emphasise safety-critical and performance-critical interventions.

Impact: Frequent safety-critical interventions during adverse weather indicate that the combined complexity of the weather and traffic environment exceeds the system's autonomous capabilities. Conversely, a low rate of critical interventions demonstrates that the system can safely handle the complete scenario without requiring human fallback.

2.7.1.4 Execution Time (S-KPI5)

Objective: Average execution times of key ROADVIEW components determine whether perception, fusion and decision-making can run in real time.

Impact: Real-time capability is necessary to react appropriately at the speeds and in the scenarios used in the demonstrations. If execution times grow too large in harsh weather (e.g. due to additional processing), safety margins shrink. Keeping these KPIs within acceptable bounds under adverse conditions is therefore an important part of demonstrating readiness for real-world use.

In the performance assessment (D8.8), these system-level KPIs will be examined for adverse weather conditions, taking into account the different types and intensities of rain, snow and fog covered by the use cases and available data. The central question will be whether the integrated system maintains high availability, low failure and intervention rates, and acceptable execution times when exposed to these adverse-weather scenarios.

2.7.2 Innovation-Level KPIs and Project Impacts

The innovation-level KPIs measure the contribution of specific technical developments that are introduced precisely to cope with harsh weather and difficult road conditions, and they form the bridge between algorithm-level performance and overall system behaviour. In this context, they can be grouped into three main impact areas. The first area is weather and environment state estimation, where the corresponding innovations provide information about the current environment (for example, accurate detection of weather type, visibility range and road surface conditions).

Beyond measuring peak performance, these I-KPIs are also used to characterize the degradation curves of the sensors and algorithms. This helps determine the specific weather intensities (e.g. rainfall rate or fog density) at which the perception and estimation modules can no longer provide reliable data.

2.7.2.1 Weather and weather-type detection (I-KPI1, I-KPI16)

Objective: Quantify how accurately the system can detect and classify weather conditions and weather type from sensor data.

Impact: Reliable weather detection is a prerequisite for adapting perception pipelines, navigation strategies and velocity. High values for OCA and $F1_{macro}$ in adverse conditions show that the system can correctly recognise when it needs to switch to "bad weather" behaviour.

2.7.2.2 Visibility estimation (I-KPI18, I-KPI19)

Objective: Measure the accuracy and reliability of the estimated visibility range around the vehicle.

Impact: The navigation and decision-making modules explicitly require knowledge of how far ahead the vehicle can reliably perceive objects. If the visibility estimator is biased or noisy, the system may either become overly conservative or drive too fast for the actual conditions. Low MAPEvis, high Percentage Within Tolerance (PWT), and good reliability metrics (R, TPR, FPR) provide evidence that visibility-based decisions rest on a solid basis.

2.7.2.3 Road grip and surface condition estimation (I-KPI6, I-KPI7)

Objective: Quantify how well the system can estimate road friction and grip variability ahead of the vehicle.

Impact: Accurate grip maps under diverse road conditions (wet, icy, snowy) enable better control decisions (e.g. braking and acceleration limits). Improvements in $RMSE_{ratio}$ and CE_{model} over baseline methods indicate that the grip estimation innovation contributes meaningfully to safe operation on slippery roads.

The second impact area is weather-aware control and manoeuvre planning. Here, the innovations use the environment and weather estimates described above (e.g. visibility range, road grip and surface conditions) to adjust the vehicle's behaviour, including its speed profile, longitudinal and lateral control, and its interaction with infrastructure-based manoeuvre coordination. In this way, the controller and planning functions can explicitly account for adverse weather and low-friction conditions rather than treating all situations as if they were clear and dry.

2.7.2.4 Weather-conditional velocity controller (I-KPI20, I-KPI21)

Objective: I-KPI20 compares the system's chosen velocity to that of an experienced human driver under the same low-friction conditions; I-KPI21 measures the rate of traction-loss events.

Impact: The intention is that, in slippery conditions, the automated controller should drive at least as cautiously as an experienced human, and should reduce traction-loss events. Values where the system velocity is consistently lower than the human reference (VR < 1) and where traction-loss event rates decrease compared to a baseline demonstrate that this objective is being met.

2.7.2.5 Weather-aware manoeuvre cooperation (I-KPI3)

Objective: Measure how often the in-vehicle system accepts minimal-risk manoeuvre (MRM) requests coming from the infrastructure side during harsh weather.

Impact: A high Manoeuvre Coordination Acceptance Rate (MCAR) in relevant scenarios indicates that infrastructure-based manoeuvre suggestions are compatible with the on-board understanding of the situation and weather, and that the cooperation mechanism is practically usable in real traffic.

The third impact area is robust perception and multi-sensor fusion in harsh weather. In this area, the innovations are designed to keep perception quality high even when individual sensors are disturbed or degraded by adverse conditions such as rain, snow and fog. By combining information from multiple complementary sensors and applying weather-aware fusion, outlier removal and free-space detection methods, the system is able to maintain a reliable understanding of its surroundings under conditions where conventional, single-sensor or non-adaptive perception stacks typically experience a significant drop in performance.

2.7.2.6 Multi-sensor fusion for 3D perception and free-space detection (I-KPI8, I-KPI9, I-KPI11, I-KPI12, I-KPI14, I-KPI15)

Objective: Quantify accuracy of 3D object detection, semantic segmentation, free-space detection and outlier removal using metrics such as mAP, mIoU and FDR.

Impact: In rain, fog and snow, individual sensors degrade differently. The fusion and perception innovations aim to compensate for this by combining information. Higher mAP/mIoU and lower FDR in adverse-weather datasets, compared to simpler baselines, show that the ROADVIEW perception stack maintains useful environment understanding where conventional methods struggle.

2.7.2.7 Latency and processing efficiency under load (I-KPI2, I-KPI8, I-KPI10, I-KPI13, I-KPI17)

Objective: Measure delays and processing loads introduced by collaborative perception, fusion, outlier detection and weather-type detection.

Impact: These functions must not only be accurate but also fast enough. By observing EVDL, AFPT, AIT and FPS in harsh-weather scenarios, we can check that the improved perception does not compromise real-time safety when the system is operating at realistic speeds.

Overall, the innovation-level KPIs explain the origins of the expected system-level improvements. By showing that each major building block of the ROADVIEW concept provides a measurable benefit over its respective baseline, particularly under adverse weather conditions where conventional CAV technology tends to degrade, they justify why better system-level performance in harsh weather can be anticipated.

2.7.3 Demonstration-Level KPIs and Project Impacts

The demonstration-level KPIs translate these technical advances into end-to-end performance in realistic test campaigns.

2.7.3.1 Scenario Success Rate (D-KPI1)

Objective: Percentage of scenarios that are completed successfully without failures.

Impact: High SSR in scenarios that explicitly include adverse weather (as defined in the use cases) shows that the integrated system can complete representative tasks (e.g. lane entering/exiting, intersection handling, driving on snow-covered roads) under harsh conditions.

2.7.3.2 Environmental Adaptability (D-KPI2)

Objective: Success rate of system adaptations to environmental variances. This is measured via dynamic transitions in simulation (e.g. clear to foggy), or by comparing performance across distinct real-world scenarios (e.g. separate runs in dry vs. snowy conditions).

Impact: A high ASR indicates that the system can remain in operation while the environment changes, instead of failing or requiring a takeover whenever weather changes. This directly supports the notion of resilience and an extended ODD.

Failed scenarios or necessary safety interventions during these demonstrations are equally valuable, as they practically define the boundaries of the ODD. They indicate the precise real-world conditions where the automated system reaches its limit and requires human fallback.

2.7.3.3 Real-Time System Performance (D-KPI3)

Objective: Measure the average system latency from input to decision and the accuracy of those decisions in the demonstration environment.

Impact: These KPIs verify that, under integrated load and realistic conditions, the system still reacts in time and makes correct decisions. This is a necessary condition for demonstrating readiness for further deployment in adverse-weather scenarios.

2.7.3.4 Safety Metrics (D-KPI4)

Objective: Quantify safety incidents and near-misses. Due to limited testing time in real-world demos, this will focus on the absolute count of interventions/incidents. Statistical rates (incidents per distance/hour) will be primarily derived from simulation scenarios.

Impact: The primary safety goal is that the ROADVIEW system does not introduce systematic safety issues in the harsh-weather scenarios. Very low (ideally zero) incident and near-miss rates over the demonstration distances support this claim.

2.7.3.5 Timeliness of Demonstrations (D-KPI5)

Objective: Measure whether demonstrations are carried out as planned, in line with the demonstration schedule.

Impact: While not weather-specific, timely delivery of demonstrations in real or emulated adverse weather is important for stakeholder engagement, project credibility and overall impact realisation.

Together, these D-KPIs allow the project to argue not only that individual functions work in isolation, but that complete vehicles equipped with ROADVIEW technologies can safely and reliably execute complex tasks under challenging weather and road conditions.

2.7.4 Interpretation of KPI Values and Thresholds

For many KPIs, it might seem straightforward to define a single numerical target (for example, “availability must be at least X%”). In reality, however, the values that are achievable and acceptable depend strongly on the context. Relevant factors include the vehicle type (e.g. truck vs. passenger car), the scenario design (urban vs. highway, traffic density and complexity), the intensity and combination of adverse weather (light rain vs. heavy snow, mixed conditions), and the speed profile and total test mileage. A fixed threshold that is reasonable for one use case may be unrealistic, too weak, or simply not relevant for another. Setting universal numerical targets at this stage would therefore either over-constrain some cases or appear arbitrary to reviewers.

For this reason, the KPI framework in this deliverable focuses on providing a clear basis for later quantitative assessment, rather than on prescribing fixed numbers. Specifically, we define what each KPI measures, state clearly in which direction improvement is desired (higher is better / lower is better), and identify suitable baselines and comparison points. These baselines include, where applicable, non-weather-aware systems, clear-weather operation, simpler or state-of-practice methods, and human-driver behaviour when data are available. This ensures that KPI results can be interpreted against realistic and transparent reference points, without committing to one-size-fits-all thresholds that may not hold across vehicles, scenarios and weather conditions.

The detailed numerical interpretation of KPI values will be done in the follow-up deliverable D8.8, once data from the demonstrations and X-in-the-loop tests are available. In D8.8, the KPI values will be compared systematically to their baselines and analysed in relation to the physical and operational constraints of each use case, such as braking distances, friction conditions, speed limits and typical human behaviour. Where relevant, existing regulations, standards and state-of-the-art performance levels will also be considered. In this way, conclusions about impact are based on the actual ROADVIEW results, not on assumptions made in advance.

Within this approach, a KPI is considered to support the intended impact if it shows a consistent and practically meaningful improvement over its baseline in the relevant harsh-weather scenarios, and its absolute value is compatible with safe and reliable operation under those conditions. Thus, both relative improvement and absolute adequacy are taken into account, case by case. This method avoids arbitrary, globally fixed thresholds that could be questioned later, while still providing a clear and defensible link between the measured KPI values and ROADVIEW’s objectives for robust automated driving in adverse and extreme weather.

3 Conclusions

The ROADVIEW project marks a significant step forward in advancing CAVs capable of operating safely and reliably under adverse weather and varying road conditions. The comprehensive KPI framework finalised in this deliverable provides a robust and validated basis for systematically evaluating system resilience, reliability, and operational efficiency. By building on the initial definitions from D8.6, incorporating performance metrics from WP7, and refining thresholds and measurement methods through real-world demonstrations, the project has ensured that its performance assessment approach is both technically sound and operationally relevant.

The finalised KPIs offer measurable, traceable, and repeatable criteria that will be applied in the forthcoming performance assessment (D8.8), enabling an objective evaluation of the integrated ROADVIEW system. These metrics will facilitate the identification of strengths, highlight areas requiring further optimisation, and validate the system’s readiness for deployment in real-world scenarios. The measurable outcomes—such as maintaining high operational uptime in severe weather, reducing subsystem failure rates, and minimising human intervention—will serve as key indicators of project success.

Through the collaborative efforts of leading research institutions, Small and Medium-sized Enterprises (SMEs), and industry partners, ROADVIEW has not only met its KPI specification objectives but also established a structured evaluation methodology that can serve as a benchmark for future weather-resilient autonomous driving projects. This deliverable completes the KPI definition phase, laying the groundwork for comprehensive performance validation and further contributing to the development of fail-safe, weather-aware CAV technologies.

4 References

[1] ISO 22400-2:2014, *Automation systems and integration — Key performance indicators (KPIs) for manufacturing operations management — Part 2: Definitions and descriptions*, International Organization for Standardization, Geneva, Switzerland, 2014.

[2] ROADVIEW Project, "Deliverable D8.6 – Initial KPI Specifications," unpublished internal project report, 2024.

[3] J. Romani, *WEDISTRICT Project Deliverable D2.2 – KPIs Definition*, 2019. [Online]. Available: https://www.wedistrict.eu/wp-content/uploads/2020/08/WEDISTRICT_WP2_D2.2-KPIs-definition.pdf. [Accessed: Aug. 26, 2025]. Project website: https://www.wedistrict.eu/project/

[4] California DMV, "Autonomous Vehicle Disengagement Reports 2023," California DMV, Sacramento, CA, USA, 2024. [Online]. Available: https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/disengagement-reports/. [Accessed: Aug. 25, 2025].

[5] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[6] ROADVIEW Project, "Deliverable D7.2 – Relevant metrics to assert the validity of the simulation-assisted test methods," unpublished internal project report, 2024.

[7] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2012, pp. 3354–3361. doi: 10.1109/CVPR.2012.6248074.

[8] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010. doi: 10.1007/s11263-009-0275-4.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91

[10] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. Melbourne, Australia: OTexts, 2018. [Online]. Available: https://otexts.com/fpp2/. [Accessed: Aug. 18, 2025].

[11] J. Levinson and S. Thrun, "Robust vehicle localization in urban environments using probabilistic maps," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, Anchorage, AK, USA, 2010, pp. 4372–4378, doi: 10.1109/ROBOT.2010.5509700

[12] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006. doi: 10.1016/j.patrec.2005.10.010