Big Data technologies and extreme-scale analytics

Multimodal Extreme Scale Data Analytics for Smart Cities Environments

D6.4: Final Assessment Report and Impact Analysis[†]

Abstract: This document provides a detailed report on the evaluation activities carried out mainly during the second reporting period (RP2) but also includes updates to the first reporting period RP1 (D6.2), to assess the implementation of the use cases and the respective societal impact, from the point of view of the use case providers, end users and the general public. The assessment is carried out against a set of pre-defined use case parameters and specific KPIs. The evaluation on the one hand leverages technological asset-specific evaluations and benchmarking carried out and reported in a number of work package deliverables, namely D3.5 and D5.5 and on the other hand makes use of qualitative surveys and interviews completed by end users and feedback obtained during project information days and events.

Contractual Date of Delivery	31/12/2023
Actual Date of Delivery	22/02/2024
Deliverable Security Class	Public
Editor	*Adrian Muscat (GRN)*
Contributors	All *MARVEL* partners
Quality Assurance	*Despina Kopanaki (FORTH)* *Dusan Jakovetic (UNS)*

[†] *The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957337.*

The *MARVEL* Consortium

Part. No.	Participant organisation name	Participant Short Name	Role	Country
1	FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS	FORTH	Coordinator	EL
2	INFINEON TECHNOLOGIES AG	IFAG	Principal Contractor	DE
3	AARHUS UNIVERSITET	AU	Principal Contractor	DK
4	ATOS SPAIN SA	ATOS	Principal Contractor	ES
5	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR	Principal Contractor	IT
6	INTRASOFT INTERNATIONAL S.A.	INTRA	Principal Contractor	LU
7	FONDAZIONE BRUNO KESSLER	FBK	Principal Contractor	IT
8	AUDEERING GMBH	AUD	Principal Contractor	DE
9	TAMPERE UNIVERSITY	TAU	Principal Contractor	FI
10	PRIVANOVA SAS	PN	Principal Contractor	FR
11	SPHYNX TECHNOLOGY SOLUTIONS AG	STS	Principal Contractor	CH
12	COMUNE DI TRENTO	MT	Principal Contractor	IT
13	UNIVERZITET U NOVOM SADU FAKULTET TEHNICKIH NAUKA	UNS	Principal Contractor	RS
14	INFORMATION TECHNOLOGY FOR MARKET LEADERSHIP	ITML	Principal Contractor	EL
15	GREENROADS LIMITED	GRN	Principal Contractor	MT
16	ZELUS IKE	ZELUS	Principal Contractor	EL
17	INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK	PSNC	Principal Contractor	PL

Document Revisions & Quality Assurance

Internal Reviewers

1. *Despina Kopanaki, FORTH*
2. *Dusan Jakovetic, UNS*

Revisions

Version	Date	By	Overview
4.0	22/02/2024	FORTH, UNS	Approval and submission
3.0	22/02/2024	GRN	Final draft
2.1	15/02/2024	FORTH	Input and Comments on the 2nd draft
2.0	9/02/2024	GRN	2nd draft
1.1	31/01/2024	UNS	Comments on the 1st draft
1.0	23/01/2024	GRN	1st Draft
0.2	28/11/2023	GRN	Final ToC
0.1	25/11/2023	FORTH	Comments on ToC
0.1	01/1120123	STPM, WPL	Comments on the ToC.
0.0	09/10/2023	GRN	ToC Draft

Disclaimer

Table of Contents

LIST OF TABLES 7

LIST OF FIGURES 9

LIST OF ABBREVIATIONS 10

EXECUTIVE SUMMARY 12

1 INTRODUCTION 13

1.1 PURPOSE AND SCOPE OF THIS DOCUMENT 13
1.2 INTENDED READERSHIP 13
1.3 CONTRIBUTION TO WP6 AND PROJECT OBJECTIVES 14
1.4 RELATION TO OTHER WPS AND DELIVERABLES 14
1.5 STRUCTURE OF THE DOCUMENT 15

2 THE MARVEL PROJECT AND ITS EVALUATION 16

2.1 SUMMARY OF THE MARVEL PROJECT AND FRAMEWORK 16
2.2 TYPES OF KPIS AND EVALUATION PARAMETERS 17
2.3 BENCHMARKING AND EVALUATION STRATEGY 19

3 USE CASE GRN1: SAFER ROADS 21

3.1 INTRODUCTION 21
3.2 SCOPE AND DESCRIPTION OF USE CASE 21
3.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 21
3.4 KPI ASSESSMENT 24
3.4.1 Use case specific KPIs 24
3.4.2 Use case specific non-functional KPIs 25
3.4.3 Additional parameters common across all GRN use cases from D6.1/D6.2 26

4 USE CASE GRN2: ROAD USER BEHAVIOUR 29

4.1 INTRODUCTION 29
4.2 SCOPE AND DESCRIPTION OF USE CASE 29
4.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 29
4.4 KPI ASSESSMENT 32
4.4.1 Use case specific KPIs 33
4.4.2 Use case specific non-functional KPIs 34
4.4.3 Additional parameters common across all GRN use cases from D6.1/D6.2 35

5 USE CASE GRN3: TRAFFIC CONDITIONS AND ANOMALOUS EVENTS 36

5.1 INTRODUCTION 36
5.2 SCOPE AND DESCRIPTION OF USE CASE 36
5.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 37
5.4 KPI ASSESSMENT 40
5.4.1 Use case specific KPIs 40
5.4.2 Use case specific non-functional KPIs 41
5.4.3 Additional parameters common across all GRN use cases from D6.1/D6.2 42

6 USE CASE GRN4: JUNCTION TRAFFIC AND TRAJECTORY COLLECTION 43

6.1 INTRODUCTION 43
6.2 SCOPE AND DESCRIPTION OF USE CASE 43
6.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 44
6.4 KPI ASSESSMENT 46
6.4.1 Use case specific KPIs 47
6.4.2 Use case specific non-functional KPIs 48
6.4.3 Additional parameters common for all GRN use cases from D6.1/D6.2 49

7 USE CASE MT1: MONITORING OF CROWDED AREAS 50

7.1 INTRODUCTION 50
7.2 SCOPE AND DESCRIPTION OF USE CASE 50
7.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 50
7.4 KPI ASSESSMENT 53
7.4.1 Use case specific functional KPIs 53
7.4.2 Use case specific non-functional KPIs 55
7.4.3 Additional parameters common across all MT use cases from D6.1/D6.2 56

8 USE CASE MT2: DETECTING CRIMINAL AND ANTI-SOCIAL BEHAVIOURS 59

8.1 INTRODUCTION 59
8.2 SCOPE AND DESCRIPTION OF USE-CASE 59
8.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 60
8.4 KPI ASSESSMENT 62
8.4.1 Use case specific KPIs 62
8.4.2 Use case specific non-functional KPIs 64
8.4.3 Additional parameters common across all MT use cases from D6.1/D6.2 65

9 USE CASE MT3: MONITORING OF PARKING PLACES 66

9.1 INTRODUCTION 66
9.2 SCOPE AND DESCRIPTION OF USE CASE 66
9.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 67
9.4 KPI ASSESSMENT 69
9.4.1 Use case specific KPIs 69
9.4.2 Use case specific non-functional KPIs 71
9.4.3 Additional parameters common across all MT use cases from D6.1/D6.2 72

10 USE CASE MT4: ANALYSIS OF A SPECIFIC AREA 73

10.1 INTRODUCTION 73
10.2 SCOPE AND DESCRIPTION OF USE-CASE 73
10.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 74
10.4 KPI ASSESSMENT 77
10.4.1 Use case specific KPIs 77
10.4.2 Use case specific non-functional KPIs 79
10.4.3 Additional parameters common across all MT use cases from D6.1/D6.2 80

11 USE CASE UNS1: DRONE EXPERIMENT 81

11.1 INTRODUCTION 81
11.2 SCOPE AND DESCRIPTION OF USE CASE 81
11.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 82
11.4 KPI ASSESSMENT 84
11.4.1 Use case specific KPIs 85
11.4.2 Use case specific non-functional KPIs 87
11.4.3 Additional parameters common across all UNS use cases 89

12 USE CASE UNS2: LOCALISING AUDIO EVENTS IN CROWDS 91

12.1 INTRODUCTION 91
12.2 SCOPE AND DESCRIPTION OF USE-CASE 91
12.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 91
12.4 KPI ASSESSMENT 93
12.4.1 Use case specific KPIs 93
12.4.2 Use case specific non-functional KPIs 95
12.4.3 Additional parameters common across all UNS use cases from D6.1/D6.2 96

13 EARLY ADOPTERS: THE GOZO MVP 97

13.1 INTRODUCTION 97
13.2 SCOPE AND DESCRIPTION OF USE CASE 97
13.3 SUMMARY OF TECHNICAL SETUP AND DECISION MAKING TOOLKIT 98
13.4 KPI ASSESSMENT 100
13.4.1 Use case specific KPIs 100
13.4.2 Use case specific non-functional KPIs 101

13.4.3 *Additional parameters common for all GRN use cases from D6.1/D6.2* ... 102

14 PROJECT WIDE KPIS AND EVALUATION PARAMETERS ... 103

14.1 PROJECT KPIS ... 103
14.1.1 *Project scientific and technical objectives* ... 103
14.1.1.1 *Objective 1: Leverage innovative technologies for data acquisition, management and distribution to develop a privacy-aware engineering solution for revealing valuable and hidden societal knowledge in a smart city environment* ... 103
14.1.1.2 *Objective 2: Deliver AI-based multimodal perception and intelligence for audio-visual scene recognition, event detection and situational awareness in a smart city environment* ... 107
14.1.1.3 *Objective 3: Break technological silos, converge very diverse and novel engineering paradigms and establish a distributed and secure Edge-to-Fog-to-Cloud (E2F2C) ubiquitous computing framework in the big data value chain* ... 108
14.1.1.4 *Objective 4: Realise societal opportunities in a smart city environment by validating tools and techniques in real- world settings* ... 112
14.1.1.5 *Objective 5: Foster the European Data Economy vision and create new scientific and business opportunities by offering the MARVEL Data Corpus as a free service and contributing to BDVA standards* 113
14.1.2 *Impact* ... 114
14.1.2.1 *Impact related to the work programme* ... 114
14.1.2.2 *Impact on innovation capacity, competitiveness and growth* ... 118
14.1.2.3 *Impact on standards* ... 121
14.2 SOCIETAL GOALS EVALUATION AND IMPACT ... 121
14.3 BUSINESS GOALS EVALUATION AND IMPACT ... 126

15 FEEDBACK FROM EXTERNAL ADVISORY BOARDS AND EXTERNAL STAKEHOLDERS 130

15.1 MARVEL EXTERNAL ADVISORY AND ETHICS BOARDS ... 130
15.2 FEEDBACK FROM EXTERNAL STAKEHOLDERS ... 134
15.2.1 *ICSC2022 (26-29.09.2022) MARV: Multimodal and AI-Responsible data processing and deliVery in smart cities* ... 134
15.2.2 *Vision zero – Belgrade* ... 135
15.2.3 *EBDVF2022 (21-23.11.2022)* ... 135
15.2.4 *Second Info Day in Malta (28.11.2022)* ... 136
15.2.5 *EBDVF2023 (October 2023)* ... 137
15.2.6 *Smart City Expo (Nov 2023)* ... 137
15.2.7 *Third Info Day – online (4.12.2023)* ... 137

16 SUMMARY AND CONCLUSIONS ... 140

17 APPENDIX : SURVEY METHODOLOGIES AND DOCUMENTS ... 141

17.1 MALTA PILOT - INTRODUCTION TO METHODOLOGIES AND DOCUMENTS ... 141
17.1.1 *Malta pilot – Online Surveys* ... 141
17.1.2 *S1 – survey for GRN1* ... 141
17.1.3 *S2 - survey for GRN2* ... 145
17.1.4 *S3 – survey for GRN3* ... 149
17.1.5 *S4 - survey for GRN4* ... 152
17.1.6 *S5 – AI Systems for Sustainable Mobility - The Malta Pilot survey* ... 157
17.1.7 *Malta pilot and Gozo MVP– Interviews* ... 165
17.2 TRENTO PILOT - INTRODUCTION TO METHODOLOGIES AND DOCUMENTS ... 166
17.2.1 *Trento pilot – Online Surveys* ... 166
17.2.2 *Survey for MT1-Crowd Monitoring - Trento Pilot* ... 166
17.2.3 *Survey for MT2-Detecting criminal and anti-social behaviours* ... 170
17.2.4 *Survey fot MT3-Monitoring of parking places* ... 173
17.2.5 *Survey for MT4-Analysis of a specific area* ... 177
17.2.6 *Trento pilot – Interviews* ... 181
17.3 UNS EXPERIMENT - INTRODUCTION TO METHODOLOGIES AND DOCUMENTS ... 181
17.3.1 *UNS1 survey* ... 182
17.3.2 *UNS2 survey* ... 185

List of Tables

Table 1: Evaluation scenario and relevant KPIs for the GRN1 (Source D6.3, Table 5) 21
Table 2: GRN1 functional KPIs (source: D6.3[6]) 24
Table 3: GRN1 non-functional KPIs (source: D6.3) 25
Table 4: Parameters that are common across GRN use cases... 26
Table 5: Evaluation scenarios and relevant KPIs for the GRN2 (source: D6.3)................................. 29
Table 6: GRN RP2 functional KPIs (Source: D6.3 and D1.2) 33
Table 7: GRN2 non-functional KPIs.. 34
Table 8: Evaluation scenario and relevant KPIs for the GRN3: Traffic Conditions and Anomalous Events (Table 6, D6.1) 36
Table 9: Use case specific functional KPIs for GRN3: Traffic Conditions and Anomalous Events (Table 6.2 in D1.2) 40
Table 10: Use case specific non functional evaluation variables for GRN3: Traffic Conditions and Anomalous Events (6.3 in D1.2).. 42
Table 11: Evaluation scenarios and relevant KPIs for the GRN4: Junction Traffic and Trajectory Collection (Table 9, D6.1) 43
Table 12: Use case specific KPIs for GRN4: Junction Traffic and Trajectory Collection (Table 6.2 in D1.2)... 47
Table 13: Use case specific non functional evaluation variables for GRN4: Junction Traffic and Trajectory Collection (Table 6.3 in D1.2).. 49
Table 14: Evaluation scenarios and relevant KPIs for the MT1: Monitoring of Crowded Areas (Table 11, D6.1).. 50
Table 15: Use case specific KPIs for MT1: Monitoring of Crowded Areas (Table 6.2 in D1.2).......... 53
Table 16: Use case specific non functional evaluation variables for MT1: Monitoring of Crowded Areas (from Table 6.3 in D1.2) 55
Table 17: Additional parameters common across all MT use cases 56
Table 18: Evaluation scenarios and relevant KPIs for the MT2 59
Table 19: Use case specific KPIs for MT2: Detecting Criminal and Anti-Social Behaviours (Table 6.2 in D1.2) 62
Table 20: Use case specific non functional evaluation variables for MT2: Detecting Criminal and Anti-Social Behaviours (from Table 6.3 in D1.2) 64
Table 21: Evaluation scenarios and relevant KPIs for the MT3, (Table 13, D6.1) 66
Table 22: Use case specific KPIs for MT3: Monitoring of Parking Spaces (Table 6.2 in D1.2) 69
Table 23: Use case specific non functional evaluation variables for MT1: Monitoring of Parking Spaces (from Table 6.3 in D1.2) 71
Table 24: Evaluation scenarios and relevant KPIs for the MT4 73
Table 25: Use case specific KPIs for MT4: Analysis of a Specific Area (From D1.2) 77
Table 26: Use case specific non functional evaluation variables for MT4: Analysis of a Specific Area (from D1.2) 79
Table 27: Evaluation scenario and relevant KPIs for the UNS1... 81
Table 28: Use case specific KPIs for UNS1: Drone Experiment... 86
Table 29: Use case specific non functional evaluation variables for UNS1: Drone Experiment (6.3 in D1.2)... 88
Table 30: Parameters that are common across all UNS use cases 89
Table 31: Evaluation scenarios and relevant KPIs for the UNS2 91
Table 32: UNS2: Localising Audio Events in Crowds KPIs (source: D6.3) 94
Table 33: UNS2 non-functional KPIs (source: D6.3)... 95
Table 34: Evaluation scenarios and relevant KPIs for the Early Adopters: The Gozo MVP (Similar to Table 9, D6.1) 97
Table 35: Use case specific KPIs for the Gozo MVP 100
Table 36: Use case specific non functional evaluation variables for The Gozo MVP (Table 6.3 in D1.2) 101
Table 37: Updated Operability and Robustness for Gozo MVP 102

Table 38: KPIs status update – Objective 1 ... 103
Table 39: KPIs status update – Objective 2 ... 107
Table 40: KPIs status update – Objective 3 ... 108
Table 41: KPIs status update – Objective 4 ... 112
Table 42: KPIs status update – Objective 5 ... 113
Table 43: KPIs status update – Impact related to the work programme ... 114
Table 44: KPIs status update – Impact on innovation capacity, competitiveness and growth ... 118
Table 45: KPIs status update – Impact on standards... 121
Table 46: Summary of the societal goals addressed by the pilots and the related measurement strategies. * indicates metrics whose evaluation would require a time span longer than the project. The tables describe the perceived impact each use case has on each societal objective. ... 122
Table 47: Mapping of the business-related use case requirements into DataBench business metrics. The table includes actual pilots' requirements as well as high-level expected goals (Table 7.1 in D1.2) .. 127
Table 48: External Advisory Board members... 131
Table 49: External Ethics Board members... 131
Table 50: MARVEL AB Feedback – M21 ... 132
Table 51: MARVEL EAB Feedback – M34... 133

List of Figures

Figure 1: The MARVEL Conceptual Architecture ... 17
Figure 2: GUI implemented on SmartViz for the use case GRN3: Traffic conditions and anomalous events... 18
Figure 3: MARVEL RP2 deployment and runtime view of the MARVEL architecture for GRN1 – Safer roads (Source: D5.6) ... 22
Figure 4: SmartViz dashboard for GRN1 ... 23
Figure 5: MARVEL RP2 deployment and runtime view of the MARVEL architecture for GRN2 – Road user behaviour (Source: D5.6) ... 30
Figure 6: SmartViz dashboard for GRN2 ... 32
Figure 7: MARVEL RP2 deployment and runtime view of the MARVEL architecture for GRN3: Traffic Conditions and Anomalous Events (Source: D5.6) ... 37
Figure 8: SmartViz dashboard for GRN3 ... 39
Figure 9: MARVEL RP2 deployment and runtime view of the MARVEL architecture for GRN4: Junction Traffic Trajectory Collection (source: D5.6)... 44
Figure 10: SmartViz dashboard for GRN4 ... 46
Figure 11: MARVEL RP2 deployment and runtime view of the MARVEL architecture for MT1 (source: D5.6)... 51
Figure 12: SmartViz dashboard for MT1... 52
Figure 13: MARVEL RP2 deployment and runtime view of the MARVEL architecture for MT2 (source: D5.6)... 60
Figure 14: SmartViz dashboard for MT2... 61
Figure 15: MARVEL RP2 deployment and runtime view of the MARVEL architecture for MT3 (source: D5.6)... 67
Figure 16: SmartViz dashboard for MT3... 68
Figure 17: MARVEL RP2 deployment and runtime view of the MARVEL architecture for MT4 (source: D5.6)... 74
Figure 18: SmartViz dashboard for MT4... 76
Figure 19: MARVEL RP2 deployment and runtime view of the MARVEL architecture for UNS1 – Drone Experiment (Source: D5.6) ... 82
Figure 20: SmartViz dashboard for UNS1 (Source: D4.6) ... 84
Figure 21: MARVEL RP2 deployment and runtime view of the MARVEL architecture for UNS2 – Localising audio events in crowds (Source: D5.6) ... 92
Figure 22: SmartViz dashboard for UNS2... 93
Figure 23: MARVEL RP2 deployment and runtime view of the MARVEL architecture for the Gozo MVP ... 98
Figure 24: SmartViz dashboard for The Gozo MVP ... 99
Figure 25: Visitors to the MARVEL Website during May 2023... 135

List of Abbreviations

AAC	Automated Audio Captioning
AB	Advisory Board
AI	Artificial Intelligence
API	Application Programming Interface
ASC	Acoustic Scene Classification
AudioAnony	Audio Anonymisation
AV	Audio-Visual
AVAD	Audio-Visual Anomaly Detection
AVCC	Audio-Visual Crowd Counting
AVDrone	Drone-based Audio-Visual data collection
AT	Audio tagging
AUC	Area Under the ROC Curve
CATFlow	Data Acquisition Framework
CCTV	Closed-Circuit Television
COVID-19	Coronavirus Disease – 2019
CPU	Central Processing Unit
D#.#	Deliverable #.#
DatAna	Data Acquisition Framework
DFB	Data Fusion Bus
DISCO	auDiovISual Crowd cOunting dataset
DL	Deep Learning
DMT	Decision-Making Toolkit
DPO	Data Protection Officer
DynHP	Compressed Models
E2F2C	Edge to Fog to Cloud
EAB	External Advisory Board
EB	Ethics Board
EdgeSec	Security Services at the edge
ELAN	EUDICO Linguistic Annotator
FedL	Framework and implementation of ML algorithms – Federated learning
FL	Federated Learning
FLOPS	Floating Point Operations Per Seconds
FPS	Frames Per Second
GA	Grant Agreement
GAN	Generative Adversarial Network
GB	Gigabyte
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GPU	Graphics Processing Unit
GPURegex	GPU Pattern Matching Framework
GUI	Graphical User Interface
H2020	Horizon 2020 Programme
HDD	Hierarchical Data Distribution
HPC	High Performance Computing
HTTP	HyperText Transfer Protocol
ICT	Information and Communication Technology
IoT	Internet of Things
IP	Internet Protocol

IT	Information Technology
KPI	Key Performance Indicator
M#	Month #
mAP	mean Average Precision
MAE	Mean Absolute Error
MB	Megabyte
MEMS	Micro Electro-Mechanical Systems
ML	Machine Learning
MQTT	Message Queuing Telemetry Transport
MVP	Minimum Viable Product
NUS	Non-Uniform Sampling
O#	Objective #
OS	Operating System
PC	Project Coordinator
R#	Release
R	Report
RAM	Random Access Memory
REST	REpresentational State Transfer
RPi	Raspberry Pi
RTSP	Real-time Streaming Protocol
S2S	Site-to-Site
S3	Simple Storage Service
sec	second
SED	Sound Event Detection
SED@Edge	Sound Event Detection at the Edge
SELD	Sound Event Localisation and Detection
SET	Sound Event Tagging
SmartViz	Advanced Visualisation Toolkit
SOTA	State-of-the-Art
T#.#	Task #.#
TAD	Text Anomaly Detection
ToC	Table of Contents
UC#	Use Case
UCSD	User Centered System Design
UI	User Interface
USB	Universal Serial Bus
VAD	Voice Activity Detection
VCC	Visual Crowd Counting
ViAD	Visual Anomaly Detection
VideoAnony	Video Anonymisation
VM	Virtual Machine
VPN	Virtual Private Network
WiFi	Wireless Fidelity
WP#	Work Package #
Y#	Year #

Executive Summary

Smart city environments generate large amounts of data from multimodal sources such as video cameras and microphones installed across the cities. Most of this data is largely underutilised and eventually deleted, mainly because of engineering and technology limitations. In an attempt to narrow the gap, MARVEL developed an experimental framework to manage the flow and processing of multimodal data over an Edge-to-Fog-to-Cloud (E2F2C) infrastructure, which would allow the end-user (e.g., researchers, engineers, managers or policy-makers) to extract useful information from the raw data via a graphical user interface (GUI). The platform has been demonstrated across three pilot cases implemented in the Municipality of Trento (Italy), the Island of Malta (Malta), and the City of Novi Sad (Serbia), the latter being more of an experimental pilot.

Throughout the project, experiments, verification, and validation steps formed part of an iterative process that drove the progress of the project. D6.2[1] reported the evaluation results up to the first prototype release (RP1), which included the implementation of five use cases in the island of Malta, the municipality of Trento and the city of Novi Sad. This document updates D6.2 with improvements carried out in RP1 and adds the evaluation of a second set of five use cases implemented during RP2.

The document first summarises the aim and objectives of the MARVEL project and provides a summary of the evaluation methodology, which was projected during the earlier stages of the project (D1.2[2]). The evaluation is carried out against a set of pre-defined parameters and KPIs, organised in a number of categories; (a) Project-related KPIs, (b) Use Case specific KPIs, (c) asset-specific KPIs, (d) MARVEL framework KPIs, (e) Business KPIs and (f) societal goals that consider the functionality and the impact of the project as a whole.

The document is written from the use case point of view and is therefore organised in separate sections per use case, whilst evaluations and parameters that cut across use cases are described collectively or per pilot. In addition, the document reports on an early adopter client of the framework, the municipality of the island of Gozo. The functional use case KPIs were aligned with technological asset-specific performance metrics and the experiments required to obtain the latter were carried out and reported in WP5, D5.5[3]. On the other hand, the non-functional KPIs are obtained through qualitative surveys and interviews as part of WP6. These were necessary to study the perceived impact of the use cases and the project.

Finally, the document reports feedback generated by the external evaluators (Advisory Board, Ethics Board, and various Info Days and events).

In summary, the perceived impact of processing data that is currently being gathered and left unused is considered highly beneficial to society by both the public, as well as the authorities responsible for various public sectors.

[1] MARVEL D6.2: Evaluation Report, 2022. https://zenodo.org/records/7296312

[2] MARVEL D1.2: MARVEL's experimental protocol, 2021. Confidential

[3] MARVEL D5.5: Technical evaluation and progress against benchmarks – final version, 2023. https://zenodo.org/records/10438311

1 Introduction

Project MARVEL considers smart cities as multimodal data pools, where data is generated from a very large number of Internet of Things (IoT) devices and sensors, such as video cameras and microphones. One of the biggest engineering challenges is the extraction of useful knowledge from this data. This presents an opportunity for developing new methodologies, techniques, and tools for information extraction and manipulation that differ from the traditional ones. MARVEL, therefore, aims to develop and harmonise techniques and technologies in the areas of Artificial Intelligence, multimodal perception, software engineering, High Performance Computing, and system architectures to process heterogeneous and distributed data in smart city environments. The project has developed an experimental prototype, the MARVEL framework, that manages the flow and processing of data over a complex Edge-to-Fog-to-Cloud (E2F2C) infrastructure. The platform receives data from a variety of IoT devices and allows the consumer (researchers, engineers, managers or policymakers) to process the data via a SmartViz implemented Graphical User Interface (GUI). The platform is showcased and evaluated over several use cases executed in the municipality of Trento (Italy), the island of Malta, and the city of Novi Sad (Serbia).

Work package WP6 was responsible to oversee the execution of the use cases and the evaluation of each, mainly from a user point of view and in so much as how the societal challenges are addressed. To do so it is also necessary to follow the progress in the technical components and the framework as a whole. The final evaluation falls under the remit of task T6.3 and this document (D6.4) reports the findings.

1.1 Purpose and scope of this document

The evaluation is carried out against the methodology defined in D1.2[4]. In summary, all Key Performance Indicators (KPIs) were monitored for each experiment in both operational and technical terms, by utilising processes that allow the systematic recording of the necessary information. This information is captured under standard operating conditions and is mainly provided by data providers and experiment owners, after consultation with the assets and framework partners. In addition, an impact analysis is carried out for each experiment and the feedback from the external evaluators is taken into consideration. The information gathered, and the analysis carried out are reported in this document (D6.4).

More specifically the objectives of this document are to:

1. Evaluate the use cases from a user point of view.
2. Summarise the evaluation of the technological assets and whether they are meeting the use case requirements.
3. Summarise the evaluation of the overall project related KPIs.
4. Evaluate to what degree societal challenges are addressed.
5. Explore the industrial potential of the results.

1.2 Intended Readership

This document is intended for a variety of readership groups, including potential end-users interested in the various use cases implemented, smart city stakeholders interested in upcoming

[4] MARVEL D1.2: MARVEL's experimental protocol, 2021. Confidential.

and cutting-edge technology and researchers from academia or industry developing or investigating similar solutions.

1.3 Contribution to WP6 and project objectives

WP6 is concerned with ensuring the implementation and evaluation of real-life societal experiments in smart city environments, or use cases, implemented in the municipality of Trento (Italy), the island of Malta, and the city of Novi Sad (Serbia).

The WP is organised across three tasks. The first task, T6.1, aligns the project activities with the experimental protocol (first developed in WP1, it includes execution timeframes, evaluation scenarios and selection of tools) to ensure the efficient execution of the use cases (i.e., the real-life societal experiments in smart city environments). Furthermore, it follows any adaption required in the execution of the use cases. The second task, T6.2, oversees the implementation of the use cases and ensures that all the steps in the data value chain are implemented for each use case. D6.1[5] and D6.3[6] defined the implementation of each use case separately and for each use case tabulates (a) all components or assets that are used in the implementation, (b) the specific datasets relevant to the use case, (c) user stories, (d) the overall evaluation parameters for all use cases and documents any revisions to use case specific functional and non-functional KPIs and asset KPIs.

Finally, T6.3 is concerned with the evaluation and impact analysis of the use cases, which also requires input from the technical evaluation and benchmarking of assets and framework, which was carried out in WP3 and WP5. This task gathers the information related to the evaluation of the assets and framework parameters and KPIs mainly from D3.5[7], D4.6[8] and D5.5[9], carries out surveys and interviews among the general public and industry experts and finally adds all feedback gathered from dissemination events and external stakeholders to evaluate the non-technical parameters and perceived impact on society and business.

1.4 Relation to other WPs and deliverables

This deliverable, D6.4, is strongly related to D1.2[4] from WP1, which planned the MARVEL Experimental protocol, D5.5[9] from WP5, which carried out the technical evaluation and benchmarking of all MARVEL components and frameworks, D3.5[7] from WP3, which reported on the AI models that support the use cases and D4.6[8] from WP4, which describes the user interface and decision making toolkit. D1.2[4] set out and described the evaluation and benchmarking strategies (section 6.3.1 in D1.2) and classifies and provides definitions for all types of KPIs (sections 6.3.3, 6.3.4 and 6.3.5, all in D1.2), which address the use case, asset, and MARVEL framework KPIs respectively. In addition, D1.2 suggests a list of parameters to study the impact on business (Table 7.1 in D1.2) and on society, i.e., how well the system addresses societal goals (Table 7.2 in D1.2). D6.4 also takes from WP7, which oversaw the organisation of events from which stakeholders' feedback was collected.

[5] MARVEL D6.1: Demonstrators execution - initial version, 2022. https://doi.org/10.5281/zenodo.6862995

[6] MARVEL D6.3: Demonstrator execution - final version, 2023. https://zenodo.org/records/8315360

[7] MARVEL D3.5 - Multimodal and privacy-aware audio-visual intelligence – final version, 2023. https://zenodo.org/records/8147164

[8] MARVEL D4.6 - MARVEL's decision-making toolkit – final version, 2023. https://zenodo.org/records/8147077

[9] MARVEL D5.5: Technical evaluation and progress against benchmarks – final version, 2023. https://zenodo.org/records/10438311

1.5 Structure of the document

This section outlined the scope of this document and how it is related to other deliverables and WPs in the project. Section 2 gives a summary of the MARVEL project, use cases, societal challenges and the overall MARVEL objectives, an overview of the evaluation and benchmarking strategy and the types of KPIs and evaluation parameters. Sections 3-13 describe the ten use cases and the Gozo Minimum Viable Product (MVP) and the respective evaluation. Section 14 reports on the project-wide KPIs and the evaluation of the business and societal goals. Section 15 describes the input from the Advisory Board (AB), the Ethics Board (EB), and the external stakeholders, whilst Section 16 summarises and concludes the report. Finally, the appendices describe methodologies used in surveys and interviews, together with a summary of responses.

2 The MARVEL project and its Evaluation

This section provides the reader with a summary of the MARVEL project and background information that may be useful to follow the rest of the document.

2.1 Summary of the MARVEL Project and Framework

This subsection is a summary of what MARVEL aims to achieve, the main objectives of MARVEL, the technical framework and its components, and the real-world demonstrators. For a more detailed treatise, the reader is referred to document D1.3[10].

The overall objective of MARVEL is to research and develop a framework that can efficiently process data collected from smart city IoT devices and sensors and produce information that is useful in addressing societal challenges. The framework is demonstrated over several use cases covering road traffic, city open spaces and large events monitoring and management to address societal challenges in sustainable mobility and climate change, and security in cities. For more details on the use cases and the societal challenges addressed, the reader is referred to sub-section 2 in each of the sections 3-12.

To achieve its overall goal, the project is based on four main pillars; (1) *Real, heterogenous, distributed Big Data in smart cities*, i.e., the real-world use case implementation, (2) *AI-based processing for multimodal (mainly video and audio) perception and situational awareness*, (3) *an Edge-to-Fog-to-Cloud (E2F2C) ubiquitous computing architecture* that allows for the distributed collection and management of data, AI model training and inference, a user interface, and the optimisation of resources for data processing, and (4) *Validation and quantitative assessment of the E2F2C and multimodal tools and methods via societal, business and industry-validated benchmarks*.

To realise the framework, it was necessary to build or modify and improve technological components and assets at every level of the E2F2C architecture. Figure 1 depicts an overview of the MARVEL conceptual architecture, which consists of thirty-two components grouped into the following seven MARVEL subsystems;

1. The **sensing and perception** subsystem includes the various sensors, mainly cameras and microphones, data capturing systems, such as the AVDrone and asset registry.
2. The **security, privacy and data protection** subsystem is responsible for preserving privacy via the anonymisation of the audio and video sources, the implementation of end-to-end communication and device security.
3. The components of the **data management and distribution** subsystem manage and optimise the data flow across the E2F2C infrastructure.
4. The **audio, visual and multimodal AI** subsystem is responsible for providing the AI models and algorithms (for example video anomaly detection and sound event detection), with the related training processes, that are necessary to implement the use cases.
5. The **optimised E2F2C processing and deployment** subsystem is responsible for the delivery of optimised AI models and includes model compression, federated learning, and GPU acceleration.

[10] MARVEL D1.3: Architecture definition for MARVEL framework, 2021. https://doi.org/10.5281/zenodo.5463897

6. The **E2F2C infrastructure** subsystem is the collection of inter-connected computational resources made available by the project partners and includes Intel NUC, RPis, PCs at the edge, workstations, and local servers at the fog and HPC infrastructure at the Cloud layers.
7. The **system outputs** consist of the **Decision-making Toolkit (DMT) and the MARVEL Data Corpus** elements. The DMT is the main interface of the MARVEL framework to the end-users (traffic managers, police and law-enforcement personnel, researchers, academics and engineers) and includes a graphical user interface (GUI) that facilitates the use of the system. The MARVEL Data Corpus provides access to labelled, anonymised training data from MARVEL pilots. Figure 2 depicts an example of the SmartViz GUI for the GRN3 use case.

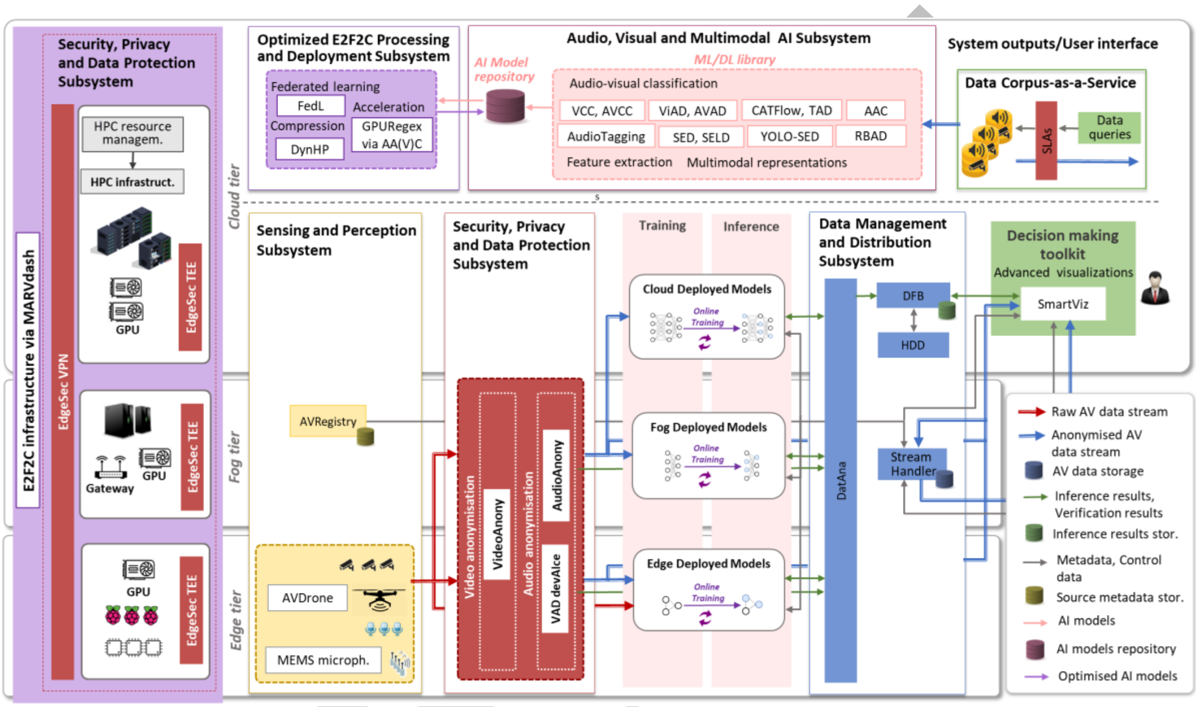


Figure 1: The MARVEL Conceptual Architecture

The MARVEL conceptual view of the architecture has been updated in RP2 to reflect some deployment consideration. Unlike the conceptual view of the architecture in RP1 (reported in D5.4), several components are excluded and/or replaced with other after lessons learned. The AV acquisition and source coder of GRNEdge was replaced by CCTV cameras, which were more robust in the outdoors condition during the summer period, whilst edge processing was implemented on a GPU device. The sensMiner and devAIce components were further developed within RP2, but they were not integrated in RP2 as there were not suitable use cases. In the end, SED@edge could not be tested in a solid way, because no microcontrollers were deployed. However, six new components were integrated in RP2: EdgeSec TEE, Automated audio captioning (AAC), Sound event localization and detection (SELD), GPURegex, and two not foreseen components, YOLO-SED and Rule-based anomaly detection (RBAD).

2.2 Types of KPIs and evaluation parameters

Due to the complex nature of the MARVEL framework, the KPIs have been organised into four categories: (a) Project-related KPIs, (b) Use Case specific KPIs, (c) Asset-specific KPIs, and (d) MARVEL Framework KPIs. Here we briefly summarise what the KPIs are about and how they have been defined. Further details are available in D1.2[4], D6.1[5] and D6.3[6].

In addition to the project-related KPIs (reported in the GA), which are associated with the technological and scientific objectives of the project as a whole, crucial for an effective assessment of the project's success are the *use case specific KPIs*. These KPIs capture the requirements of the end-users (the pilots) and are expected to evaluate whether the project solved or not the technical and societal challenges and issues that are behind the participation of the pilots in the project. For each use case, we collected technological goals as well as societal and business expectations. These have been defined via questionnaires and have been categorised into two groups: *Functional and non-Functional KPIs.*

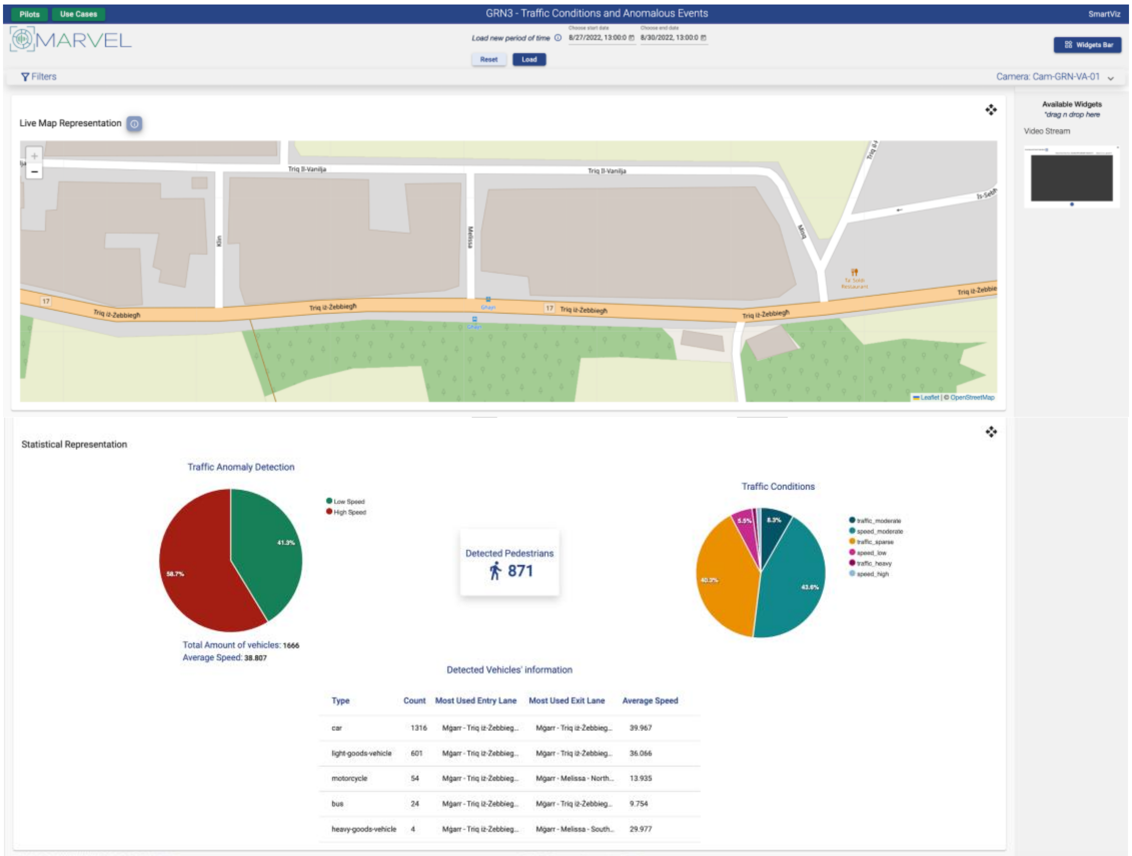


Figure 2: GUI implemented on SmartViz for the use case GRN3: Traffic conditions and anomalous events

<u>Functional KPIs</u> measure if and how the MARVEL platform solves the pilot issues. Examples are:

- GRN-KPI1: to increase the cyclists detection rate by 10% with respect to current technology.
- MT-KPI6: to reduce the reaction time in case of issues with respect to a single person monitoring all the cameras.

These objectives are in general evaluated by internal personnel of the pilots.

<u>Non-Functional KPIs</u> refer instead to other features of the solution, not directly related to the final technological goals. Examples are:

- Usability and satisfaction for the end-users.
- Acceptance and satisfaction by the public.

- Scalability, costs, economic benefits.

These non-functional KPIs are particularly challenging because they either often involve external evaluators whose engagement is not always straightforward or the evaluation or estimation of long-term impacts following a widespread rollout of the technology. In spite of the diverse nature of the pilot application scenarios, several non-functional requirements are similar across pilots.

Besides delivering end-to-end solutions for the pilot end-user, one of the goals of the project is to advance all the technological assets provided by the partners. These improvements were initially defined in the GA and finalised in D1.2[4], including for each asset:

- one or more target improvements
- evaluation metrics
- benchmarking datasets
- comparison baselines
- the related project KPIs.

Asset KPIs have been collected directly from each involved partner, following a bottom-up strategy. As the state-of-the-art rapidly evolves in most scientific fields involved in MARVEL, a periodic revision of these KPIs was carried out as part of the benchmarking process in WP5 and WP6.

From the pool of asset and use case KPIs, the evaluation parameters and the evaluation variables of the overall *MARVEL framework* were identified. Overall, five evaluation parameters: performance, usability, scalability, trustworthiness and reliability, and fifteen evaluation variables, were defined. In addition, for each evaluation variable, metrics and affected MARVEL subsystems were defined. These are elaborated in the use case evaluation sections 3-12.

2.3 Benchmarking and Evaluation Strategy

Three types of benchmarks to assess and evaluate the MARVEL system were highlighted in D1.2[4];

- **Technical benchmarks:** Assessing how the MARVEL framework and the individual components go beyond the state-of-the-art on technical excellence. This type of benchmarks is mostly related to the work carried out in the scope of WP5.
- **User experience benchmarks:** Looking at the usability of the system by the end-users and therefore related mostly to the "User interaction and decision-making" layer of the MARVEL architecture.
- **Business (model) benchmarks:** Validating the business-related aspects, such as processes, sustainability of the solutions and cost perspective. This has to do with the work carried out in exploitation (WP7) but also encompasses the assessment of some business metrics that can help to understand the industrial potential of the results.

The first two were used to assess the technical and usability aspects, while the final set of benchmarks helps to understand the business value of the MARVEL solution.

Since WP5 performed dedicated technical benchmarking for the components and the MARVEL framework as a whole, reported in deliverable D5.5[9], the strategy in WP6 related to the technical evaluation focused mainly on the overall perception of the system behaviour from the perspective of the fulfilment of the use cases, looking at aspects such as general performance, scalability, robustness, accountability, or privacy awareness, leaving more technical details of the various components to WP5 technical assessment.

As suggested in D1.2[4], usability was assessed via dedicated questionnaires or interviews with the users involved in the use cases to validate mainly the current version of the user interface in a qualitative fashion. From the business perspective, D1.2[4] suggested a set of typical metrics to measure business performance. Some of these metrics were assessed to some extent in this deliverable, especially those related to efficiency (Time Efficiency, Service Quality) that are directly related to costs (Cost Reduction, Profit and Revenues Increase). However, a more in-depth assessment of the business metrics has been carried out in the final release version of the platform in collaboration with WP7 (deliverable D7.6[11]) in relation to the business value and innovation of the MARVEL framework with regard to the pilots' use cases.

The sections related to the specific use cases in this document present a summary of some of these metrics, such as Efficiency, Operability, Usability, Robustness and Performance, along with an overview of specific KPIs related to the pilots. Evaluations with regard to the specific pilot use cases developed during the RP1 period were reported in D6.2[1], which document served as an initial assessment of the system and acted as a basis and feedback for the final assessment carried out at the end of RP2 and reported in this document.

[11] MARVEL D7.6: Final Business Model and Long-term Sustainability Report, 2023. Confidential.

3 Use case GRN1: Safer Roads

3.1 Introduction

This section gathers all the information pertaining to use-case GRN1: Safer Roads, which was implemented during reporting period RP2. The first two sections summarise the scope and description of the use case (section 3.2), and the component assets and framework configuration (Section 3.3). These sections provide a background to the use case evaluations reported in Section 3.4.

3.2 Scope and description of use case

The *Safer Roads* use case addresses the need to increase safety on urban roads for vulnerable road users (VRUs), with the aim of encouraging the uptake of active travel modes in Malta. More specifically, this use case mainly targets cycling but can also be applied to walking. Malta has witnessed a significant effort, from both the authorities and the bicycle commuting lobby, in encouraging cycling and walking, mainly through physical infrastructural changes. The use case takes this effort further and aims at adding road-side information technology to deliver safer roads. More specifically the use case aims at detecting cyclists (and pedestrians), and alerting car and other motorised-vehicle drivers of the presence of VRUs via variable message / smart signs with the hope that vehicle drivers take note, apply greater care, and concentrate on the road in such circumstances. This use case should therefore contribute towards an increase in the perceived safety on the roads and potentially encourage commuters to consider cycling/walking as an alternative mode of transportation. To determine the societal impact of this use case, surveys to gauge citizens' perceptions of safety with this device are conducted.

From a technological perspective, the use case makes use of both cameras and microphones to detect the VRUs and minimise the number of false positives. It should be noted that a high rate of false positives can potentially result in vehicle drivers losing faith in the system. Therefore the expected result is to note an improvement in the detection rate when multimodal detection is deployed (compared to single mode vision systems). In addition, the detection should occur in real-time, i.e., latency in both the detector and overall system should be very small. Two evaluation scenarios have therefore been defined for this use case and are related to two use case KPIs as reported in Table 1. These evaluation scenarios and KPIs are further discussed in Section 3.4.1.

Table 1: Evaluation scenario and relevant KPIs for the GRN1 (Source D6.3, Table 5)

Evaluation Scenario	Target	Relevant KPI
Testing the various AI models on a labelled dataset to determine the detection rate and F1 score achieved by the models	Improvement in detection rate due to multimodal detection	**GRN-KPI1: Detect Vulnerable Road Users (VRUs) at any time of day, including low-light conditions**
Observing the time taken to detect an anomaly/entity through measuring the system’s latency	The aim is to detect entities 2 seconds after the start of the event	**GRN-KPI2: low latency (time between detecting the cyclist and informing the road users)**

3.3 Summary of Technical Setup and Decision Making Toolkit

The Safer Roads use case samples vision (video camera) and audio (microphone) data to detect bicycles (and pedestrians) and disambiguates cycles from internal combustion engine (ICE) motorcycles that may be picked up by the vision detector. An IP camera with integrated microphone is used in the pilot implementation. Two independent systems are used to process the data, one based on the Jetson Apollo Dev Kit edge device (NVIDIA Jetson Xavier NX) and the other on the GPU-enabled edge PC. The former is responsible for executing the VRU detector model, while the latter anonymises the video such that it can be displayed on the SmartViz DMT[8]. The YOLO-SED AI model is used in the use case, where the YOLO part processes the vision data, the SED part processes the audio data and a logic component combines the multi-modal input. The YOLO component is based on the CATFlow fine-tuned model, whilst the SED component is trained on the GRN-AV-traffic-entity dataset described in D2.1[12]. For this dataset, audio-visual snippets were obtained from various locations around Malta and from the GRN static cameras. A small-scale multimodal dataset was used for testing the YOLO-SED component. This dataset contains examples of pedestrians, bicycles, and motorcycles labelled with bounding boxes and sound events. All video streamers are anonymised using VideoAnony. The version used blurs number plates and faces resulting in AV streams that contain no identifiable information.

The system architecture is depicted in Figure 3. During operation, the system illuminates a traffic sign (replaced by a set of LEDs in the project) when it detects a VRU thus alerting vehicle drivers of the VRU. The detections from YOLO-SED are also transmitted to the MARVEL platform for visualisation in SmartViz, enabling remote monitoring as well as data collection for further analysis. The Real-time alerts are sent to SmartViz (the dashboard), Figure 4, through Kafka messages via the Data Fusion Bus (DFB) component. These alerts can be visualised in the *Alerts* widget, providing users with immediate information. Additionally, the *Details* widget displays the last detected events in a textual format, utilising historical data accessed from an Elasticsearch database. This database is continuously updated by the audio, visual, and multimodal AI subsystem components.

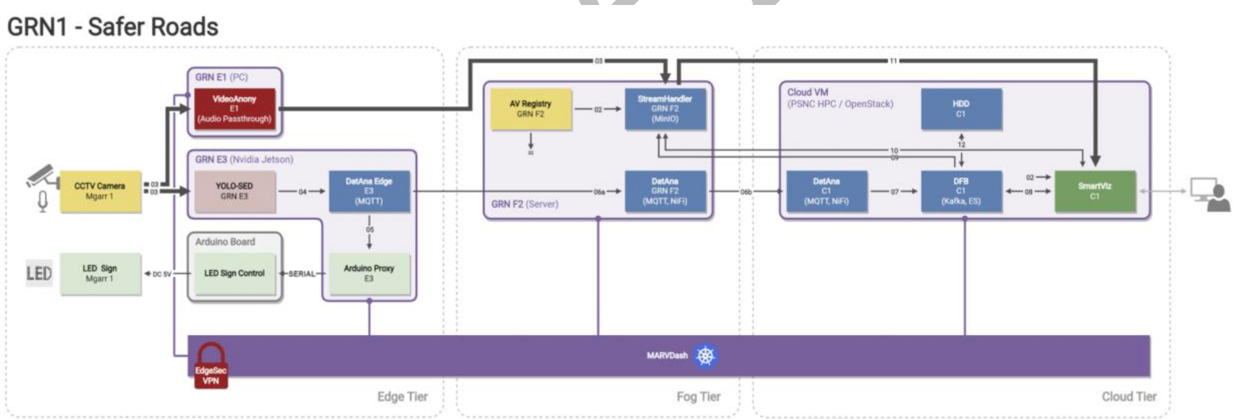


Figure 3: MARVEL RP2 deployment and runtime view of the MARVEL architecture for GRN1 – Safer roads (Source: D5.6)

Within SmartViz, users have the capability to select an event displayed in the widget and play the corresponding video snippet. The StreamHandler component facilitates this functionality by retrieving the relevant video from the camera stream within a specific time period. The segmented stream generates a URL containing the video for the selected event, which is played in SmartViz using the "Anomaly and Event Detection player" widget.

During video playback, users have the option to validate the inference result by marking them as accurate or not. This validation process contributes to the continuous improvement of the

[12] MARVEL D2.1: Collection and Analysis of Experimental Data, 2021. https://zenodo.org/records/7543662

system and the training of AI models. The inference result verification is sent through Kafka messages from SmartViz to DFB, which updates the status of the corresponding event and stores the information in an Elasticsearch index accessed by the Data Corpus.

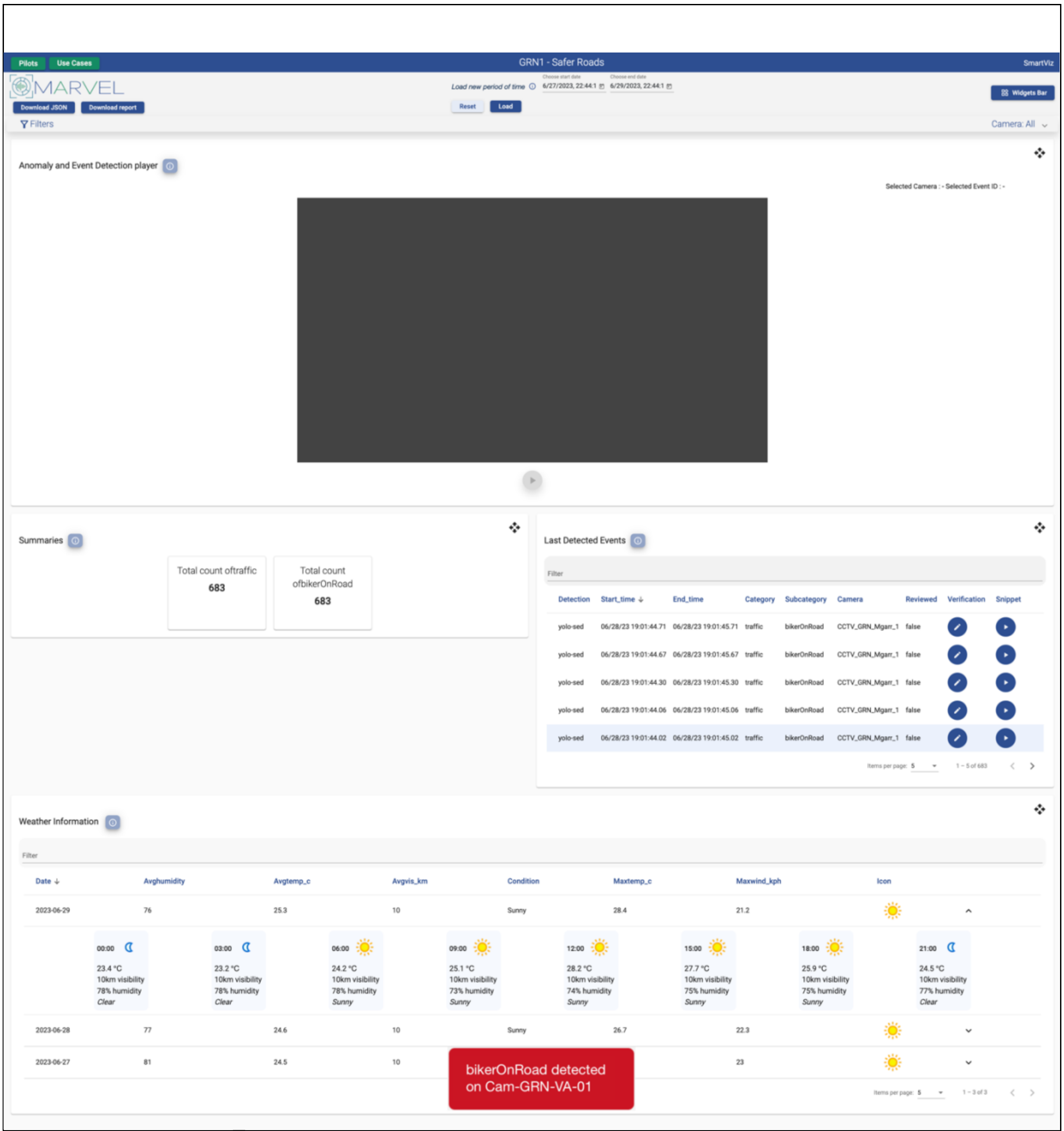


Figure 4: SmartViz dashboard for GRN1

In addition to widget arrangement and resizing, the dashboard view offers functionality for users to download the visualised data in JSON format. This feature allows for easy retrieval and utilisation of the data for further analysis and reference purposes. Furthermore, users have the option to save the entire dashboard as a PDF file, enabling offline access and sharing based on their individual preferences.

Finally, the *Weather Information* widget in this use case provides users with a representation of weather-related data. It allows users to view and explore weather information for a selected time period. By visualising weather variables such as visibility, humidity, temperature, and overall weather conditions, users can gain insights and uncover potential correlations between detections of events, anomalies and weather data that may influence the traffic data.

3.4 KPI assessment

This section tabulates the results obtained from the evaluations of the use case parameters and KPIs. These tables are obtained from D1.2[4], D6.1[5], and D6.3[6] and the evaluations are specific to the use case. Section 14 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs, Business KPIs and societal goals benchmarks, all of which cut across use cases and pilots.

3.4.1 Use case specific KPIs

The detailed results for the functional KPIs are tabulated in Table 2. The system achieves the target of the use case if the cyclist detection rate is high and the delay between detection and notification of vehicle drivers is low, typically less than two seconds. The detection rate achieved is 88.26% and the system delay is 1.15 seconds. The AI model implemented has therefore improved the detection rate by a factor of 5.4, from 16.34% (for the general YOLO model) and from 56.39% using CATFlow weights to 88.26% using object suppression rules. In practice, the system detects almost 90% of cyclists and vehicle drivers are notified within almost 1 second. The system performance is therefore highly satisfactory for the use case to be useful in practice.

Table 2: GRN1 functional KPIs (source: D6.3[6])

<table>
<tr><th colspan="5">GRN1: Safer Roads: functional KPIs</th></tr>
<tr><th>KPI</th><th>Metric</th><th>Baseline</th><th>Expected result/ Improvement</th><th>Evaluators</th></tr>
<tr><td rowspan="2">GRN-KPI1: Detect Vulnerable Road Users (VRUs) at any time of day, including during low-light conditions</td><td>Detection rate/F1</td><td>Labelled dataset</td><td>Improvement in detection rate due to multimodal detection</td><td>GRN developers</td></tr>
<tr><td colspan="4">Evaluation and Comments

The detailed technical evaluation methodology for the YOLO-SED model is documented in Section 4.2.4 of D3.5[7]. The evaluation here is restricted to the application of the AI model to the use case.

The YOLO-SED model consisted of an object detector, a motorcycle sound detector, various fusion rules and an object detection suppression mechanism, and the evaluation is performed on a dataset of eight audio-visual sequences, recorded from the camera of the use case. The highest accuracy achieved with the final model is 88.26% when the vision part is the YOLO model fine-tuned with images sourced from various sources and the object detection suppression rule-based mechanism is deployed. The system therefore improved the detection rate from 16.34% (for the general YOLO model) to 88.26%.

The object detection suppression rules improve the performance of the general model by a factor of 2.64 to 59.4% whether SED is used or not. Similarly, the fine-tuned model improves the performance of the general model to 56.39% with or without SED. Combining SED with the best model does not improve the performance and the accuracy is 86.4%. Error analysis concluded that most errors come from unrelated objects and most of these are fixed with the object suppression mechanism.</td></tr>
<tr><td>GRN-KPI2: low latency (time between detecting the cyclist and informing the road users)</td><td>Time in seconds</td><td>No baseline</td><td>2 seconds</td><td>GRN developers and managers</td></tr>
</table>

	Evaluation and Comments *The technical evaluation methodology for determining the system latency is documented in Section 4.4.1 of D5.5*[3]. The overall latency is dependent on the various components operating at the edge, namely the YOLO-SED model, DatAna Edge, and Arduino sub-system. The overall latency achieved is 1.15 seconds, which is less than the target of 2 seconds. The YOLO-SED model handles 1.98FPS on the edge device, which when installed in the system introduces a latency of 580ms. The other components contribute 333ms (DataAna edge) and 233ms (Arduino sub-system) to the overall latency. The camera latency is insignificant.

3.4.2 Use case specific non-functional KPIs

The detailed evaluations for the GRN1 non-functional KPIs, tabulated in Table 3, are mostly obtained from surveys and interviews (Appendix section 17.1).

Table 3: GRN1 non-functional KPIs (source: D6.3)

GRN1: Safer Roads: Non-functional KPIs			
Evaluation variable	**How to measure**	**Internal evaluators**	**External Evaluators**
End-user Experience: Perceived Safer cycling and walking. System welcomed by cyclists and pedestrians	Survey	GRN managers	Cyclists and pedestrians
	Evaluation and Comments The evaluation is carried out from surveys S1, S5 (section 17.1.1), and interviews (section 17.1.7). Details of survey results are given in section 17.1.1. From S5, 41.4% and 35.7% of respondents from the general public agree and strongly agree that if car drivers actually take greater care, following alerts from the system, cyclists and pedestrians will indeed feel safer on the road. In addition, 37.1% and 21.4% agree and strongly agree that this would convince more commuters to cycle. Furthermore, 45.7% and 32.9% agree and strongly agree that they would be inclined to consider cycling. This corroborates previous research on active commuting[13]. From S1, 60% of expert respondents think the system is an effective tool to educate drivers and 40% think that it must be coupled with law enforcement. All agree that storing all detections of VRUs together with weather data is useful in planning for active travel. There is strong interest, 80-100%, in using the system to support variable speed limits, which can smoothen a transition to active travel. From interviews: Such a system can help pedestrians in areas where there are no pavements. A strong feature is the fact that near misses can also be collected with such a system, whereas accident statistics provide data only when an accident occurs. The system is also useful during sparse traffic when cars overspeed most, whereas in heavy traffic drivers are more cautious.		
Efficacy: Car drivers welcome the system and agree that it will help them be more careful.	Through a survey	GRN managers	Car drivers
	Evaluation and Comments The evaluation is carried out from surveys S1, S5 (section 17.1.1), and interviews (section 17.1.7).		

[13] S. Maas and M. Attard, "Shared Mobility Services in Malta: User Needs and Perceptions", in Sustainable Mobility for Island Destinations, 2022: DOI:10.1007/978-3-030-73715-3_5

<table>
<tr><td></td><td colspan="3"> From S5: In general, car drivers agree that such a system should either help them drive more carefully when observing the smart signs or would help them achieve more careful driving in the longer term. S1 and interview: The system was compared or likened to over-speeding smart signs already installed on the roads and the new Intelligent Speed Assistance (ISA) installed in new vehicles. </td></tr>
<tr><td rowspan="2">Scalability
Cost to add new devices/junctions</td><td>Cost estimate</td><td>GRN managers</td><td>Third party service Providers</td></tr>
<tr><td colspan="3"> Evaluation and Comments Scalability here refers to the cost of installing the system at a new location. The cost varies according to mainly the number of cameras per location and the active sign deployed (maintenance is not included in table). The cost needs to be compared to the cost, including opportunity cost, of physical infrastructure.
<table>
<tr><td>Hardware</td><td>Cost (in euros)</td><td>Streams Processed</td><td>Typical Cost (in euros)</td></tr>
<tr><td>Jetson Xavier</td><td>800</td><td>8</td><td>100-800</td></tr>
<tr><td>Fog Sever</td><td>2000</td><td>3</td><td>666</td></tr>
<tr><td>IP camera</td><td>500</td><td>1-3</td><td>500-1500</td></tr>
<tr><td>Variable Message sign</td><td>200-2000</td><td>1-3</td><td>200-2000</td></tr>
<tr><td colspan="3">Ancillary equipment</td><td>500</td></tr>
<tr><td colspan="3">Total Cost for one camera + edge processing</td><td>2000-3800</td></tr>
</table></td></tr>
<tr><td rowspan="2">Usability</td><td>Survey</td><td>GRN Managers</td><td>Transport Experts</td></tr>
<tr><td colspan="3"> Evaluation and Comments The evaluation is carried out from surveys S1 (section 17.1.1) and interviews (section 17.1.7). In general, all respondents agree that the SmartViz toolkit (a) effectively supports the user reach the end goal of the use case, (b) the summaries widget is useful, (c) the availability of a video snippet helps them understand better the data and the user can contribute to the system by validating or otherwise the detections. 20% of the respondents are not sure of the usefulness of the Alert widget, possibly because it may be only useful in live troubleshooting the system. In addition, GRN engineers think that the alert widget should be useful during system installation and testing. </td></tr>
</table>

3.4.3 Additional parameters common across all GRN use cases from D6.1/D6.2

A subset of parameters suggested in D6.1 are common for all use-cases. Table 4 tabulates these parameters and their assessment, which have been updated since D6.2.

Table 4: Parameters that are common across GRN use cases

<table>
<tr><th>Parameter</th><th>How to measure</th><th>Target to be achieved</th></tr>
<tr><td>Operability and Robustness
Operability is related to the ability of the components to keep functioning together. This can be quantified by recording downtime for any of the components (assets) along</td><td colspan="2"> Evaluation and Comments The primary AV raw data source in the GRN pilot consists of three IP cameras which have no ability to process data at the Edge and limited to transmitting audio and video via an IP connection, typically a 4G wireless connection. Two PC computers are installed at two locations and act as edge devices. These devices are used to either execute AI models or carry out data anonymisation at the edge. The third edge device is the Jetson Apollo </td></tr>
</table>

the system pipeline as a percentage of total time. *Robustness is related to how robust the system components are during the period of operation. This can be quantified by observing whether performance is sustained in various weather conditions (Power source, sensor, and CPU board operation).*	Dev Kit, which is a development kit for the NVIDIA Jetson Xavier NX. This device is connected to an Arduino nano device which controls an LED board for use in GRN1 use case. Further information about the GRN edge infrastructure can be found in D5.6[14]. The GRN edge infrastructure was robust and operable for most of the time, with the exception of the following events. One of the cameras, although rated for outside use, sustained water damage that resulted in a blurred image and the device was not available for a few days whilst being repaired. The damage could have been caused by wet gale force winds that are common at the location, or alternatively, the product was faulty. During the summer months of 2023 one of the edge PCs had to be switched off and moved to a room with air conditioning to prevent overheating because of the extreme ambient temperatures. This resulted in a downtime of a couple of weeks until the new location was set up. During this time, a second camera was malfunctioning, for unknown reasons, probably due to overheating. The camera was replaced within a week. Following this experience, edge devices performing computations are better installed in a temperature-controlled environment to maximise robustness. The GRN fog infrastructure is hosted in a temperature-controlled server room. Hence the fog servers did not experience any downtime due to extreme temperatures. Nevertheless, the server sustained 8 hours of downtime, due to loss of electrical power in the building. The 4G wireless internet connections occasionally experienced short-term interruptions, resulting in the loss of a few frames, whilst wired connections were much more stable and reliable.
Accountability *Related to the system being able to explain results or decisions.*	**Evaluation and Comments** With the exception of the rule-based model (RBAD), the AI models deployed do not explicitly explain the reasoning process behind the prediction. However, this was not a problem in the use cases since in case of the anomaly and event detection models, the system stores a video snippet of the event such that the human (via the SmartViz DMT) can then validate or otherwise. This approach has been confirmed by experts in transport as not only satisfactory but also desirable since the final decision is anyway taken by the human.
Transparency *Related to the description of the processes or algorithms that are used to generate system output*	**Evaluation and Comments** The processes and algorithms are explained in detail in deliverables D3.5[7] (AI Models), D4.6[8] (Decision Making Toolkit) and D5.6[14] (Integrated Framework).

[14] MARVEL D5.6: MARVEL Integrated framework – final version, 2023. https://doi.org/10.5281/zenodo.8315386

Privacy awareness *Related to the provision of adequate governance mechanisms that ensure privacy in the use of data.*	**Evaluation and Comments** Three methods have been adopted in the GRN use cases; (a) AV data is processed at the edge and non-AV data is forwarded to the next node; (b) AV data is anonymised prior to being stored at the fog or cloud layer. Details on the anonymisation techniques are given in D3.3[15] and D3.5[7] Benchmarking results for VideoAnony yielded the following results: the AI model anonymises videos with a precision of 0.797, Recall of 0.492, 0.614 mAP@0.5 and 0.29mAP@0.95; and for car plate detection using Yolov5X6 (pre-trained on COCO) and fined-tuned using the GRN annotated data, using images of 1280 by 1280 yielded: - Average precision: 0.82, - Average recall: 0.67. GRN engineers manually sampled a number of videos in which most face and number plates were detected. Error analysis revealed that non-detected faces or number plates were located in the background of the image, i.e., small and difficult to resolve. In addition, from surveys S4 and S5 (section 17.1.1): 50% of experts think that privacy is well catered for in the system, but 50% are more cautious and took a neutral stance; 61.4% of the general public are satisfied with having the data anonymised, 25.7% took a neutral stance and 12.9% are not.

[15] MARVEL D3.3: E2F2C Privacy preservation mechanisms, 2023: https://zenodo.org/records/7541694

4 Use case GRN2: Road User Behaviour

4.1 Introduction

This section gathers all the information pertaining to use case GRN2: Road User Behaviour, which was implemented during RP2. The first two sections summarise the scope and description of the use case (Section 4.2), and the component assets and framework configuration (Section 4.3). These sections provide a background to the use case evaluations reported in Section 4.4.

4.2 Scope and description of use case

This use case addresses the need to monitor the behaviour of road users at junctions and on roads. This use case demonstrates tools that are potentially useful in education and law enforcement campaigns targeting responsible driving, cycling, and other uses of the roads. Malta has experienced fast changes in the transport landscape, to which human response often lags technical progress. Educational campaigns are one way to fill in the gap and have been shown to be effective in the past. This use case involves the classification of actions into a spectrum of examples demonstrating inappropriate behaviour. This use case will not be implementing the latter campaigns or policies; however, it provides a demonstration. Surveys will be used to study how this tool will be able to help local authorities. As with educational campaigns, this use case implementation could equally be used to plan and study the impact of law enforcement.

Examples of actions include the way pedestrians cross over the intended crossings, whether cyclists use the infrastructure as intended and whether car drivers stop in the delineated zone at junctions. The system will be able to count the number of times inappropriate behaviour is detected before and after the execution of education campaigns or policy changes.

The data is then summarised by the DMT, following which one can compare the data before and after an educational campaign or assess the impact of a recent installation of a traffic calming measure, whether it is law enforcement or a passive measure. The impact of this use case will be measured through interviews with authorities or other third parties to determine if this tool will help them evaluate their campaigns and policies.

The usefulness of this use case relies on the ability of the system to automatically capture improper behaviour. For this use case, a realistic expected result was set to detect at least four improper types of behaviours. Table 5 summarises the evaluation scenario defined for this use case and the relevant KPI originally defined in D1.2[4].

Table 5: Evaluation scenarios and relevant KPIs for the GRN2 (source: D6.3)

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario includes testing the various AI models to confirm the detection of at least four improper behaviours	Four types of improper behaviours detected	**GRN-KPI3: Automatically detect and label or quantify actions that determine driver behaviour**

4.3 Summary of Technical Setup and Decision Making Toolkit

GRN2 focuses on monitoring the behaviour of road users at a junction to support law enforcement and educational campaigns in promoting responsible driving, cycling, and walking. The use of AI models through the MARVEL platform enables automatic detection of common inappropriate behaviours, providing counts and comparisons of occurrences over time. The GRN sensory components provided for this use case consist of three IP cameras streaming audio and video to the framework. The framework, Figure 5, implemented in this use case is very flexible and can handle AI model processing at both the edge and the fog layer.

The TAD and RBAD AI components are used to detect anomalies in the video stream. Both components ingest output from CATFlow detailing the object detected and tracked from the video and then use statistical and rule-based anomaly detection methods to flag improper behaviour. SED is also used to detect the usage of horns from the audio stream as horn usage normally indicates improper behaviour on the road. GRN provided a dataset for the testing of the TAD component in this use case. This dataset was collected from the Mgarr location. The annotators were instructed to go through a series of five-minute videos and find cases where vehicles underwent unusual behaviours, given the context, for example, a vehicle making a U-turn or stopping at the roadside, or just taking too long to proceed at a stop sign. The annotators also labelled examples of vehicles that do not occur too often on this road, such as bicycles and scooters. The datasets consist of one hundred and thirty-seven examples of labelled anomalous events which amount to more than eleven hours of data. In addition, a GRN-provided small-scale dataset was used for fine-tuning and testing the SED horn detection model.

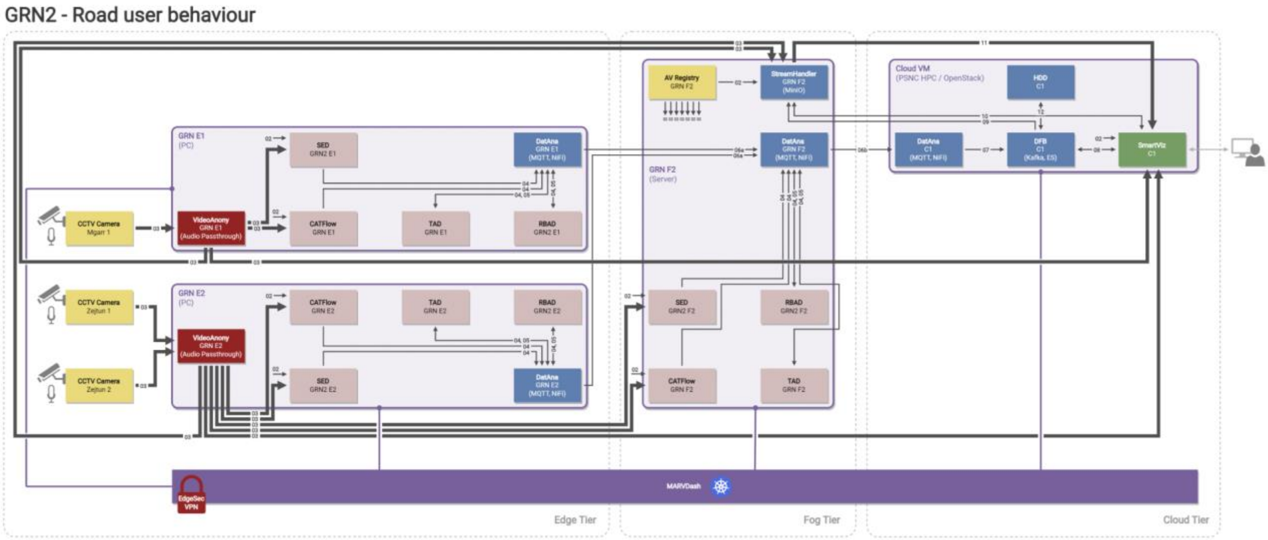


Figure 5: MARVEL RP2 deployment and runtime view of the MARVEL architecture for GRN2 – Road user behaviour (Source: D5.6)

The SmartViz platform, an example of which is shown in Figure 6, displays the detected behaviours and allows users to select two time periods for comparison. The Summaries widget presents the total number of detected anomalies, utilising outputs from SED, RBAD, and only the anomalous events of TAD. The Statistics widget visualises bar charts indicating the total number of detections per anomalous event.

The Temporal Representation widget presents the information of detections by all AI components used in this use case in a temporal format. The Details widget visualises the detections in a textual format, with CATflow and TAD displayed in the "Traffic Events Detection" table, and RBAD and SED displayed in the "Sound Events and Anomalies Detection" table.

All data feeding the widgets in this use case are historical and accessed through an Elasticsearch database, which is constantly updated by the audio, visual, and multimodal AI subsystem components.

Within the audio and video player functionality in SmartViz, users can select an event and play the corresponding snippet. The StreamHandler component facilitates this feature by retrieving the stream for the relevant time period of the detection. The segmented stream generates a URL containing the snippet for the selected event, which is played in the Audio or Video player widget within SmartViz.

Users have the ability to validate the inference result by marking them as accurate or not. The inference result verification is sent through Kafka messages from SmartViz to the Data Fusion Bus (DFB), which then updates the status of the corresponding event and stores the information in an Elasticsearch index accessed by the Data Corpus.

In addition to rearranging and resizing widgets, the dashboard view offers functionality for users to download the visualised data in JSON format. Furthermore, users can save the entire dashboard as a PDF file, allowing for offline access and sharing.

Finally, the Weather Information widget in this use case provides users with a representation of weather-related data, allowing them to view and explore weather information for a selected time period.

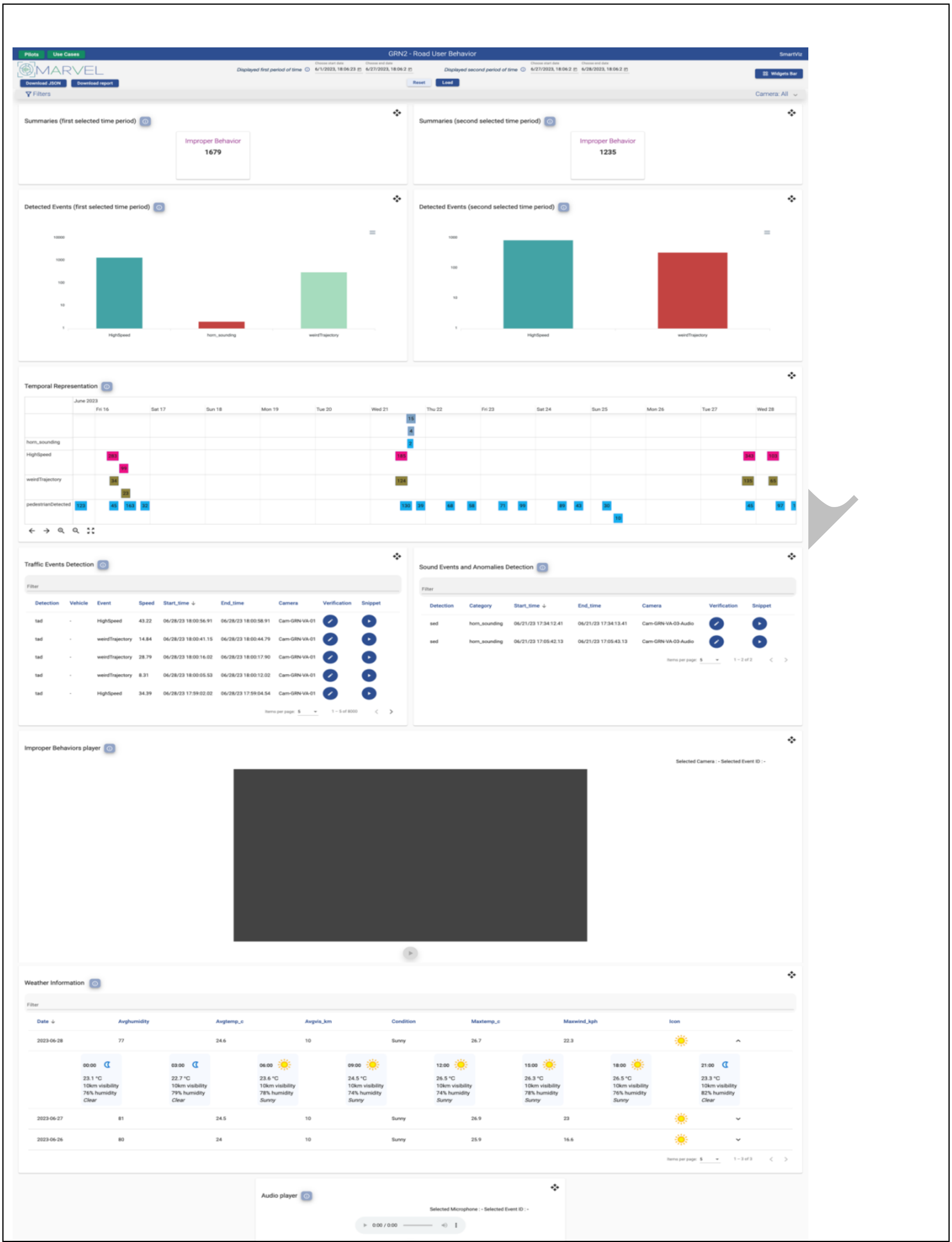


Figure 6: SmartViz dashboard for GRN2

4.4 KPI assessment

This section tabulates the results obtained from the evaluations of the use case parameters and KPIs. These tables are obtained from D1.2[2], D6.1[5] and D6.3[6] and the evaluations are specific to the use case. Section 14 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs, Business KPIs and societal goals benchmarks.

4.4.1 Use case specific KPIs

The detailed results for the functional KPIs are tabulated in Table 6. The system achieves the target of the use case (i.e., to detect four improper behaviours). Three types of models have been evaluated; TAD and RBAD that take CATFLow output as input and SED, which makes use of the audio stream to detect vehicle horn sounds. Some behaviours are better detected by TAD, whereas others are better detected by RBAD. Some examples of successful detections include U-turns (TAD), double parking (RBAD), bicycle on footpath (RBAD) and anomalous high speeds (TAD). In addition, the system detects almost 50% of horn sounds, which can add to counts of improper behaviour after a human validates the events. As is evident, the horn detector requires further development. Overall, the model performance is therefore satisfactory and is useful for the use case.

Table 6: GRN RP2 functional KPIs (Source: D6.3 and D1.2)

<table>
<tr><th colspan="5">GRN2: Road User Behaviour: functional KPIs</th></tr>
<tr><th>KPI</th><th>Metric</th><th>Baseline</th><th>Expected result/ Improvement</th><th>Evaluators</th></tr>
<tr><td rowspan="2">GRN-KPI3: Automatically detect and label or quantify actions that determine driver behaviour</td><td>Human evaluation on the basis of accuracy/ precision/ recall</td><td>No baseline</td><td>Four improper behaviours detected</td><td>GRN developers and managers</td></tr>
<tr><td colspan="4">Evaluation and Comments

The correct detection of improper behaviours depends on the specific model deployed and as such there is no one model that performs well for all behaviours. In this use case, the models used in detecting improper behaviours are TAD, RBAD, and SED. The best four successful detections are U-turns (TAD), double parking (RBAD), bicycle on footpath (RBAD) and anomalous high speeds (TAD). Below are the evaluations for each method.

TAD (CATFlow detections are used as inputs)

The detailed technical evaluation methodology for the TAD model is documented in Section 8.1.5 in D5.5[3].

The CATFlow speed estimation is used to detect anomalous speeds. The precision per vehicle type when detecting anomalous high-speed vehicles is: Car: 97%, Light Good Vehicle: 82%, Motorcycle: 99%, Bus: 100%, Heavy goods vehicle: 100%, and the precision per vehicle type when detecting anomalous low speed vehicles is: Car: 83%; Light Goods Vehicle: 75%; Motorcycle: 56%; Bus: 73%; Heavy Goods Vehicle: 76%. It is therefore concluded that anomalous high speeds can be detected with high precision.

The CATFlow trajectory estimation is used as an input to detect trajectory-based anomalous behaviour. The True Positive Rate per event is: vehicle crossing centre line - 5.1%; overtaking - 22.2%; significant traffic - 100.0 %; bicycle - 50.0 %; parking - 7.1%; obstruction - 57.9 %; wrong lane - 10.5%; U-turn - 95.2 %, and the overall False Positive Rate is 1.12% (20% reduction in false detections when compared to the available one)

It can be concluded that the TAD can successfully detect anomalous events characterised by rather slow-moving vehicles such as significant traffic, obstructions and U-turns. Anomalies that involve either no movement or small lateral movement, like parking and overtaking are detected less successfully. Changing the position of the camera to an over-head view may help in improving the latter detections.

Rule Based anomaly detector (anomalies are considered inappropriate behaviour)</td></tr>
</table>

	The technical description for the RBAD model is documented in Section 3.5.1, D3.5[7] The evaluation here is restricted to the application of the AI model to the use case. The RBAD model, which is a logic-based anomaly detector, employs a predefined ruleset to detect very specific anomalies, based on the input messages coming from CATFlow. The messages contain information about the objects detected in the video frame, including their location and time of detection. Based on this information, combined with the predefined rules, RBAD creates specific anomaly alert. RBAD has been set to detect pedestrians jaywalking, buses not arriving on schedule, heavy-weight vehicles present on residential roads, and cyclists using the pavement intended for pedestrians. For all anomalous situations, a human expert creates the required rules. The precision of RBAD depends on the accuracy of the CATFlow detections as well as on the quality of the rules provided by human experts. Detecting very specific behaviours with RBAD results in a high precision rate and often compliments the TAD detector. SED (Horns are considered as a sign of improper behaviour) *The technical evaluation methodology for the SED model is documented in Section 8.4.6 in D5.5[3].* The audible sound of vehicle horn events are considered as indicating some improper behaviour and therefore such detections point to time periods where events can be further observed and studied on the video stream. SED Evaluation results for GRN use case are: F1 score of 47.62% compared to 33.9% with baseline, resulting in a relative increase of 40%. Evaluation has been carried out with macro-averaged metrics to score under-represented classes. As is evident, the horn detector requires further development and improvement.

4.4.2 Use case specific non-functional KPIs

The detailed evaluations for the GRN1 non-functional KPIs, tabulated in Table 7, are mostly obtained from surveys and interviews (Appendix Section 17.1). Table 7 shows the new updated non-functional KPIs for GRN2. In general, experts agreed that the comparative tool, together with the standard widgets, is a useful addition to the framework. Furthermore, the behaviour detection models need to be specific and accurate and this may require customisation of the models.

Table 7: GRN2 non-functional KPIs

GRN2: Road User Behaviour: non-functional KPIs			
Evaluation variable	**How to measure**	**Internal evaluators**	**External Evaluators**
Potential end-user	Survey	GRN Managers	Transport authorities Road users
	Evaluation and Comments The evaluation is carried out from survey S2 (section 17.1.1) and interviews (section 17.1.7). Details of survey results are given in section 17.1.1. GRN Evaluators:_The GRN2 DMT simplifies the comparative analysis and visualisation task, which now requires office level skills. Some end users will find this tool useful. Survey S5: 37.1% and 22.9% of respondents from the general public agree and strongly agree that targeted educational campaigns can be effective at reducing certain types of improper behaviour on the road. More than 80% of respondents think that society benefits most from detecting the following (a) Vehicle drivers ignoring "stop" signs, (b) Vehicles not giving way to pedestrians at zebra crossings,		

	and (c) Double parking. 37.1% and 42.9% of respondents agree and strongly agree that such a system should also be used in law enforcement. Survey S2: 75% and 25% of respondents agree and strongly agree that educational campaigns and law enforcement go together, i.e., for maximising efficacy. From the interviews: The methods developed are in general interesting for use in law enforcement. The methods do not need to be perfect, but a certain level of accuracy is desired. Short and Long term and seasonal analysis is also useful in law enforcement. For example, certain trajectories can lead to certain accidents and this type of analysis will show whether it is the user who is at fault or the road designer. Educational campaigns are useful to a certain extent, for example, when laws change and in promoting road etiquette since many accidents are due to a series lack of road etiquette. In terms of usability, from survey S2: The evaluators agree that the system is straightforward and intuitive to use. All agree or strongly agree that the “comparison” feature is very useful in performing the required tasks. 25% of respondents do not find it convenient to download the data to a file and use third-party analysis tools and prefer a comprehensive dashboard as the one in GRN2. All agree or strongly agree that the timeline feature synchronised to all other widgets enhances the ease of data exploration. Finally, the evaluators found the customisation feature offers flexibility.		
Efficacy	System evaluated by an NGO/public authority to determine the efficacy of the system when compared to what they use now to judge the success of a campaign	No baseline	Effectiveness verified by road experts in terms of the value it adds to a potential education campaign
	Evaluation and Comments From survey S2: Three external evaluators had been involved in the design and implementation of educational campaigns in transport. However, none of them have used a tool similar to the one presented in GRN2. In general, they usually find it difficult to review the outcome of the campaign in terms of impact on behavioural change and agreed or strongly agreed that the MARVEL framework is definitely useful for analysing the impact. They also agree that the same framework can be used to a certain extent during the design of the campaign and the framework should also be effective in law enforcement. From interviews: Educational campaigns need to reflect the real data. So collecting long-term data to be used in the design of educational campaigns is crucial. Especially so during the design of such campaigns.		

4.4.3 Additional parameters common across all GRN use cases from D6.1/D6.2

Table 4, in Section 3.4.3, tabulates the parameters that are common for all GRN use cases.

5 Use case GRN3: Traffic Conditions and Anomalous Events

5.1 Introduction

This section gathers all the information pertaining to use case GRN3: Traffic Conditions and Anomalous Events, which was implemented during RP1, and following feedback updated during reporting period RP2. The first two sections summarise the scope and description of the use case (Section 5.2), and the component assets and framework configuration (Section 5.3). These sections provide a background to the use case evaluations reported in Section 5.4.

5.2 Scope and description of use case

The intended application for this use case is the monitoring of traffic conditions and the detection of anomalous events, for example: traffic jams, accidents, cars stuck and obstructing a junction, very slow vehicles and service vehicles parked on the side or obstructing a carriageway. In general, this output would find application in, for example, systems intended to inform drivers near the detected anomaly or to infer possible issues in adjacent areas and inform drivers of obstacles ahead. In addition, the detection of anomalous events can be used to alert personnel stationed at traffic management control rooms, who can then interpret the data and take any necessary action. This latter application could potentially improve the efficiency of operations.

The accurate detection and detection time are two crucial metrics to evaluate this use case. The former is related to the AI model accuracy, whilst the latter is dependent on the complexity of the AI model and the system specifications. Whenever an anomaly occurs, the anomaly is flagged in the traffic control room such that actions can be taken. In particular, a video clip of the anomaly is saved and made available to the traffic control personnel, thus gaining visual insight into the anomaly, which can help in determining the right course of action. This feature is essential for the evaluation of both metrics since the traffic personnel would be able to validate both the anomaly detected as well as the traffic conditions.

The users of this system are intended to be traffic managers who can pass on information to other authorities to react to any disruptive event, for example, a traffic incident. The system would partially replace manual systems, where personnel would typically observe multiple video screens to pick up anomalies. This use case could potentially reduce the time taken to respond to disruptive events and therefore ease traffic congestion, which is one of the societal challenges.

Table 8: Evaluation scenario and relevant KPIs for the GRN3: Traffic Conditions and Anomalous Events (Table 6, D6.1)

Evaluation Scenario	Target	Relevant KPI
Testing the various AI models on a labelled dataset to determine the detection rate and F1 score achieved by the models.	The aim is to obtain a 70% detection of anomalous events.	**GRN-KPI4: Correctly detect various anomalous events on the road**
Observing the time taken to detect an anomaly through measuring the system's latency.	The aim is to detect anomalies 2 minutes after the start of the event.	**GNR-KPI5: Detection time**

5.3 Summary of Technical Setup and Decision Making Toolkit

Figure 7 depicts the framework configuration for the GRN3 use case. Three cameras are deployed in the implementation of this use case, one at Mgarr and two at the Zejtun location. The three cameras deliver continuous AV streams with both modalities present (audio, video). The AV stream from the Mgarr camera is first transmitted to the GRN edge PC (GRN E1) where it is consumed by the VideoAnony component. One AV stream from the Zejtun Cameras is anonymised at the edge (GRN E2) whilst the second one is anonymised at the fog layer (GRN F2), via the EdgeSec component. Each of the three VideoAnony instances produces an RTSP stream that is consumed by several components further up in the pipeline as follows; 1) the AI components (CATFlow+TAD, AVAD, and AT), to produce inference results; 2) StreamHandler, for AV data storage; and, finally, 3) SmartViz, for real-time visualisation of the AV streams. StreamHandler and an instance of AV Registry are both deployed at the fog layer. DFB, Data Corpus, and SmartViz are all implemented at the Cloud layer, (PSNC HPC via OpenStack).

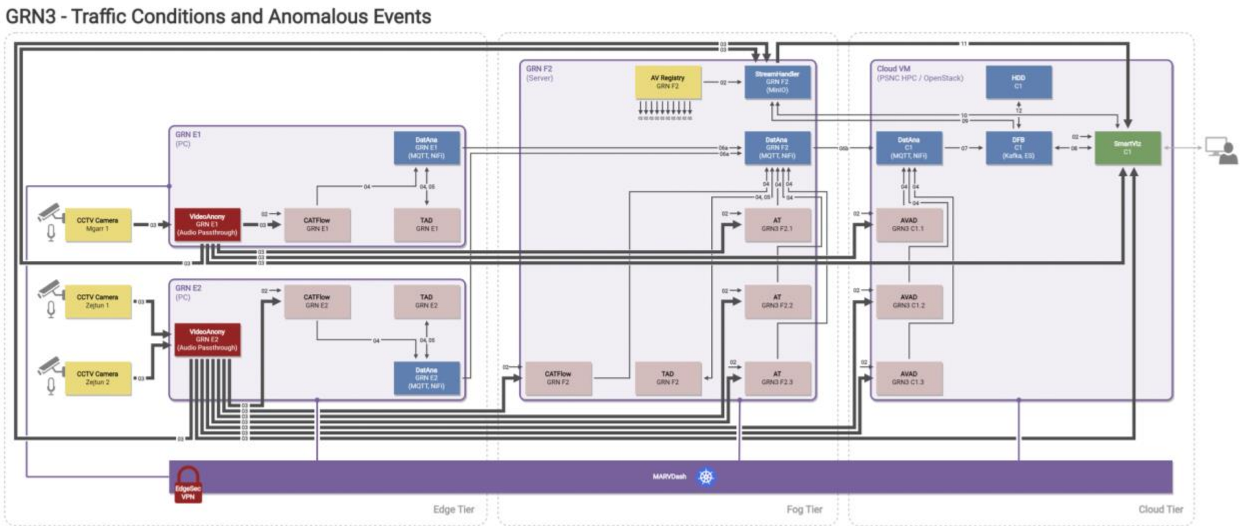


Figure 7: MARVEL RP2 deployment and runtime view of the MARVEL architecture for GRN3: Traffic Conditions and Anomalous Events (Source: D5.6)

DatAna is deployed at all layers of the architecture including the edge. After the inference results have been generated, either at the edge, fog or cloud, they are submitted to the respective MQTT broker of the DatAna agent residing at the same node. Each of the DatAna agents, after receiving the inference results, applies the appropriate transformation to standard data models, and relays the results to the next DatAna agent in the infrastructure, in the case of DatAna edge or DatAna fog, or sends these to the appropriate Kafka topics.

The AI components deployed in this use case are CATFlow+TAD, AVAD, and AT. The goal of the Text Anomaly Detection (TAD) component is to detect anomalies in vehicle trajectories after CATFlow processes the anonymised AV stream. The Audio-Video Anomaly Detection (AVAD) component is trained on normal traffic situations such that it can recognise any deviation from normal in previously unseen data, i.e., events that are not present in the training data are flagged as anomalous. The Audio Tagging (AT) component is deployed in GRN3 to enable tagging of audio traffic events at equally spaced time intervals. Data for the training of AT, AVAD, and VideoAnony was sourced from GRN sensors and is organised in three datasets. The GRN-AV-traffic-state dataset is used for training the AT model. The AVAD dataset on the other hand consisted of Audio-Video data that does not include anomalies (training set) and Audio-Video data that included anomalies (test set). Anomalies included temporary obstructions on the road, bicycles and U-turns. The fine-tuning of VideoAnony required an annotated dataset with bounding boxes over the vehicles' number plates.

SmartViz, Figure 8, in this use case incorporates both real-time and historical data feeds. The real-time messages generated by AVAD are transmitted to SmartViz through Kafka messages via the DFB. These real-time messages are utilised to populate and update the real-time map representation, providing immediate visibility into ongoing traffic anomalous events. On the other hand, the visualisations of the rest inference result in this use case draw upon the constantly updated historical data stored in the Elasticsearch database.

The Statistics widget features two line charts—one for Low-speed events and another for High-speed events—highlighting the five highest speeds detected. Users can zoom in and hover over the peaks to view additional information such as the exact speed and timestamp of each event. Additionally, a bar chart is used to display the output of Low and High speed events, along with the calculated average speed detected by TAD. Audio-Tagging (AT) data is visualised on a bar chart, yielding a clear representation of the different entities and their corresponding counts. Furthermore, the total number of detected pedestrians, calculated by CATflow, is displayed.

To enhance the organisation and relevance of events, the details widget is split into two: "Audio-visual anomaly detection," which visualises data from AVAD and AT, and "Traffic events detection," which visualises CATflow and TAD events. The Audio player widget allows users to select an event detected in a microphone stream and play the corresponding audio snippet. The functionality to rearrange and resize widgets, download visualised data in JSON format, and save the dashboard as a PDF file is also incorporated into this use case, providing users with greater flexibility and options for data analysis and offline access.

Finally, the Weather information widget in the use case provides users with a representation of weather-related data, allowing them to view and explore weather information for a selected time period. By visualising weather variables such as visibility, humidity, temperature, and overall weather conditions, users can gain insights and uncover hidden correlations between detections of events and anomalies and weather data that may influence them.

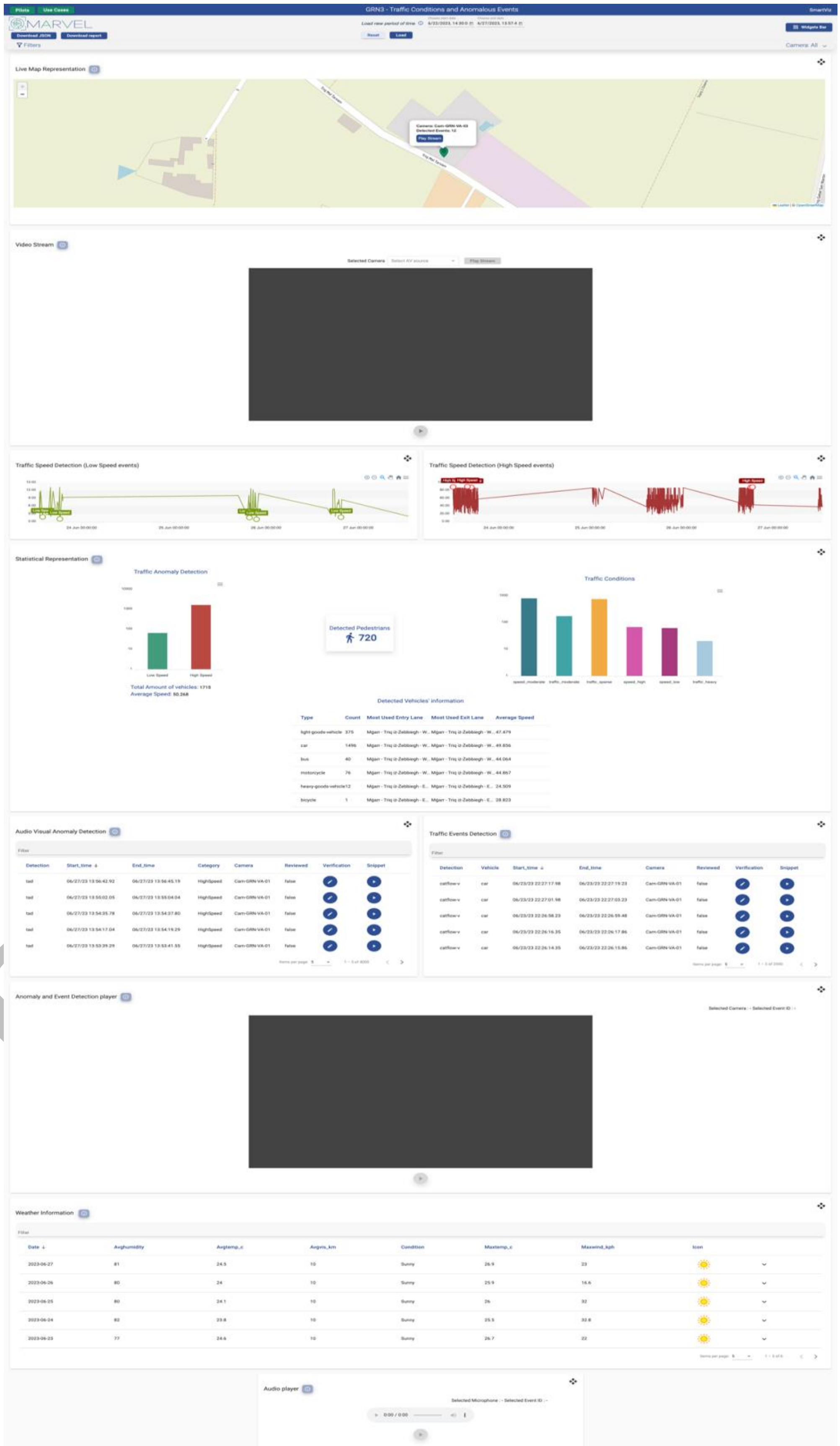


Figure 8: SmartViz dashboard for GRN3

5.4 KPI assessment

This section evaluates the use case against a set of functional KPIs (Table 9) and a set of non-functional parameters (Table 10). These tables are sourced from D1.2 and D6.1 and the evaluations are specific to the use case. Section 14 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs, Business KPIs and societal goals benchmarks.

5.4.1 Use case specific KPIs

The detailed results for the functional KPIs are tabulated in (Table 9). The performance of three types of models is evaluated; TAD/CATFlow, AVAD, and AT. The system achieves the target of the use case (i.e., correctly detecting various anomalies on the road at least 70% of the time). While the results have been achieved due to the performance of the TAD/CATFLow and AVAD models AT (which detected 48% of the occurrences) can potentially provide a lower cost system to deploy. Furthermore, the system latency measured (3.42sec) satisfies the requirement of the use case.

Table 9: Use case specific functional KPIs for GRN3: Traffic Conditions and Anomalous Events (Table 6.2 in D1.2)

<table>
<tr><th>GRN3</th><th></th><th></th><th></th><th></th></tr>
<tr><th>KPI</th><th>Metric</th><th>Baseline</th><th>Expected result/ Improvement</th><th>Evaluators</th></tr>
<tr><td rowspan="2">GNR-KPI4: Correctly detecting various anomalies on the road</td><td>Detection rate/F1</td><td>Labelled Dataset</td><td>70% detection of anomalous events</td><td>GRN Developers External traffic experts</td></tr>
<tr><td colspan="4">Evaluation and Comments

This use case makes use of a set of AI models namely, TAD (Text Anomaly Detection on CATFLow obtained data), AVAD (Audio Visual Anomaly Detection), and AT (Audio Tagging), to read the traffic state on the road and signals any anomalies that may be disruptive to traffic flow. In addition, visualisation plays a role during operation.

<u>TAD (CATFlow detections are used as inputs)</u>

The detailed technical evaluation methodology for the TAD model is documented in Section 8.1.5 of D5.5[3]

The CATFlow speed estimation is used to detect anomalous speeds. The precision per vehicle type when detecting anomalous high-speed vehicles is: Car: 97%, Light Good Vehicle: 82%, Motorcycle: 99%, Bus: 100%, Heavy good vehicle: 100%, and the precision per vehicle type when detecting anomalous low speed vehicles is: Car: 83%; Light Goods Vehicle: 75%; Motorcycle: 56%; Bus: 73%; Heavy Goods Vehicle: 76%. It is therefore concluded that anomalous high speeds can be detected with high precision, while low-speed precision is less. Nonetheless, the usefulness of the detections lies in the visualisation, i.e., a human would optically observe on the monitor significant anomalies being detected.

The CATFlow trajectory estimation is used as an input to detect trajectory based anomalous behaviour. The True Positive Rate per event is: vehicle crossing centre line - 5.1%; overtaking- 22.2%; significant traffic - 100.0 %; bicycle- 50.0 %; parking - 7.1%; obstruction -57.9 %; wrong lane - 10.5%; U-turn - 95.2 %, and the overall False Positive Rate is 1.2% (20% reduction in false detections when compared to the available one).

<u>AVAD (Audio Visual Anomaly Detection)</u>

The technical evaluation methodology for the AVAD model is documented in Section 4.1.4 of D3.5[7]

The AVAD method has been evaluated on the MARVEL - Malta Audio Visual Anomaly Dataset (MAVAD), which contains multiple recordings of situations considered normal, as well as anomalous ones. The dataset contains data from three scenes: Mgarr, Zejtun Scrapyard, and Zejtun Field, and the model is evaluated with the AUC (Area Under the ROC</td></tr>
</table>

	Curve) metric for various configurations of the audio component (zeroed audio, focused audio path, and plain audio path). The method achieved highest performance (89.76%) on the Mgarr scene when the focused audio path is used and 87.91% when the audio is zeroed. In general, including the audio path improves the performance by 1-4%. The performance achieved for the Zejtun Field scene is 80.45% and for the Zejtun Scrapyard, scene the performance dropped to 64.69%. The reason for this drop in performance is probably due to the lower quality video (camera). Furthermore, the effect of data anonymisation on the performance of the model is minor. For example, for the Mgarr scene the AUC is 91.49%, for Zejtun Scapyard scene AUC is 66.5%, and for Zejtun Field AUC is 80.29%. <u>AT (Audio Tagging)</u> *The technical evaluation methodology for the TAD model is documented in Section 8.4.7 of D5.5*[3]. The AT model is an audio only model that is intended for low power consumption. The model performance is evaluated on mean average precision (mAP). Average precision (AP) is the area under the recall-precision curve of a given class, while the mean AP is the average over all the classes available. GRN3 yielded a mAP of 0.48 compared to 0.39 with *dcase2019* baseline, relative increase of 23%. The considered classes in the AT models are (speed_high, speed_low, speed_moderate, speed_standstill, speed_none), (traffic_heavy, traffic_jam, traffic_moderate, traffic_sparse, traffic_none). The accuracy of the model when classifying the speed is 63.1% and when classifying traffic is 64.8%.			
GRN-KPI5: detection time	Time in minutes	No baseline	2 minutes from start of the anomalous event	GRN Developers
	Progress, Results and Comments	**Evaluation and Comments** Detection time is the time that elapses from the start of the anomalous event to the time when the event is posted on the SmartViz user interface and this is equal to the system latency. *The system latency for the GRN3 use case is measured and documented in Section 4.4.2 of D5.5*[3] The total system latency achieved is 3.42 seconds. Furthermore, the computational load at the Edge layer has very little effect on the global delay value.		

5.4.2 Use case specific non-functional KPIs

The detailed evaluations for the GRN3 non-functional KPIs, tabulated in Table 10, are extracted from surveys and interviews (Appendix section 17.1) as well as GRN Managers. In general, the system is considered to be very useful by experts and should have a significant impact on the day-to-day operations in the control room. In addition, the ability to share the data gathered on anomalous events with other transport entities was highlighted. From the IT infrastructure perspective, the cost of data processing needs to be considered. The use case considered the use of edge processing in addition to fog nodes. In the case of edge computational nodes, there is the added problem of extreme temperatures, which may result in the shutting down of the resource and loss of data. This issue requires longer-term studies.

Table 10: Use case specific non functional evaluation variables for GRN3: Traffic Conditions and Anomalous Events (6.3 in D1.2)

<table>
<tr><th>GRN3</th><th></th><th></th><th></th></tr>
<tr><th>Evaluation variable</th><th>How to measure</th><th>Internal evaluators</th><th>External Evaluators</th></tr>
<tr><td rowspan="2">End-user experience</td><td>Survey</td><td>GRN Managers</td><td>Transport Experts (Traffic Control Room Personnel)</td></tr>
<tr><td colspan="3">Evaluation and Comments
The evaluation is carried out from survey S3 (section 17.1.1) and interviews (section 17.1.7). Details of survey results are given in section 17.1.1.
Experts consider traffic jams, accidents on the road and vehicles obstructing the carriageway as the three most important anomalies to detect in real-time. 33.3% and 66.7% of respondents strongly agree or agree that the system would help personnel at the control room monitor a larger number of locations, decrease the time of response and therefore improve efficiency at the control room.
In terms of usability, all respondents agree that the system is easy and intuitive to use. In particular, the Real-time Map widget and the Video Stream Player are seen as essential and useful tools. A useful addition to the statistical tool is the potential inclusion of accident coordinates and clustering of accidents in the report generation task.
From an IT perspective, GRN engineers found it straightforward to add a new edge device to the system.</td></tr>
<tr><td rowspan="2">Scalability</td><td>Cost to add new video/audio feed</td><td>GRN Managers</td><td></td></tr>
<tr><td colspan="3">Evaluation and Comments
Financial Costs to set up new Camera node:
If not already available, the cost of installing the sensor node has to be factored in. Below are some typical costs at the time of writing:
• IP camera with A/V Stream Cost: 150-1000€
• Ancillary equipment and installation (Power supply, modem, casing): €500-€1000
• Internet Access: €20-€25 euros/month per shared node

Computational Cost to ingest stream with VideoAnony, CATFlow+TAD, AVAD, and AT:
The cost of continuously processing AV data needs to be considered
Below are equipment characteristics that can be used to assess cost.
• VideoAnony: 20GB Storage
• CATFlow + TAD: 50GB Storage, 1-2 CPUs, 3-4GB GPU RAM
• AVAD: 10GB Storage, 4 CPU cores, 16GB GPU RAM
• AT:4-16 GB Storage, 1-4 CPU cores, 8GB GPU RAM
Note: all components require a GPU for best Performance.</td></tr>
</table>

5.4.3 Additional parameters common across all GRN use cases from D6.1/D6.2

Table 4, in Section 3.4.3, tabulates the parameters that are common for all GRN use cases.

6 Use case GRN4: Junction Traffic and Trajectory Collection

6.1 Introduction

This section gathers information pertaining to use case GRN4: Junction Traffic and Trajectory Collection, which was implemented during RP1, and following feedback updated during RP2. The first two sections summarise the scope and description of the use case (Section 6.2) and the components, including relevant datasets and the configuration of the framework (Section 6.3). These sections provide a background to the use case evaluations reported in Section 6.4.

6.2 Scope and description of use case

Junction Traffic Trajectory collection (GRN4 use case) is focused on the requirement of long-term data analytics that shed light on both the behaviour of road users (e.g., car drivers, motorcyclists, cyclists, pedestrians, etc.) and on gathering traffic statistics on road network junctions. This use case is of interest in both short and long-term transport planning and evaluation. In particular, there is currently significant interest in studying active travel modes, such as cycling, walking, and micro-mobility, more generally. Authorities in Malta are interested in, for example, finding the optimal position of pedestrian crossings, whether provisions for cyclists at complex junctions are adequate, and whether installed provisions are being used as intended. Sustainable mobility is the societal challenge addressed by this use case.

From a technical point of view, this use case requires entity detection and its trajectory across a junction or road segment followed by descriptive statistics of network junction traffic. It therefore follows that entity detection and tracking models can be potentially used as a first processing stage, followed by further processing to generate descriptive statistics. The innovation that is targeted with this use case is the construction of a database that can be used to look up long-term historical data on the trajectories of vehicles and pedestrians at junctions. The trajectories and data generated from the CATFlow algorithm are saved on the MARVEL Data Corpus such that the data can be accessed and processed by the end-user. Patterns of trajectories can be detected through visual inspection of the trajectories, which is a feature of the system, whilst the system can also be used to detect anomalous ones.

The users for this system are intended to be transport policy-makers, transport engineers who need data to make informed decisions about infrastructure changes and upkeep, as well as transport researchers.

Table 11: Evaluation scenarios and relevant KPIs for the GRN4: Junction Traffic and Trajectory Collection (Table 9, D6.1)

Evaluation Scenario	Target	Relevant KPI
Detection of the trajectories and the storage of these trajectories to be used later in data-driven decision-making.	The aim is to have a 50% detection rate of the trajectories.	**GRN-KPI6 Availability of historical video samples of pedestrian and vehicle trajectories at two junctions.**
Surveys with relevant traffic experts to determine if this data will help in decreasing the planning time.	The aim is to have confirmation that this data is helpful.	**GRN-KPI7 Increased efficiency in the planning of roads**

6.3 Summary of Technical Setup and Decision Making Toolkit

Figure 9 is a diagram of the GRN4 frame configuration, which gives information on how the various components interact with each other. As can be inferred, the architecture is similar to that of GRN3, however, some of the components are different, namely the AI models. In addition to the basic framework, GRN4 makes use of the full CATFlow+TAD asset, the SED and the AVCC AI components. The DatAna component at each layer receives the results from the AI model instances at the respective layer, transforms them, and sends results to the appropriate Kafka topics of DFB.

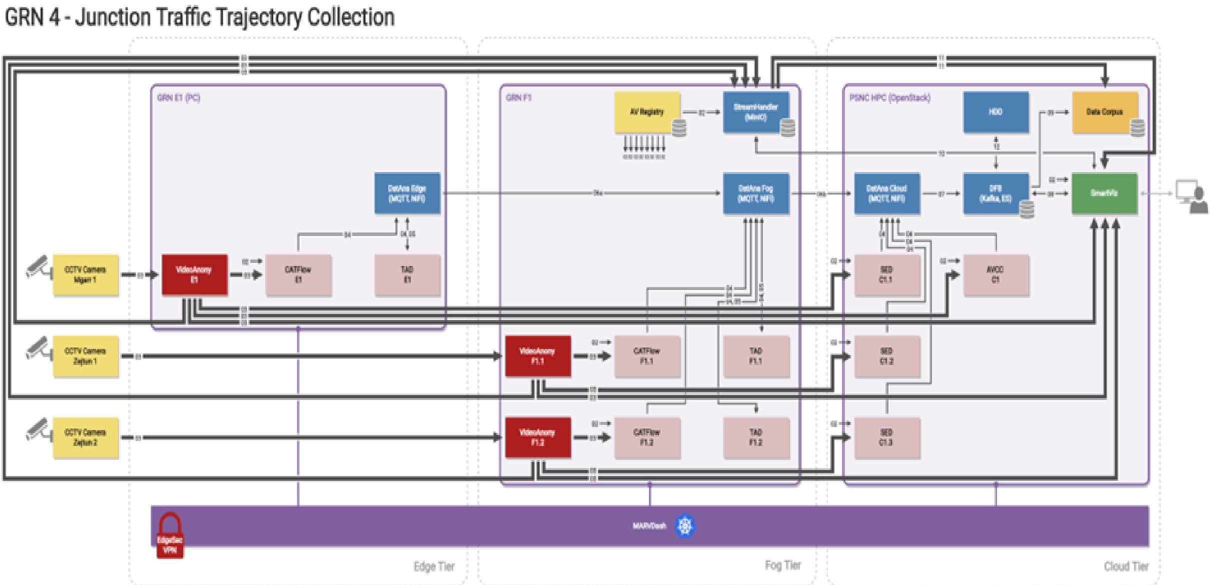


Figure 9: MARVEL RP2 deployment and runtime view of the MARVEL architecture for GRN4: Junction Traffic Trajectory Collection (source: D5.6)

The AI components deployed in GRN4 are CATFlow, TAD, SED and AVCC.

CATFlow classifies and tracks both vehicles and pedestrians. In addition, for implementation in GRN4, CATFlow is internally configured such that it outputs the trajectories of each vehicle as well as the entry and exit points within the camera field of view (FoV). TAD can be used to detect anomalous trajectories and speeds.

The Sound Event Detection (SED) component is deployed in GRN4 for the detection and classification of vehicles, such as cars, buses, and motorcycles, from the audio stream. Three instances of SED are deployed at the cloud (PSNC HPC via OpenStack) to process all three microphones at the Mgarr and the Zejtun locations.

The Audio Visual Crowd Counting (AVCC) component is deployed to process the Mgarr video stream to estimate the number of people present in the visual live feed. Crowd counting is the problem of identifying the number of people present in a visual scene. The technical details are given in D3.1[16]. The aim in the GRN4 is to count the pedestrians walking through the frame, thus for example helping in determining if the area being monitored requires additional pedestrian infrastructure. A second example is the case of using AVCC to count the number of pedestrians waiting at the bus stop, thus providing data for sustainable mobility research. The benefit of AVCC is that it should give accurate counts of pedestrians in crowds (i.e., when the

[16] MARVEL D3.1: Multimodal and privacy-aware audio- visual intelligence – initial version, 2022. https://doi.org/10.5281/zenodo.6821318

people are partially hidden by other people) which is often difficult for normal object detectors like CATFlow.

GRN-AV-traffic-entity dataset is used for SED training. For this dataset, audio-visual snippets were obtained from various locations around Malta and from the GRN static cameras. The audio track was manually annotated. All AV clips are 3-5 minutes long. Approximately 180 snippets were annotated for this dataset, 44 from different locations around Malta and the rest from the GRN static cameras. This amounts to more than 13 hours of annotated data.

In addition, GRN contributed 71 snippets, that contain pedestrians, to the MARVEL AVCC dataset. The AVCC dataset is annotated with the position of pedestrians' heads. These types of annotations are very time-consuming since a short video involves a large number of frames. The CVAT[17] software was used to streamline the process as much as possible. This software partially automates the annotation process by guessing the location of the pedestrian's head in the next frame.

Figure 10 is an example illustration of the SmartViz dashboard or DMT for GRN4. The user interface is split into four sections representing the variety of outputs from the various AI models. The data of detected vehicles from the Sound Event Detection (SED) component is presented to the user in a temporal form, through the Temporal Representation widget as well as the Details widget which also makes use of CATFlow+TAD data. The Details Widget presents incoming events in tabular format and provides a standalone text filtering capability that facilitates a rapid search through the available information. In the Vehicle Trajectories widget, the paths traversed by the vehicles detected by the CATflow component are drawn on the image of the camera feed. The paths are grouped and colour-coded depicting different types of vehicles. The users are also able to change the time period and filter by the type of vehicle or lane to investigate further. The statistics widget provides the counts (a distribution) of the different types of vehicles passing through a road junction in the form of a colour-coded bar chart. This will allow the user to track how frequently a road is being used and by what type of vehicle, which could then be used to make decisions on the road infrastructure, for example, decisions related to maintenance and infrastructure upgrades, for example installing a cycle lane. Additionally, the total number of detected pedestrians, calculated by CATflow, is displayed in the statistics widget. This provides users with a comprehensive overview of pedestrian-related data and insights. Furthermore, the Crowd Density Heatmap widget provides a heatmap of the AVCC output over the original camera image. Finally, the Weather information widget in the use case provides users with a representation of weather-related data, allowing them to view and explore weather information for a selected time period. By visualising weather variables such as visibility, humidity, temperature, and overall weather conditions, users can gain insights and uncover hidden correlations between detections of events and weather data that may influence them. Moreover, the DMT provides functionalities to rearrange and resize widgets, download visualised data in JSON format, and save the dashboard as a PDF file, thus providing users with greater flexibility and options for data analysis and offline access.

[17] https://cvat.org/

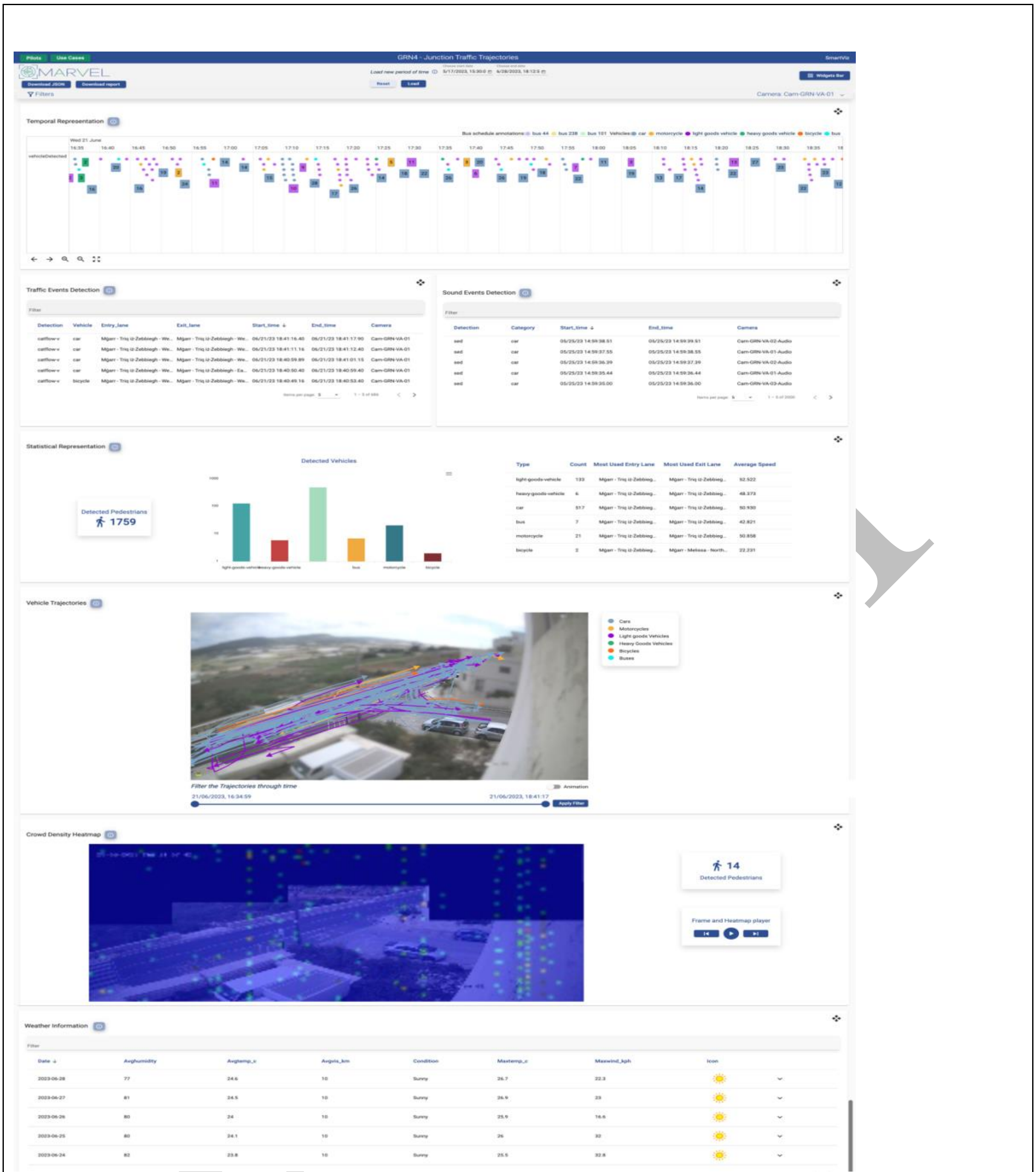


Figure 10: SmartViz dashboard for GRN4

6.4 KPI assessment

This section evaluates the use case against a set of functional KPIs (Table 12) and a set of non-functional parameters (Table 13). These tables are sourced from D1.2[2], D6.1[5] and D6.3[6] and the evaluations are specific to the use case. Section 14 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs, Business KPIs, and societal goals benchmarks.

6.4.1 Use case specific KPIs

The detailed results for the functional KPIs are tabulated in Table 12. The performance of three types of models is evaluated: TAD/CATFlow, SED, and AVCC. The system achieves the target of the use case (i.e., correctly detecting trajectories at least 50% of the time, as set in D1.2). The CATFlow model achieved an averaged object count accuracy of 90.82% and correctly extracted 84.03% of trajectories. The AVCC model used to count persons achieved good relative performance and is an alternative for monitoring specific areas, such as bus stops. On the other hand, the SED model achieved a performance score of 47.62%. The SED model can potentially provide a lower-cost system to deploy and improvements to this model are needed. It can therefore be concluded that the performance achieved across models satisfies the requirement of the use case.

Experts agree that increased efficiency is achieved when wider and/or long-term data is used in transport planning and when a disruptive event that require study occur. Vision models make it possible to count alternative modes and the possibility of long-term data not only improves the quality of planning but also allows much needed wider planning that is currently lacking. Current planning mostly depends on narrow studies using ad-hoc equipment, such as the temporary installation of a portable camera to collect data.

Table 12: Use case specific KPIs for GRN4: Junction Traffic and Trajectory Collection (Table 6.2 in D1.2)

GRN4				
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**
GRN-KPI6 Availability of historical video samples of pedestrian and vehicle trajectories at two junctions.	Automatic detection of patterns	No baseline	50% of the trajectories detected	GRN managers
	Evaluation and Comments This use case makes use of a set of AI models namely, CATFLow (Vision based Object detection and tracking), SED (Sound Event Detection), and AVCC (Audio Video Crowd Counting), to collect traffic and road junction trajectory data for use in long-term planning. CATFlow (Object Detection and Tracking) *The detailed technical evaluation methodology for the CATFlow model is documented in Section 8.1.2 of D5.5*[3] The model mAp (mean Average Precision across all classes) the detector achieved is 0.693, for a confidence threshold of 0.7. The mAp increases to 0.881 for a confidence threshold of 0.5. The mAp considers the overlap of the bounding boxes with the ground truth in the test set. The MOTA (Multi-Object Tracking Accuracy) achieved by the tracker is 54.33%. More importantly, 84.03% of trajectories averaged over various road scenes were correctly detected and the average count accuracy is 90.82%, depending on the camera position and look angle and environmental variables. SED (Sound Event Detection) *The detailed technical evaluation methodology for the SED model is documented in Section 8.4.6 of D5.5*[3]. The model detects occurrences of specific sounds on the road, for example, the vehicle engine, and can potentially be used as a vehicle counter. The F1- score (harmonic mean of precision and recall) achieved is of 47.62%, which may seem low but this is a 40% relative improvement when compared to 33.9% with the baseline. Furthermore, evaluation was done with macro-averaged metrics to score under-represented classes. AVCC (Audio-Visual crowd counting) *The detailed technical description for the AVCC model is documented in Section 3.4.4 of D5.5*[3]			

GRN4				
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**
	AVCC produces heatmaps (one in 100 frames by default) with the locations of the person heads highlighted and every output contains an estimate of the number of persons. As a comparison, it is closer to simply a counting task, than to object detection which usually includes location detection. AVCC was trained and tested on data from the Mgarr Camera. Evaluation carried out on the test set yielded a Mean Average Error (MAE) of 0.099 and a Mean Squared Error (MSE) of 0.158. This is considered a good performance when compared to the MAE of 13.0 on the DISCO benchmark dataset (Wang et al, "Audiovisual Crowd Counting dataset"[18].			
GRN-KPI7 Increased efficiency in the planning of roads	Surveys with road planners	No baseline	potential decrease in time for planning through the availability of data	External traffic experts
	Evaluation and Comments The evaluation is carried out from survey S4 (section 17.1.1) and interviews (section 17.1.7). Details of survey results are given in section 17.1.1. Experts indicated that currently not enough data is collected prior to planning, since collection of data occurs with the use of portable equipment, over a limited amount of time and space during the design stage of a project. This is basically due to the limitations of the data collecting equipment and budget. It follows that collecting the same amount of data via the MARVEL framework may not necessarily result in a significant decrease in time required in planning and design, unless the areas being studied are already covered by CCTV cameras and the video data has been stored, in which case the historical data is then processed with the AI models with the potential added benefit of sampling a larger area. On the other hand, the experts agreed that long-term data is necessary throughout the life-cycle of transport policy and infrastructure, contributing to the analysis of travel mode choice trends, planning infrastructural changes and developing mobility policies and evaluating the same. This is where the MARVEL framework will contribute to an increase in overall efficiency. One expert emphasised the importance of data sharing amongst the various transport entities, like planning, enforcement, and management and noted that MARVEL facilitates this requirement since a road engineer can for example use data generated from other entities to potentially discover bad designs in the infrastructure. One expert noted that a huge advantage of the MARVEL framework is the ability to process historical data. Most public CCTV systems store the data for a period of two to three weeks. Therefore, when an anomalous disruptive event occurs, it will be possible to process the data straight away and get insights into what triggered the event. Finally, one expert noted that the audio data collected can help with noise pollution studies. This aspect would require further development.			

6.4.2 Use case specific non-functional KPIs

The detailed evaluations for the GRN4 non-functional KPIs, tabulated in Table 13, are mostly obtained from surveys and interviews (Appendix section 17.1). Respondents emphasised that the possibility of collecting long-term data (especially over all users of the road infrastructure) is a strong point of MARVEL. SmartViz offers basic level useful tools for data analysis using

[18] https://zenodo.org/records/3828468

office level skills in the transport industry. For further analysis and correlation with other additional datasets (not from MARVEL), it is necessary to export the data and use third-party analysis tools. Finally, data anonymisation and privacy issues should not be overlooked.

Table 13: Use case specific non functional evaluation variables for GRN4: Junction Traffic and Trajectory Collection (Table 6.3 in D1.2)

GRN4 Evaluation variable	How to measure	Internal evaluators	External Evaluators
End-user experience	Survey	GRN Managers	Transport Experts (Traffic Engineer)
	Evaluation and Comments The evaluation is carried out from survey S4 (section 17.1.1) and interviews (section 17.1.7). Details of survey results are given in section 17.1.1. Experts fully agree that the data collected should cover the whole spectrum of road infrastructure users, as is the case in MARVEL, rather than limiting it to only motorised vehicles, as is the case in current adopted practices. Regarding usability, 75% of experts agree and strongly agree that SmartViz is intuitive and easy to use, whereas 25% took a neutral stance. The experts agree that the SmartViz user interface provides a good number of tools and widgets to enable the exploration of data, such as the selection and filtering options. One expert noted that it will still be necessary to export the data to other dashboards which read other datasets. Experts agree that the trajectory widget is a good and useful concept. However, one expert noted that [for certain studies] other types of data (not collected from the road, e.g., events data) would also be needed together in addition to the trajectory data.		
Data protection, privacy preservation	Manually evaluate sample of AV data.	GRN DPO	
	Evaluation and Comments See Table 4.		
Secure data transmission and cybersecurity in general	Draw a list of secure data characteristics and evaluate each.	GRN Engineers	
	Evaluation and Comments The EdgeSec Virtual Private Network (EdgeSec VPN) component was deployed on the GRN use cases to secure the communications between the interconnected components of the MARVEL platform within the three distinct layers (i.e., Edge, Fog, Cloud). Section 2.3.3 (D4.5[19]) and Section 3.2.3 (D5.5[3]) give the details on the use of EdgeSec VPN in the GRN pilot. EdgeSec VPN by its nature removes ISP and any other middleman from the list of components that need to be trusted. Threats such as ISP snooping, attacks over insecure wireless networks as well as compromised networking equipment are avoided. Tests on the MARVEL framework have demonstrated that data in transit is encrypted. Full details on EdgeSec VPN tests are given in Section 8.2.3 (D5.5[3]).		

6.4.3 Additional parameters common for all GRN use cases from D6.1/D6.2

Table 4, in section 3.4.3, tabulates the parameters that are common for all GRN use cases.

[19] MARVEL D4.5 - Security assurance and acceleration in E2F2C framework – final version, 2023: https://zenodo.org/records/8147058

7 Use case MT1: Monitoring of Crowded Areas

7.1 Introduction

This section gathers all the information pertaining to use case MT1: Monitoring of Crowded Areas, which was implemented during RP1, and following feedback updated during RP2. The first two sections summarise the scope and description of the use case (Section 7.2) and the components, including relevant datasets and the configuration of the framework (Section 7.3). These sections provide a background to the use case evaluations reported in Section 7.4.

7.2 Scope and description of use case

The goal is to select specific cameras (among all those available) that allow monitoring and verifying events like exceptional crowds, suspicious or anomalous behaviour of individuals or unusual crowd movements.

Piazza Fiera, a square renowned for hosting the annual "Christmas Markets" in Trento, is one such area. These markets draw in thousands of visitors, and the heightened crowd density increases the risk of thefts and assaults, as well as the potential necessity for medical assistance for individuals experiencing discomfort or fainting. Another site warranting vigilant observation is Piazza Duomo, a square in the city centre that hosts a weekly market. Similar to the Christmas Markets, this location is susceptible to overcrowding.

The MARVEL framework was deployed to detect the above-mentioned situations or anomalous events that occur and alert the operational centre of the Local Police and consequently the policeman/woman on or close to the site in quasi real-time. In addition to sending alerts, the MARVEL framework also activates custom views in the control room managed by the local police. Two evaluation scenarios have been defined for this use case related to two use case KPIs, as reported in Table 14. The MARVEL solution improves the local territory, the city surveillance, the quality of life of the citizens, their perceived security, and the experience of tourists.

Table 14: Evaluation scenarios and relevant KPIs for the MT1: Monitoring of Crowded Areas (Table 11, D6.1)

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario includes testing the various AI models on a labelled dataset to determine the detection rate and score achieved by the models.	Single person observing multiple cameras improve at least 10% the detection of anomalous events.	**MT-KPI1 increase the accuracy in detecting targeted events in crowds**
The evaluation scenario here involves observing the time taken to detect an anomaly through measuring the system's latency.	The aim is to detect anomalies 5 minutes from the start of the anomalous event.	**MT-KPI2 Detection time reduction.**

7.3 Summary of Technical Setup and Decision Making Toolkit

The MT1 infrastructure consists of six IP cameras transmitting video. The data flow starts from the IP cameras (three in Piazza Fiera and other three in Piazza Duomo). Each camera produces continuous video footage. FBK provides the fog tier for all the MT use cases. A secure transmission by VPN access between MT and FBK allows raw video to be sent to the FBK network. In order to comply with the constraints in the agreement that granted FBK access to the raw data of the MT’s sensors and to satisfy the requirements of the MARVdash Kubernetes cluster, FBK deploys two workstations, both with GPU.

At the fog tier, the data are anonymized thanks to the VideoAnony component. The anonymised streams are received and processed by CATFlow, ViAD, and VCC.

For the development and evaluation of the MT1, a dataset was collected using the video feeds from the selected cameras. The dataset was anonymised, annotated and made available to the partners. The dataset was named “TrentoOutdoor – real recording”, and its details are given in D2.1[20] and D8.2[21]. MT, in collaboration with FBK, collected and anonymised more than 500 videos of 3 minutes each, in total 165 GB used for the training of CATFlow (detection of pedestrians), ViAD (detection of anomaly situations), and VCC (number of detected pedestrians).

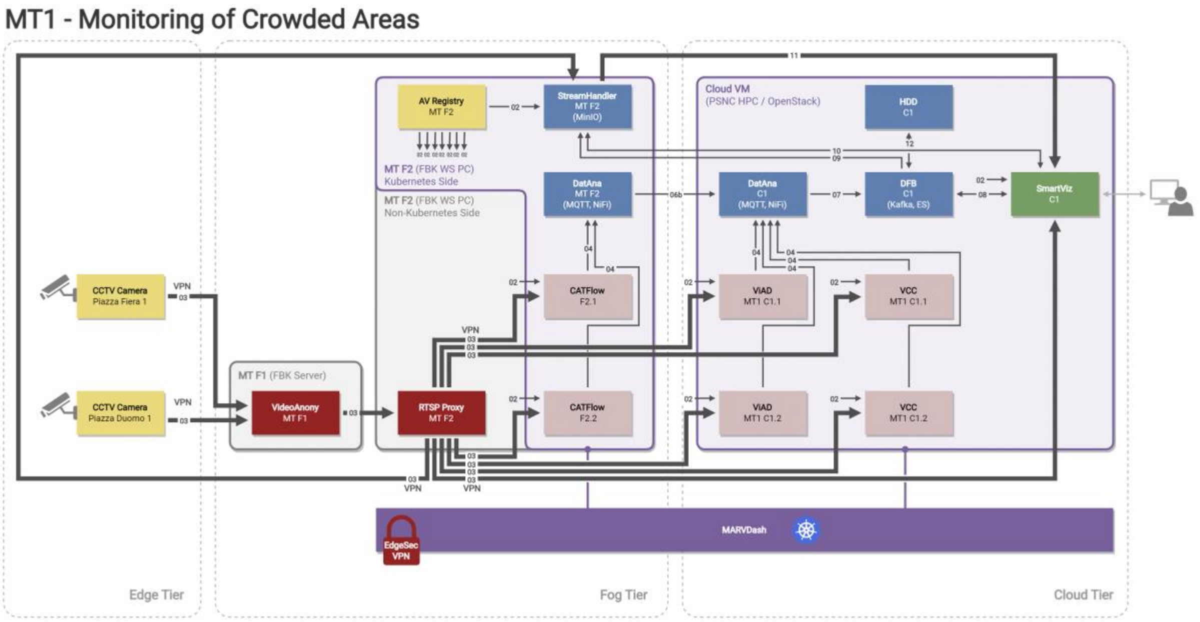


Figure 11: MARVEL RP2 deployment and runtime view of the MARVEL architecture for MT1 (source: D5.6)

SmartViz, the User Interface of the platform, supported several functionalities: 1) advanced visualisations, with a dashboard for user configurations and user interactions, 2) live AV feed for real-time inspection of the monitored areas, 3) on-request inspection of stored AV data, and 4) user-based verification of AI inference results.

[20] MARVEL D2.1: Collection and analysis of experimental data, 2021. https://doi.org/10.5281/zenodo.5052713

[21] MARVEL D8.2: MARVEL Data Management Plan, 2021. Confidential.

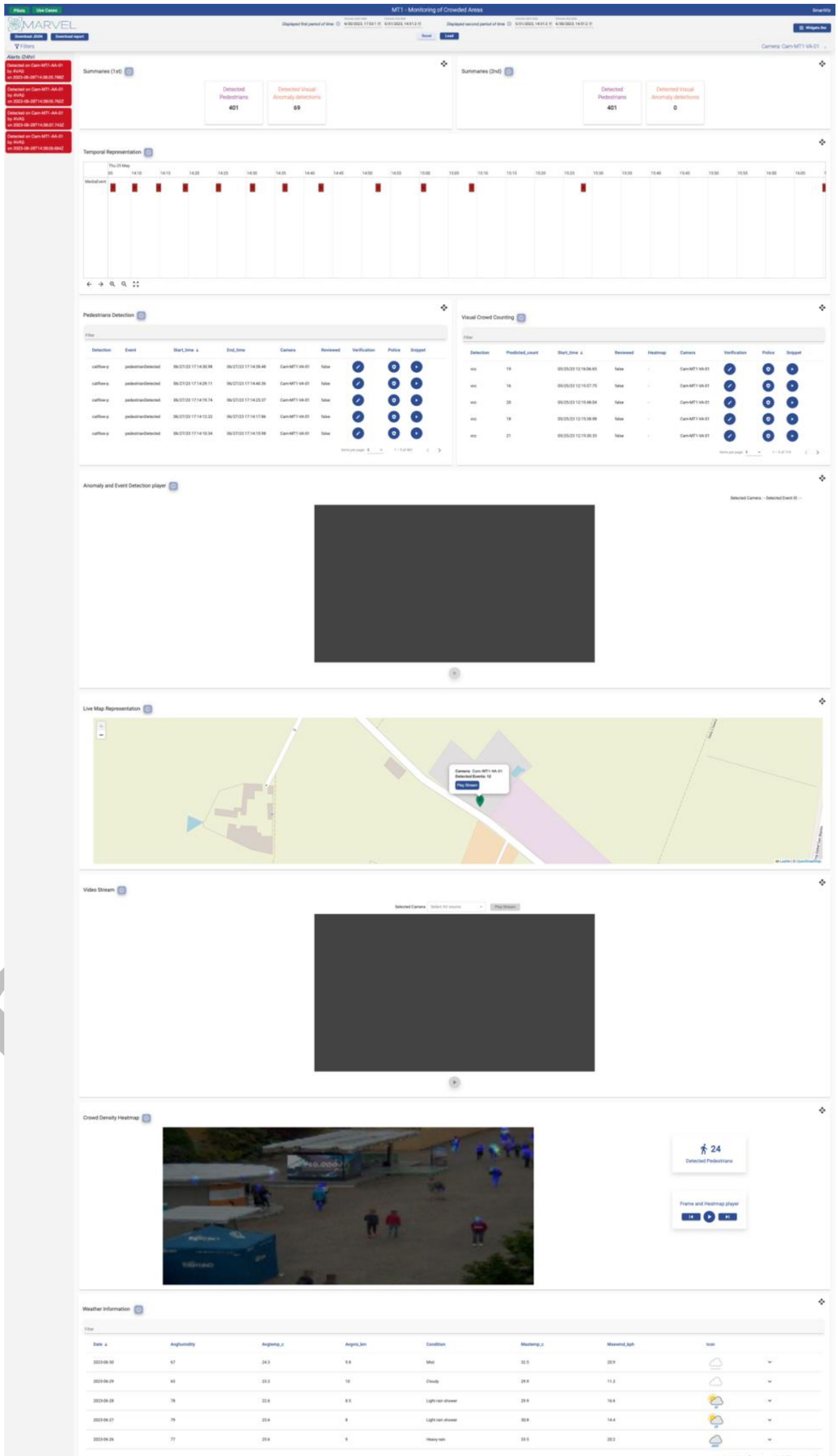


Figure 12: SmartViz dashboard for MT1

The Alerts widget has been integrated into this use case to address the user's requirement of being alerted when the number of people present exceeds a predetermined maximum limit (80 people in Piazza Fiera and 80 in Piazza Duomo). Real-time alerts are detected by VCC and ViAD and then transmitted to SmartViz through Kafka messages via the DFB component. These alerts are then visualised in the Alerts widget, enabling users to quickly respond and take appropriate actions based on the detected anomalies.

To facilitate comparative analysis, the comparison view has been integrated into the MT1. In SmartViz, users can select two different time periods and observe the events in different contexts, allowing for a comprehensive understanding of temporal patterns and changes.

The police intervention functionality is also included in the use case. After reviewing the events, users can mark an event as important and in need of police intervention.

In addition, the VCC component provides data on each frame of a video, indicating the likelihood of a pedestrian being present at that specific point. To visualise this data, the Crowd Density Heatmap has been added to the dashboard of this use case. The Crowd Density Heatmap widget overlays the heatmap onto the original frames, offering a visual representation of the crowd density.

Furthermore, the functionality to rearrange and resize widgets, download visualised data in JSON format, and save the dashboard as a PDF file has been incorporated into the system.

Finally, the Weather information widget in the use case provides users with a representation of weather-related data, allowing them to view and explore weather information for a selected time period. By visualising weather variables such as visibility, humidity, temperature, and overall weather conditions, users can gain insights and uncover hidden correlations between detections of events and anomalies and weather data that may influence them.

7.4 KPI assessment

This section evaluates the use case against a set of functional KPIs (Table 15) and a set of non-functional parameters (Table 16). These tables are sourced from D1.2[2], D6.1[5] and D6.3[6] and the evaluations are specific to the use case. Section 14 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs, Business KPIs and societal goals benchmarks.

7.4.1 Use case specific functional KPIs

This section tabulates the progress in evaluating the use case against a set of functional KPIs (described in Table 14 and evaluated in Table 15) and a set of non-functional parameters (Table 16).

Table 15: Use case specific KPIs for MT1: Monitoring of Crowded Areas (Table 6.2 in D1.2)

MT1				
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**
MT-KPI1 increase the accuracy in detecting targeted events in crowds.	Classification accuracy	Single person observing multiple cameras	10% improvement	MT managers
	Evaluation and Comments The local police operations centre currently employs two people who monitor 700 surveillance cameras from 7:00a.m. to 7:00p.m. The monitoring occurs retrospectively, following a dangerous situation. The operators inspect the images saved by the system, attempting to identify the footage of the incident for analysis. In only rare cases, when a demonstration or authorised public event is scheduled in			

<table>
<tr><th>MT1</th><th></th><th></th><th></th><th></th></tr>
<tr><th>KPI</th><th>Metric</th><th>Baseline</th><th>Expected result/ Improvement</th><th>Evaluators</th></tr>
<tr><td></td><td colspan="4">advance, operators dedicate themselves to real-time monitoring of the cameras assigned to monitor the areas affected by the event, seeking to identify potential danger. In these instances, the number of surveillance network cameras is a subset (on average 10-20) monitored by the operations centre operators. Even in this scenario, identifying a dangerous situation proves to be particularly challenging and it occurs in most cases retrospectively.

Only in a residual portion of cases, quantifiable at about 5%, can operators identify a dangerous situation if they don't know the location, date, and time of the event (and therefore only retroactively).

This use case makes use of a set of AI models namely, CATFLow (Vision based Object detection), VCC (Video Crowd Counting) and ViAD (Visual Anomaly Detection), to collect anomaly situations and presence of people in the square.

<u>CATFlow (Object Detection and Tracking)</u>

The detailed technical evaluation methodology for the CATFlow model is documented in Section 8.1.2 of D5.5.

The model mAp (mean Average Precision across all classes) the detector achieved is 0.693, for a confidence threshold of 0.7. The mAp increases to 0.881 for a confidence threshold of 0.5. The mAp considers the overlap of the bounding boxes with the ground truth in the test set. The MOTA (Multi-Object Tracking Accuracy) achieved by the tracker is 54.33%. More importantly for this use case, 84.03% of trajectories averaged over various road scenes were correctly detected and the average count accuracy is 90.82%.

<u>VCC (Visual Crowd Counting)</u>

VCC component solves the problem of counting the number of people presented in a camera image. The models are evaluated on a private MARVEL dataset and on a public ShanghaiTech dataset with Mean Average Error (MAE) and Mean Squared Error (MSE) metrics for accuracy and Floating-Point Operations (FLOPs) for model complexity.

The detailed technical evaluation methodology for the VCC model is documented in Section 8.4.3 of D5.5[3]

<u>ViAD (Visual Anomaly Detection)</u>

ViAD component is aimed at detecting novel situations in the scene of interest that can be described as unusual or anomalous.

The performance of the models is evaluated using an Area Under the ROC Curve (AUC) metric for accuracy and Floating-Point Operations (FLOPs) for model complexity. The benchmarks are performed on the MAVAD dataset.

The detailed technical evaluation methodology for the CATFlow model is documented in Section 8.4.1 of D5.5[3]

In parallel, MT staff checked the system stores video snippets of targeted events and that these are properly classified.

Finally, the functionality developed via SmartViz to validate the anomalies recorded by the system enables continuous improvement of the entire framework.</td></tr>
<tr><td rowspan="2">MT-KPI2 Detection time reduction.</td><td>Time</td><td>Single person observing multiple cameras</td><td></td><td>MT managers</td></tr>
<tr><td colspan="4">Evaluation and Comments</td></tr>
</table>

<table>
<tr><th>MT1</th><th></th><th></th><th></th><th></th></tr>
<tr><th>KPI</th><th>Metric</th><th>Baseline</th><th>Expected result/ Improvement</th><th>Evaluators</th></tr>
<tr><td></td><td colspan="4">MT staff investigated if the system could enable the Local Police to intervene in a shorter time than in the situation where an operator monitors the cameras without the help of the MARVEL framework.

Currently, the intervention time depends on various factors (number of available patrols, distance of the patrol from the scene of the event, etc.), but it occurs after receiving a request for intervention from the operational center and is estimated at around 20 minutes on average.

Thanks to the survey submitted to the Local Police of the Municipality of Trento, it emerged that the system could reduce intervention times:
• less than 10% by 18,2% of the voters;
• between 10% and 30% by 27,3% of the voters;
• between 30% and 50% by 36,4% of the voters;
• more than 50% by 18,2% of the voters.

Lastly, the total latency for the MT1 pipeline is 1.57 seconds, which can be considered very good for a processing inference pipeline. It means that in a few seconds after an event is caught in the AV streams, the event is processed and is available for the human operator to be visualised in SmartViz.</td></tr>
</table>

7.4.2 Use case specific non-functional KPIs

Table 16: Use case specific non functional evaluation variables for MT1: Monitoring of Crowded Areas (from Table 6.3 in D1.2)

<table>
<tr><th>MT1</th><th></th><th></th><th></th></tr>
<tr><th>Evaluation variable</th><th>How to measure</th><th>Internal evaluators</th><th>External Evaluators</th></tr>
<tr><td rowspan="2">End-user experience</td><td>Periodic Surveys</td><td>MT staff</td><td>Local Police</td></tr>
<tr><td colspan="3">Evaluation and Comments
The survey confirmed that the MARVEL framework is easy and helpful to use.
The final users consider that the framework enables Local Police to focus on relevant views, analyse detected anomalies, and determine appropriate courses of action during public events in crowded areas:
• very effectively by 9,1% of the voters;
• effectively by 63,6% of the voters;
• neutral 0% of the voters;
• ineffectively by 18,2% of the voters;
• very ineffectively by 9,1% of the voters;
Moreover, the alerting functionality enable Local Police to promptly respond to anomalous events in considered manner:
• very high by 18,2% of the voters;
• high by 45,5,% of the voters;
• neutral by 9,12% of the voters;
• low by 18,2% of the voters;
• other by 9,1% of the voters;</td></tr>
</table>

Data protection, privacy preserving	Data breach reports	IT managers/infrastructure managers	DPO
	Evaluation and Comments No data breach was reported. MT staff and FBK staff have complied with all the precepts established by the MT DPO.		
Scalability/Modularity	Extend the solution to other cameras and microphones (other places of city surveillance system)	IT managers/infrastructure managers	Local Police
	Evaluation and Comments The MARVEL framework is highly modular: and scalable adding a sensor is possible without impacting the system. Obviously, adding a sensor implies ensuring that the processing capabilities are available either at the fog or cloud.		

7.4.3 Additional parameters common across all MT use cases from D6.1/D6.2

A subset of parameters suggested in D6.1 are common for all use-cases. Table 17 tabulates these parameters and their assessment, which have been updated since D6.2.

Table 17: Additional parameters common across all MT use cases

Parameter	How to measure	Target to be achieved
Efficiency *Related to the efficiency of the system as used in the use case*	Efficiency is largely dependent on how long it takes to detect and flag targeted events. The average time taken to detect and flag the event	On average less time taken when compared to a single person observing multiple cameras
	Evaluation and Comments: Human operator might not notice relevant events by observing several streams, whereas system performs automatic detection of anomalous event, this according to the opinions collected through the survey, considerably decreases intervention time Currently, the intervention time depends on various factors (number of available patrols, distance of the patrol from the scene of the event, etc.), but it occurs after receiving a request for intervention from the operational center and is estimated at around 20 minutes on average. Thanks to the survey submitted to the Local Police and the Mobility Department of the Municipality of Trento, it emerged that the system could reduce intervention times in 50% of cases for more than 30% in all use cases in the Municipality of Trento. The latencies in the MT use cases can be considered higher than expected, most likely due to the complex solution to restream data from different MQTTs to isolate the edge and fog infrastructures. The fact that MT runs were scheduled first during the benchmarking tests, also impacted the resolution of bugs and other issues, which were already solved when running the rest of the pilots. At the same time, the system is able to identify potentially dangerous situations with certainly greater accuracy compared to the current situation, where the Local Police operators of the Municipality of Trento are only able to identify potential dangers retrospectively or with a 5% accuracy rate by observing the cameras in real time.	

Parameter	How to measure	Target to be achieved
Operability *Related to the ability of the components to keep functioning together*	Install additional cameras and microphones	System discovers sensors or adding sensors does not disrupt operation
	Evaluation and Comments: The MARVEL framework is highly modular: and scalable adding a sensor is possible without impacting the system. Obviously, adding a sensor implies ensuring that the processing capabilities are available either at the fog or cloud.	
Usability *Related the how well the system helps the users to achieve a task in a given use case*	Interview MT staff and local police to measure the end-user experience	End-users finds the system easy to use
	Evaluation and Comments: Survey confirm that the potential users consider the MARVEL framework easy and helpful to use. The opinions of the Commander of the Local Police and the Mobility Office, also surrogated by the feedback collected through the surveys, indicate how the graphical interface of the MARVEL framework helps to understand the collected data in a simple and intuitive way. The functionalities developed for the UI allow potential users to fully understand the potential of the system, resulting in an improved decision-making process and, consequently, in better quality of the local territory, the city surveillance, the quality of life of the citizens, their perceived security, and the experience of tourists. The functionalities developed for the UI allow potential users to fully understand the potential of the system, resulting in an improved decision-making process	
Robustness *Related to how robust are the system components during the period of operation*	Scale system to other cameras and microphones	Performance sustained as additional sensors are added
	Evaluation and Comments: The MARVEL framework is highly modular: and scalable adding a sensor is possible without impacting the system. Obviously, adding a sensor implies ensuring that the processing capabilities are available either at the fog or cloud.	
Performance *Related to how well the system performs the intended task*	Counting the number of detected targeted events in crowds	10% increase in the detection of events when compared to a single person observing multiple cameras
	Evaluation and Comments: MT staff checked the system stores video snippets of targeted events and that these are properly classified. The system is able to identify potentially dangerous situations with certainly greater accuracy compared to the current situation, where the Local Police operators of the Municipality of Trento are only able to identify potential dangers retrospectively or with a 5% accuracy rate by observing the cameras in real time. The functionality developed via SmartViz to validate the anomalies recorded by the system enables continuous improvement of the entire framework.	
Accountability *Related to the system being able to explain results or decisions.*	System stores video snippet of targeted event	Number of times system fails to store video snippet
	Evaluation and Comments: In RP2, the data storage functionalities (StreamHandler, also MARVEL Data Corpus) were focused on configurable and event-driven snippet selection. MT staff checked the system stores video snippets of targeted events and that these are properly classified. The functionality developed via SmartViz to validate the anomalies recorded by the system enables continuous improvement of the entire framework.	

<table>
<tr><th>Parameter</th><th>How to measure</th><th>Target to be achieved</th></tr>
<tr><td rowspan="2">Transparency

Related to the description of the processes or algorithms that are used to generate system output</td><td>Decision processes are described in a document.</td><td>Document availability</td></tr>
<tr><td colspan="2">Evaluation and Comments:

All inference results are stored in DFB and in addition, the end-user is given an explanation of how the system functions. The user are able to verify each inference result during system operation, which is subsequently captured in the relevant field of the Elasticsearch database.</td></tr>
<tr><td rowspan="2">Privacy awareness

Related to the provision of adequate governance mechanisms that ensure privacy in the use of data.</td><td>Count data breach reports.</td><td>Minimisation of count.</td></tr>
<tr><td colspan="2">Evaluation and Comments:

No data breach was reported .

MT staff and FBK staff were complied with all the precepts established by the MT DPO.</td></tr>
</table>

8 Use case MT2: Detecting Criminal and Anti-Social Behaviours

8.1 Introduction

This section gathers all the information pertaining to use case MT2: Detecting Criminal and Anti-Social Behaviours. The first two sections summarise the scope and description of the use case (Section 8.2) and the components, including relevant datasets and the configuration of the framework (Section 8.3). These sections provide a background to the use case evaluations reported in section 8.4.

8.2 Scope and description of use-case

The objective is to monitor specific areas for the purpose of identifying criminal or anti-social activities. The MARVEL framework was implemented to detect potentially dangerous situations. These situations include gatherings, robberies, aggressions, and drug trafficking. The system analyses the visual and audio data streams of the designated location and promptly alerts the local police operational centre, enabling them to dispatch a team to the scene.

Additionally, the streams are saved on the local police server for future investigations. The real-time analysis covers a daily timeframe. In the latter scenario, the data will be retained for seven days as per the privacy regulations outlined in the GDPR 2016/679. After this period, the data will be deleted unless the police receive specific requests for timely investigations.

In order to properly monitor the squares, the Local Police can display the total number of events and anomalies detected, the statistics of behaviour detected, the most frequently used descriptions associated with the events detected and the information on the observed detections.

Once an anomaly has been reported, the Local Police are alerted and notified which live feed of which camera needs to be closely monitored. Operators are able to access the audio and video feed a few minutes before the anomaly occurs, to accurately assess the event.

Finally, there is the possibility to mark the event as abnormal or not and to report whether the police have intervened or not.

Two evaluation scenarios have been defined for this use case related to two use case KPIs as reported in Table 18.

The MARVEL solution improves the local territory, the city surveillance, the quality of life of the citizens, their perceived security, and the experience of tourists.

Table 18: Evaluation scenarios and relevant KPIs for the MT2

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario involves observing the reduction in reaction time and taking appropriate intervention measures from the time of detection of the above events	The aim is to react to the detected anomalies in 5 minutes from the start of the anomalous event	**MT-KPI3: Reaction time in case of issues**
The evaluation scenario includes testing the various AI models on a labelled dataset to determine the detection rate and score achieved by the models	Single person observing multiple cameras improves at least 50% the detection of anomalous events	**MT-KPI4: Increase the detection of targeted events**

8.3 Summary of Technical Setup and Decision Making Toolkit

The MT2 infrastructure consists of two IP cameras transmitting video, two Raspberry Pis devices transmitting audio data collected from the connected IFAG microphones and two workstations at the fog layer managed by FBK.

The MT2 edge layer infrastructure consists of two IP cameras transmitting video and two Raspberry Pi with a connected IFAG microphone. The Raspberry Pi runs a voice activity detector and a speech anonymiser and streams the anonymised audio via RTSP. The upload function is a secure transmission by VPN access between MT and FBK in which raw data will be sent to the data lake in FBK.

FBK provides the Fog tier for the MT use cases. To comply with the constraints of the agreement between MT and FBK (FBK was nominated as data processor) that granted FBK access to the raw data of the MT's sensors and to satisfy the requirements of the MARVdash Kubernetes cluster, FBK deploys two workstations, both with GPU.

MT has provided data for various components. MT has provided the dataset "TrentoOutdoor – real recording" and "TrentoOutdoor – staged recording" (as defined in D2.1) processing the streams of cameras and microphones from Piazza Santa Maria Maggiore.

MT, in collaboration with FBK, collected, anonymised, and annotated more than 180 audio and video files, used for the training of AAC associated at GPURegex (textual description of audio anomaly situations), SED (detection of audio anomaly situations), AT (classification of audio scenarios) and AVAD (detection of audio-video anomaly situations).

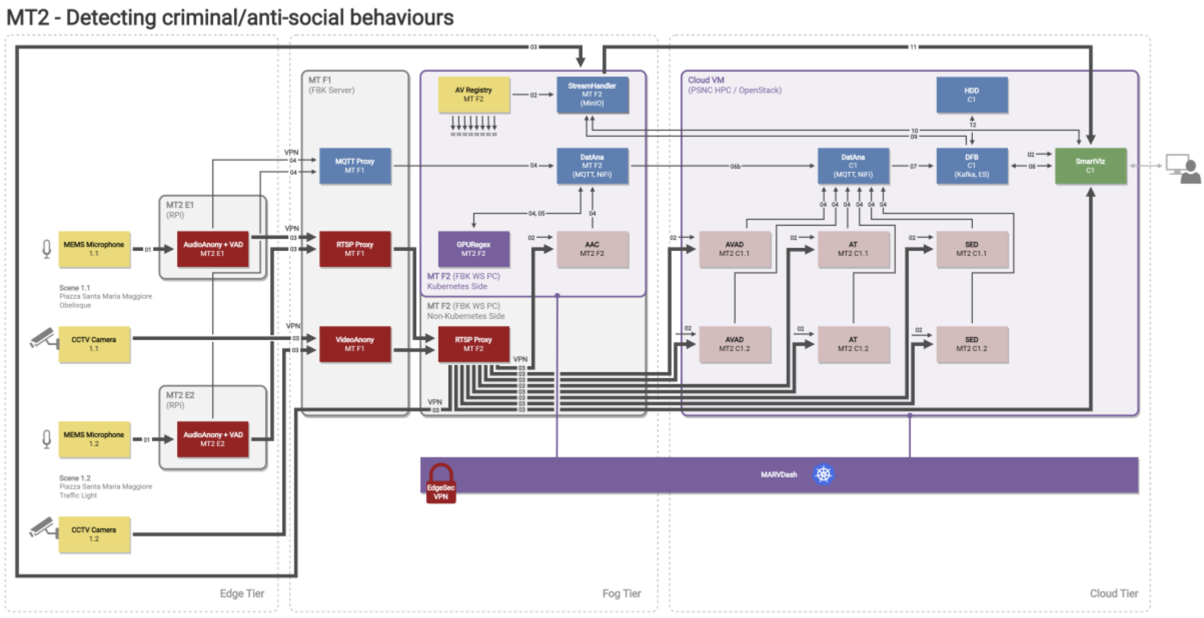


Figure 13: MARVEL RP2 deployment and runtime view of the MARVEL architecture for MT2 (source: D5.6)

SmartViz, the User Interface of the platform, supported several functionalities: 1) advanced visualisations, with a dashboard for user configurations and user interactions, 2) live AV feed for real-time inspection of the monitored areas, 3) on-request inspection of stored AV data, and 4) user-based verification of AI inference results.

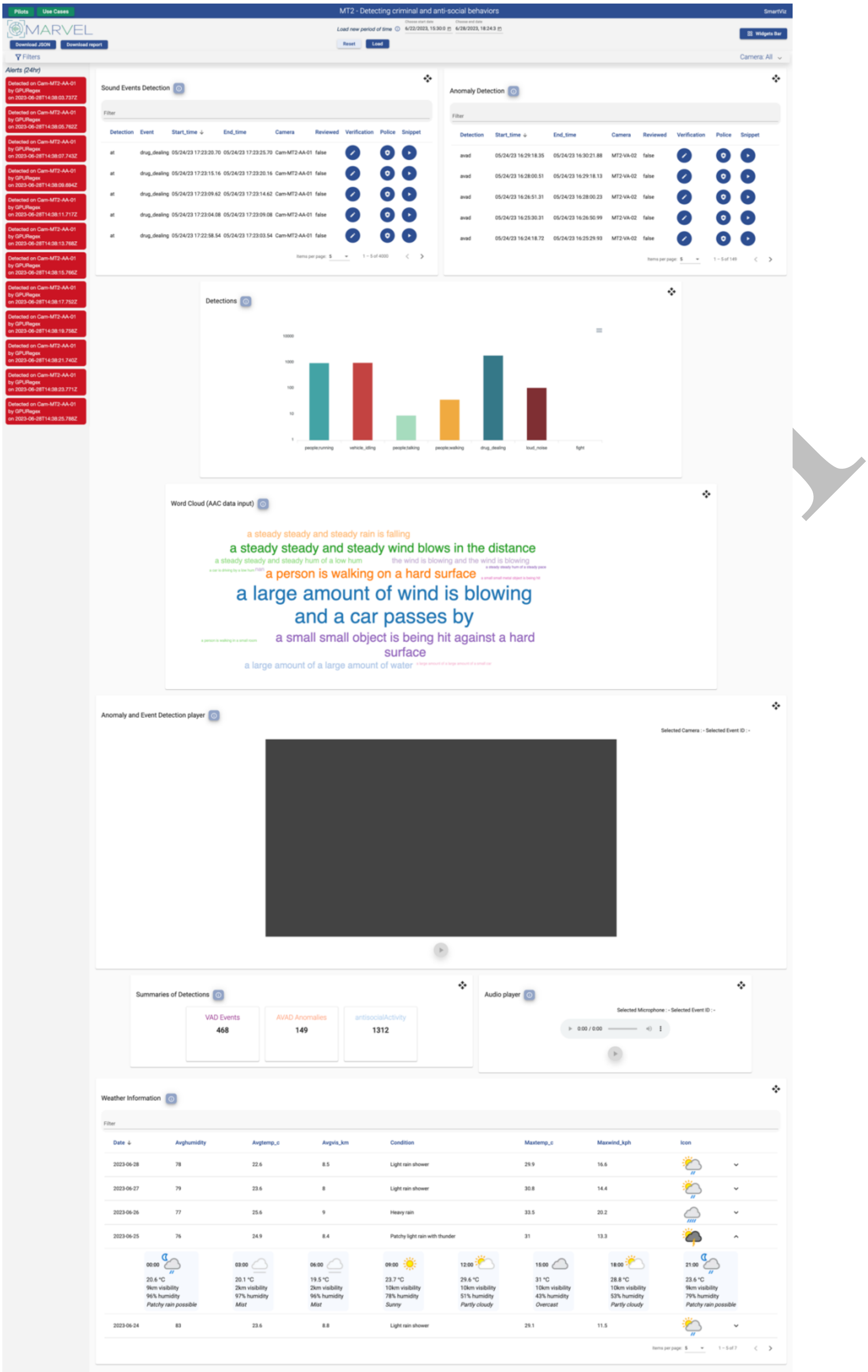


Figure 14: SmartViz dashboard for MT2

The Alerts widget has been integrated into this use case to address the user's requirement of being alerted when an anomaly occurs. Real-time alerts are detected by GPURegex and transmitted to SmartViz through Kafka messages via the DFB component. These alerts are then visualised in the Alerts widget, enabling users to quickly respond and take appropriate actions based on the detected anomalies.

The Summaries widget provides an overview of the total number of detected anomalies and events, utilising outputs from SED, GPURegex and AVAD components.

The Word Cloud widget, developed specifically for the AAC component, allows users to visualise the most common keywords and descriptions associated with the detected events.

Within the audio and video player functionality in SmartViz, users can select an event and play the corresponding audio or video snippet. The StreamHandler component retrieves the stream for the relevant time period of the detection, segments it, and plays the snippet in the Audio or Video player widget within SmartViz.

The Details widget visualises the detections in a textual format, with SED and AT displayed in the "Sound events and anomalies detection" table and AAC displayed in the "Anomalous Audio Captioning" table.

Users have the ability to validate the inference results by marking them as accurate or not. The verification of inference results is sent through Kafka messages from SmartViz to the DFB, which then updates the status of the corresponding event and stores them in an Elasticsearch index accessed by the Data Corpus. The police intervention functionality is also included in the use case. After reviewing the events, users can mark an event as important and in need of police intervention, facilitating appropriate action.

In addition to rearranging and resizing widgets, the dashboard view offers the functionality to download the visualised data in JSON format. Furthermore, users have the option to save the entire dashboard as a PDF file, allowing for offline access and convenient sharing based on individual preferences.

Finally, the Weather information widget in the use case provides users with a representation of weather-related data, allowing them to view and explore weather information for a selected time period. By visualising weather variables such as visibility, humidity, temperature, and overall weather conditions, users can gain insights and uncover hidden correlations between detections of events and anomalies and weather data that may influence them.

8.4 KPI assessment

This section evaluates the use case against a set of functional KPIs (Table 19) and a set of non-functional parameters (Table 20). These tables are sourced from D1.2[2], D6.1[5] and D6.3[6] and the evaluations are specific to the use case. Section 14 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs, Business KPIs and societal goals benchmarks.

8.4.1 Use case specific KPIs

Table 19: Use case specific KPIs for MT2: Detecting Criminal and Anti-Social Behaviours (Table 6.2 in D1.2)

MT2					
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**	**Results and Comments**
MT-KPI1 increase the accuracy in	Classification accuracy	Single person observing	10% improvement	MT managers	

<table>
<tr><td rowspan="2">detecting targeted events in crowds.</td><td></td><td>multiple cameras</td><td></td><td></td><td></td></tr>
<tr><td colspan="5">
Evaluation and Comments

The local police operations centre currently employs 2 people who monitor 700 surveillance cameras from 7:00 a.m. to 7:00 p.m. The monitoring occurs retrospectively, following a dangerous situation. The operators inspect the images saved by the system, attempting to identify the footage of the incident for analysis. In only rare cases, when a demonstration or authorized public event is scheduled in advance, operators dedicate themselves to real-time monitoring of the cameras assigned to monitor the areas affected by the event, seeking to identify potential danger. In these instances, the number of surveillance network cameras is a subset (on average 10-20) monitored by the operations centre operators. Even in this scenario, identifying a dangerous situation proves to be particularly challenging and it occurs in the majority of cases retrospectively.

Only in a residual portion of cases, quantifiable at about 5%, can operators identify a dangerous situation if they don't know the location, date, and time of the event (and therefore only retroactively)

This use case makes use of a set of AI models namely AAC (Automatic Audio Captioning), AT (Audio Tagging), SED (Sound event detection) and AVAD (Audio-Visual Anomaly Detection), to collect anomaly situations in the square.

<u>AAC (Automatic Audio Captioning)</u>

AAC provides textual descriptions for segments of audio, offering valuable content descriptions for humans accessing audio-visual data.

To assess the performance of AAC systems, several metrics originally designed for image captioning have been adapted. The selected one to evaluate the model is SPIDEr and the result achieved is 0.546

The detailed technical evaluation methodology for the AAC model is documented in Section 8.4.5 of D5.53

<u>AT (Audio Tagging)</u>

AT aims to identify and classify sounds within audio segments of predefined fixed length. This functionality offers the ability to recognise sounds related to actions or events where the start or end timestamps of these sounds are not important.

The model achieved Accuracy 38.54%, F1-score 39.71% and mAP 0.54

The detailed technical evaluation methodology for the AT model is documented in Section 8.4.7 of D5.5[3]

<u>SED (Sound event detection)</u>

SED component can identify specific sounds within brief time intervals, such as one second. These specific sounds are referred to as characteristic sounds, each of which can be described by a unique label, known as a sound event.

The algorithm achieved F1-score 45.23% (macro)

The detailed technical evaluation methodology for the SED model is documented in Section 8.4.6 of D5.5[3]

<u>AVAD (Audio-Visual Anomaly Detection)</u>

AVAD component uses a combination of audio and video data sequences to analyse a specific time frame for any anomalous events, providing a score which represents the likelihood of an anomaly happening in this time frame.

The detailed technical evaluation methodology for the AVAD model is documented in Section 8.4.2 of D5.5[3]

In parallel, MT staff checked the system stores video snippets of targeted events and that these are properly classified.
</td></tr>
</table>

<table>
<tr><td></td><td colspan="5">Finally, the functionality developed via SmartViz to validate the anomalies recorded by the system enables continuous improvement of the entire framework.</td></tr>
<tr><td rowspan="2">MT-KPI2 Detection time reduction.</td><td>Time</td><td>Single person observing multiple cameras</td><td></td><td>MT managers</td><td></td></tr>
<tr><td colspan="5">Evaluation and Comments

MT staff investigated if the system could enable the Local Police to intervene in a shorter time than in the situation where an operator monitors the cameras without the help of the MARVEL framework.

Currently, the intervention time depends on various factors (number of available patrols, distance of the patrol from the scene of the event, etc.), but it occurs after receiving a request for intervention from the operational center and is estimated at around 20 minutes on average.

Thanks to the survey submitted to the Local Police of the Municipality of Trento, it emerged that the system could reduce intervention times:
• between 10% and 30% by 36.4% of the voters;
• between 30% and 50% by 45.5% of the voters;
• more than 50% by 18.2% of the voters.

Lastly, the total latency for the MT2 pipeline is high due to the individual latency introduced by VAD components, around 30 seconds. After investigation, the bottleneck seems to be related to the re-streaming of MQTT messages between the two fog servers used to isolate the MARVEL from the MT infrastructure, affecting VAD, as the only inference component running at the edge in this scenario.</td></tr>
</table>

8.4.2 Use case specific non-functional KPIs

Table 20: Use case specific non functional evaluation variables for MT2: Detecting Criminal and Anti-Social Behaviours (from Table 6.3 in D1.2)

<table>
<tr><th>MT2</th><th></th><th></th><th></th></tr>
<tr><th>Evaluation variable</th><th>How to measure</th><th>Internal evaluators</th><th>External Evaluators</th></tr>
<tr><td rowspan="2">End user experience</td><td>Periodic Surveys</td><td>MT staff</td><td>Local Police</td></tr>
<tr><td colspan="3">Evaluation and Comments
The survey confirmed that the MARVEL framework is easy and helpful to use.
The final users consider that the framework enables Local Police to functionality contribute to the user's ability to flag specific events within the visualisation interface and highlight those events as significant for further attention from authorities:
• very high by 36,4% of the voters;
• high by 45,5% of the voters;
• neutral by 18,2% of the voters;
Moreover, the statistics widget, in visually representing the numbers of categories associated with detected criminal or anti-social activities are considered:
• very effective by 27,3% of the voters;
• effective by 54,5,% of the voters;
• neutral by 18,2% of the voters;</td></tr>
</table>

	Data protection, privacy preserving	Data breach reports	IT managers/infrastructure managers	DPO
	Evaluation and Comments No data breach was reported. MT staff and FBK staff have complied with all the precepts established by the MT DPO			
	Scalability/Modularity	Extend the solution to other cameras and microphones (other places of city surveillance system)	IT managers/infrastructure managers	Local Police
	Evaluation and Comments The MARVEL framework is highly modular: and scalable adding a sensor is possible without impacting the system. Obviously, adding a sensor implies ensuring that the processing capabilities are available either at the fog or cloud.			

8.4.3 Additional parameters common across all MT use cases from D6.1/D6.2

A subset of parameters suggested in D6.1 are common for all use-cases. Table 17 tabulates these parameters and their assessment, which have been updated since D6.2.

Table 17, in Section 7.4.3, tabulates the parameters that are common for all MT use cases.

9 Use case MT3: Monitoring of Parking Places

9.1 Introduction

This section gathers all the information pertaining to use case MT3: Monitoring of Parking Places, which was implemented during RP1, and following feedback updated during RP2. The first two sections summarise the scope and description of the use case (Section 9.2) and the components, including relevant datasets and the configuration of the framework (Section 9.3). These sections provide a background to the use case evaluations reported in Section 9.4.

9.2 Scope and description of use case

This use case concerns audio-visual monitoring of a parking lot, including analysis of car trajectories, detection of cars moving out of the parking slots, car damages and robberies, obstructions, etc. The target of this use case is the “Ex Zuffo” Parking Area which is one of the largest parking lots in Trento (around 1000 parking places). It is typically used by citizens who park their cars and then move around the city centre using public transportation, bike-sharing services or e-scooters.

The aim of the use case is to detect anomalies, for example, vandalism and other anomalous behaviour, the timeline distribution of parking activity and the clustering of vehicles or events. When an anomaly is detected, an alert is sent to the Local Police headquarter, who could then check the live feed from the camera where the event is occurring to accurately assess the event. The MARVEL framework will therefore support the prevention of robberies or damages to parked cars through audio-video analytics using the existing cameras and the microphones that have been installed thanks to the MARVEL project. The audio-visual analysis will be carried out in real-time and also on recorded data saved on the servers of the Local Police.

Finally, there is the possibility to mark the event as abnormal or not and to report whether the police have intervened or not.

Two evaluation scenarios have been defined for this use case related to two use case KPIs as reported in Table 21.

The MARVEL solution improves the local territory, the city surveillance, the quality of life of the citizens, their perceived security, and the experience of tourists.

Table 21: Evaluation scenarios and relevant KPIs for the MT3, (Table 13, D6.1)

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario involved the increased detection of targeted events, like car damages, checking the number of campers, and the average length of stay.	Single person observing multiple cameras improve at least 50% the detection of anomalous events.	**MT-KPI5 Increase the detection of targeted events**
The evaluation scenario here involves observing the reduction of detection time needed to identify the events mentioned above.	The aim is to detect anomalies 5 minutes from the start of the anomalous event.	**MT-KPI6 Detection time reduction.**

9.3 Summary of Technical Setup and Decision Making Toolkit

The MT3 infrastructure consists of one IP camera transmitting video, one Raspberry Pi device transmitting audio data collected from the connected IFAG microphone and two workstations at the fog layer managed by FBK.

The MT3 edge layer infrastructure consists of one IP camera transmitting video and two Raspberry Pi with a connected IFAG microphone. The Raspberry Pi runs a voice activity detector and a speech anonymiser and streams the anonymised audio via RTSP. The upload function is a secure transmission by VPN access between MT and FBK in which raw data will be sent to the data lake in FBK.

FBK provides the Fog tier for the MT use cases. To comply with the constraints of the agreement between MT and FBK (FBK was nominated as data processor) that granted FBK access to the raw data of the MT's sensors and to satisfy the requirements of the MARVdash Kubernetes cluster, FBK deploys two workstations, both with GPU.

MT has provided data for various components. MT has provided the dataset "TrentoOutdoor – real recording" and "TrentoOutdoor – staged recording" (as defined in D2.1) processing the streams of cameras and microphones from Piazzale ex Zuffo.

MT, in collaboration with FBK, collected, anonymised, and annotated more than 200 audio and video files, used for the training of SED (detection of audio anomaly situations), AT (classification of audio scenarios) and AVAD (detection of audio-video anomaly situations).

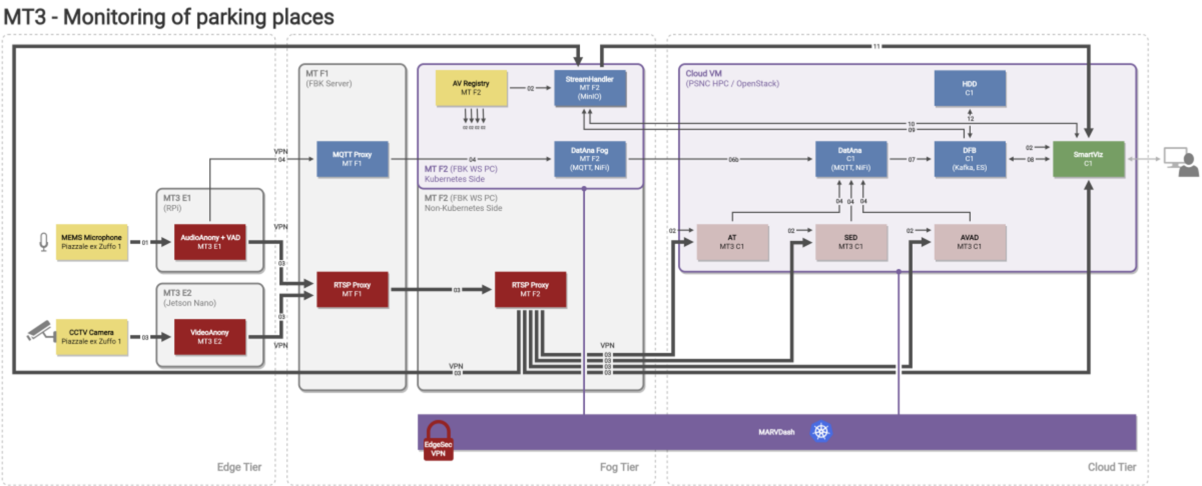


Figure 15: MARVEL RP2 deployment and runtime view of the MARVEL architecture for MT3 (source: D5.6)

SmartViz, the User Interface of the platform, supported several functionalities: 1) advanced visualisations, with a dashboard for user configurations and user interactions, 2) live AV feed for real-time inspection of the monitored areas, 3) on-request inspection of stored AV data, and 4) user-based verification of AI inference results.

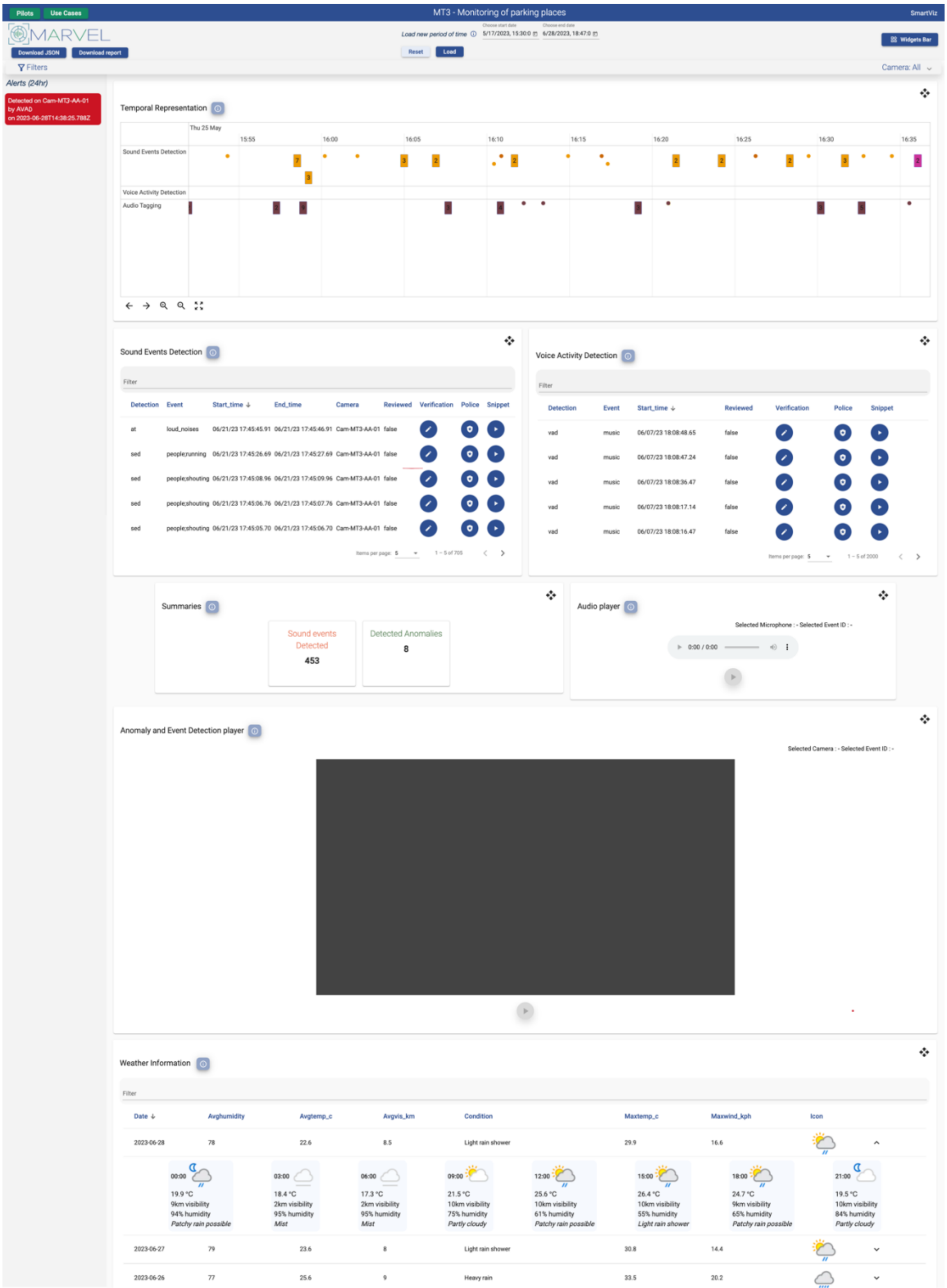


Figure 16: SmartViz dashboard for MT3

The Alerts widget has been integrated into this use case to address the user's requirement of being alerted when an anomaly occurs. Real-time alerts are detected by AVAD and transmitted to SmartViz through Kafka messages via the DFB component. These alerts are then visualised in the Alerts widget, enabling users to quickly respond and take appropriate actions based on the detected anomalies.

The addition of the Audio player widget is also included in RP2 for MT3 since AT and SED are dealing particularly with audio streams. Within the audio player functionality in SmartViz, users can select an event detected in a microphone stream and play the corresponding audio snippet. This feature is facilitated by the StreamHandler component, which retrieves the audio from the microphone stream for the relevant time period of the detection. The audio stream is segmented, and a URL containing the audio for the selected event is played in the Audio player widget within SmartViz.

The police intervention functionality is also included in the use case. After reviewing the events, users have the ability to mark an event as important and in need of police intervention.

Moreover, the functionality to rearrange and resize widgets, download visualised data in JSON format, and save the dashboard as a PDF file has been incorporated, providing users with greater flexibility and options for data analysis and offline access.

Finally, the Weather information widget in the use case provides users with a representation of weather-related data, allowing them to view and explore weather information for a selected time period.

9.4 KPI assessment

This section evaluates the use case against a set of functional KPIs (Table 22) and a set of non-functional parameters (Table 23). These tables are sourced from D1.2[2], D6.1[5] and D6.3[6] and the evaluations are specific to the use case. Section 13 on the other hand deals with KPIs across the project or the pilots, i.e project KPIs, Business KPIs and societal goals benchmarks.

9.4.1 Use case specific KPIs

This section tabulates the progress in evaluating the use case against a set of functional KPIs (described in Table 21 and evaluated in Table 22) and a set of non-functional parameters (Table 23).

Table 22: Use case specific KPIs for MT3: Monitoring of Parking Spaces (Table 6.2 in D1.2)

MT3				
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**
MT-KPI5 Increase the detection of targeted events	Classification accuracy	Single person observing multiple cameras	50% of dangerous situations are correctly noted	MT managers
	Evaluation and Comments The local police operations centre currently employs 2 people who monitor 700 surveillance cameras from 7:00 a.m. to 7:00 p.m. The monitoring occurs retrospectively, following a dangerous situation. The operators inspect the images saved by the system, attempting to identify the footage of the incident for analysis. In only rare cases, when a demonstration or authorized public event is scheduled in advance, operators dedicate themselves to real-time monitoring of the cameras assigned to monitor the areas affected by the event, seeking to identify potential danger. In these instances, the number of surveillance network cameras is a subset (on average 10-20) monitored by the operations center operators. Even in this scenario, identifying a dangerous situation proves to be particularly challenging and it occurs in the majority of cases retrospectively.			

<table>
<tr><td></td><td colspan="4">
Only in a residual portion of cases, quantifiable at about 5%, can operators identify a dangerous situation if they don't know the location, date, and time of the event (and therefore only retroactively)
This use case makes use of a set of AI models namely AT (Audio Tagging), SED (Sound event detection) and AVAD (Audio-Visual Anomaly Detection), to collect anomaly situations in the square.
<u>AT (Audio Tagging)</u>
AT aims to identify and classify sounds within audio segments of predefined fixed length. This functionality offers the ability to recognise sounds related to actions or events where the start or end timestamps of these sounds are not important.
The model achieved Accuracy 52.69%, F1-score 56.36% and mAP 0.67.
The detailed technical evaluation methodology for the AT model is documented in Section 8.4.7 of D5.5[3]
<u>SED (Sound event detection)</u>
SED component can identify specific sounds within brief time intervals, such as one second. These specific sounds are referred to as characteristic sounds, each of which can be described by a unique label, known as a sound event.
The algorithm achieved F1-score 50.08% (macro).
The detailed technical evaluation methodology for the SED model is documented in Section 8.4.6 of D5.5[3]
<u>AVAD (Audio-Visual Anomaly Detection)</u>
AVAD component uses a combination of audio and video data sequences to analyse a specific time frame for any anomalous events, providing a score which represents the likelihood of an anomaly happening in this time frame.
The detailed technical evaluation methodology for the AVAD model is documented in Section 8.4.2 of D5.5[3]
In parallel, MT staff checked the system stores video snippets of targeted events and that these are properly classified.
Finally, the functionality developed via SmartViz to validate the anomalies recorded by the system enables continuous improvement of the entire framework.
</td></tr>
<tr><td rowspan="2">MT-KPI6 Detection time reduction.</td><td>Processing time</td><td>Current situation</td><td></td><td>MT managers</td></tr>
<tr><td colspan="4">
Evaluation and Comments
MT staff investigated if the system could enable the Local Police to intervene in a shorter time than in the situation where an operator monitors the cameras without the help of the MARVEL framework.
Currently, the intervention time depends on various factors (number of available patrols, distance of the patrol from the scene of the event, etc.), but it occurs after receiving a request for intervention from the operational center and is estimated at around 20 minutes on average.
Thanks to the survey submitted to the Local Police of the Municipality of Trento, it emerged that the system could reduce intervention times:

less than 10% by 9,1% of the voters;
between 10% and 30% by 36,4 of the voters;
between 30% and 50% by 45,5% of the voters;
more than 50% by 9,1% of the voters.

</td></tr>
</table>

	Lastly, the total latency for the MT3 was around 40 seconds. After investigation, the bottleneck seems to be related to the re-streaming of MQTT messages between the two fog servers used to isolate the MARVEL from the MT infrastructure, affecting VAD, as the only inference component running at the edge in this scenario.

9.4.2 Use case specific non-functional KPIs

Table 23: Use case specific non functional evaluation variables for MT1: Monitoring of Parking Spaces (from Table 6.3 in D1.2)

MT3			
Evaluation variable	**How to measure**	**Internal evaluators**	**External Evaluators**
End-user experience	Periodic Surveys	MT staff	Local Police
	Evaluation and Comments The survey confirmed that the MARVEL framework is easy and helpful to use. The final user considers the features provided by the MARVEL framework, like the timeline distribution of vehicles, total number of vehicles, clustering of vehicles/events, and information on detections, contributing to the Local Police officer's ability to properly monitor the parking area: • very well by 18.2% of the voters; • well by 54,5% of the voters; • neutral by 18,2% of the voters; • not well by 9.1% of the voters. Additionally, the end users agree the summaries widget fulfil its purpose of providing a quick and easy overview of anomalies in the parking lot, supporting Local Police officers in their monitoring and corrective actions: • strongly agree by 36.4% of the voters; • agree by 54,5% of the voters; • disagree by 9.1% of the voters.		
Data protection, privacy preserving	Data breach reports	IT managers/infrastructure managers	DPO
	Evaluation and Comments No data breach was reported. MT staff and FBK staff have complied with all the precepts established by the MT DPO.		
Scalability/Modularity	Extend the solution to other cameras and microphones (other places of city surveillance system)	IT managers/infrastructure managers	Local Police
	Evaluation and Comments The MARVEL framework is highly modular: and scalable adding a sensor is possible without impacting the system. Obviously, adding a sensor implies ensuring that the processing capabilities are available either at the fog or cloud.		

9.4.3 Additional parameters common across all MT use cases from D6.1/D6.2

A subset of parameters suggested in D6.1 are common for all use-cases. Table 17 tabulates these parameters and their assessment, which have been updated since D6.2.

Table 17, in section 7.4.3, tabulates the parameters that are common for all MT use cases.

10 Use case MT4: Analysis of a Specific Area

10.1 Introduction

This section gathers all the information pertaining to use case MT4: Analysis of a Specific Area. The first two sections summarise the scope and description of the use case (section 10.2) and the components, including relevant datasets and the configuration of the framework (section 10.3). These sections provide a background to the use case evaluations reported in section 10.4.

10.2 Scope and description of use-case

The Municipality of Trento aims to enhance its decision-making process by monitoring key locations within the city. To achieve this, the MARVEL framework helps in counting people, cars, buses, taxis, and bikes, as well as calculating their trajectories and identifying noteworthy events during specific timeframes or throughout the day to facilitate effective decision-making.

The identified area of interest is the vicinity of the Trento train station, encompassing the road and a portion of Piazza Dante, spanning from the traffic lights on Via Dogana to the traffic lights on Via Pozzo. This scenario is likely to be integrated into the larger project called the "Smart City Control Room," which is being launched in the Municipality of Trento. This initiative aims to gather the necessary data for formulating and monitoring sustainable mobility plans and energy transition actions in the urban area.

The audio-video analysis will be performed in real-time and on recorded data stored on the servers of the Local Police.

The data collected in this use case allow to:

- by the side of Local Police, analyse the monitored area through the timeline distribution of the anomalies and events, the clustering of anomalies and/or events, the severity, type and information of anomalies observed, and the statistics of the anomalies and/or events. Once the Local Police reviews the video or audio snippet of a detected anomaly/event, they have the option to mark this as anomalous or not and report on whether police intervened or not. These options allow for continuously improving the system;
- on the side of policy-maker/mobility manager: investigate the timeline distribution of the vehicles, the total number of vehicles, the statistics of vehicles and/or events, the trajectories by traffic entities, and the total number of persons. This enhanced efficiency will contribute to more effective traffic management and urban planning processes.

Two evaluation scenarios have been defined for this use case related to two use case KPIs as reported in Table 24.

The MARVEL solution improves the local territory, the city surveillance, the quality of life of the citizens, their perceived security, and the experience of tourists.

Table 24: Evaluation scenarios and relevant KPIs for the MT4

Evaluation Scenario	Target	Relevant KPI
The evaluation scenario involves observing the timeline distribution of the vehicles, the total number of vehicles, the statistics of vehicles and/or events, the trajectories by traffic entities, the total number of persons and other security-related events so the Municipality can increase the	The aim is to recognise 50% of mobility patterns and recurrent dangerous events detected	**MT-KPI7: collection of trajectories and snippets of anomalous events related to mobility as well as other security-related events**

management of traffic and security near the train station		
The evaluation scenario includes testing the searchable database that will provide insights enabling long-term decision-making support for public authorities	The goal is the reduction of travel time through the city centre and the decrease in urban and mobility planning time	**MT-KPI8: Increased efficiency in the urban planning**

10.3 Summary of Technical Setup and Decision Making Toolkit

The MT4 infrastructure consists of four IP cameras transmitting video, two Raspberry Pis devices transmitting audio data collected from the connected IFAG microphones, and two workstations at the fog layer, managed by FBK.

The MT4 edge layer infrastructure consists of four IP cameras transmitting video and two Raspberry Pi with a connected IFAG microphone. The Raspberry Pi runs a voice activity detector and a speech anonymiser and streams the anonymised audio via RTSP. The upload function is a secure transmission by VPN access between MT and FBK in which raw data will be sent to the data lake in FBK.

FBK provides the Fog tier for the MT use cases. To comply with the constraints of the agreement between MT and FBK (FBK was nominated as data processor) that granted FBK access to the raw data of the MT's sensors and to satisfy the requirements of the MARVdash Kubernetes cluster, FBK deploys two workstations, both with GPU.

MT has provided data for various components. MT has provided the dataset "TrentoOutdoor – real recording" and "TrentoOutdoor – staged recording" (as defined in D2.1) processing the streams of cameras and microphones from Piazza Dante.

MT, in collaboration with FBK, collected, anonymised, and annotated more than 180 audio and video files, used for the training of SED and AVAD (detection of audio-video anomaly situations).

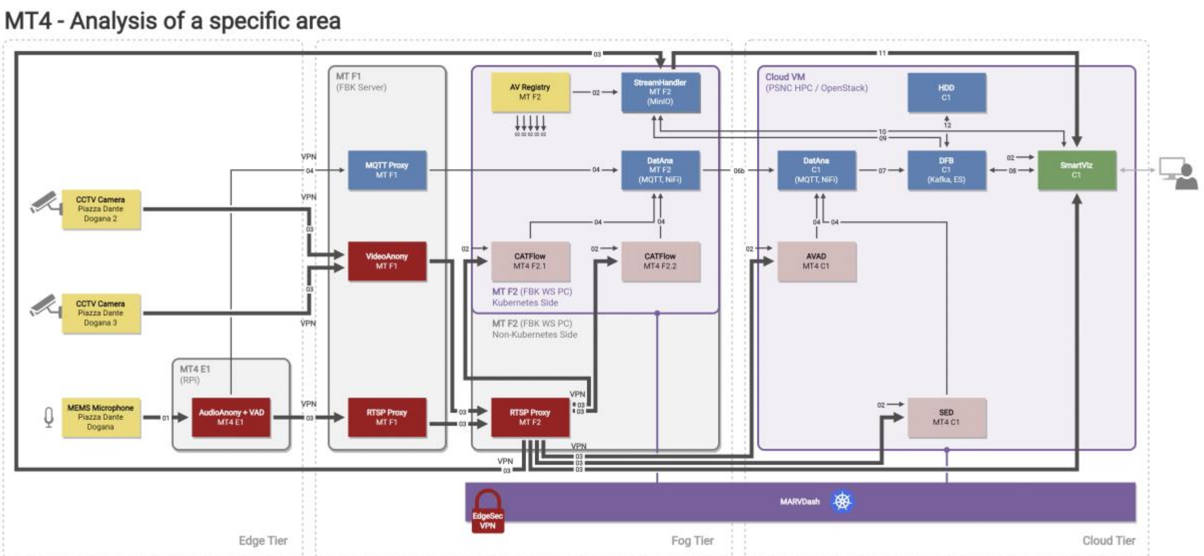


Figure 17: MARVEL RP2 deployment and runtime view of the MARVEL architecture for MT4 (source: D5.6)

SmartViz, the User Interface of the platform, supported several functionalities: 1) advanced visualisations, with a dashboard for user configurations and user interactions, 2) live AV feed for real-time inspection of the monitored areas, 3) on-request inspection of stored AV data, and 4) user-based verification of AI inference results.

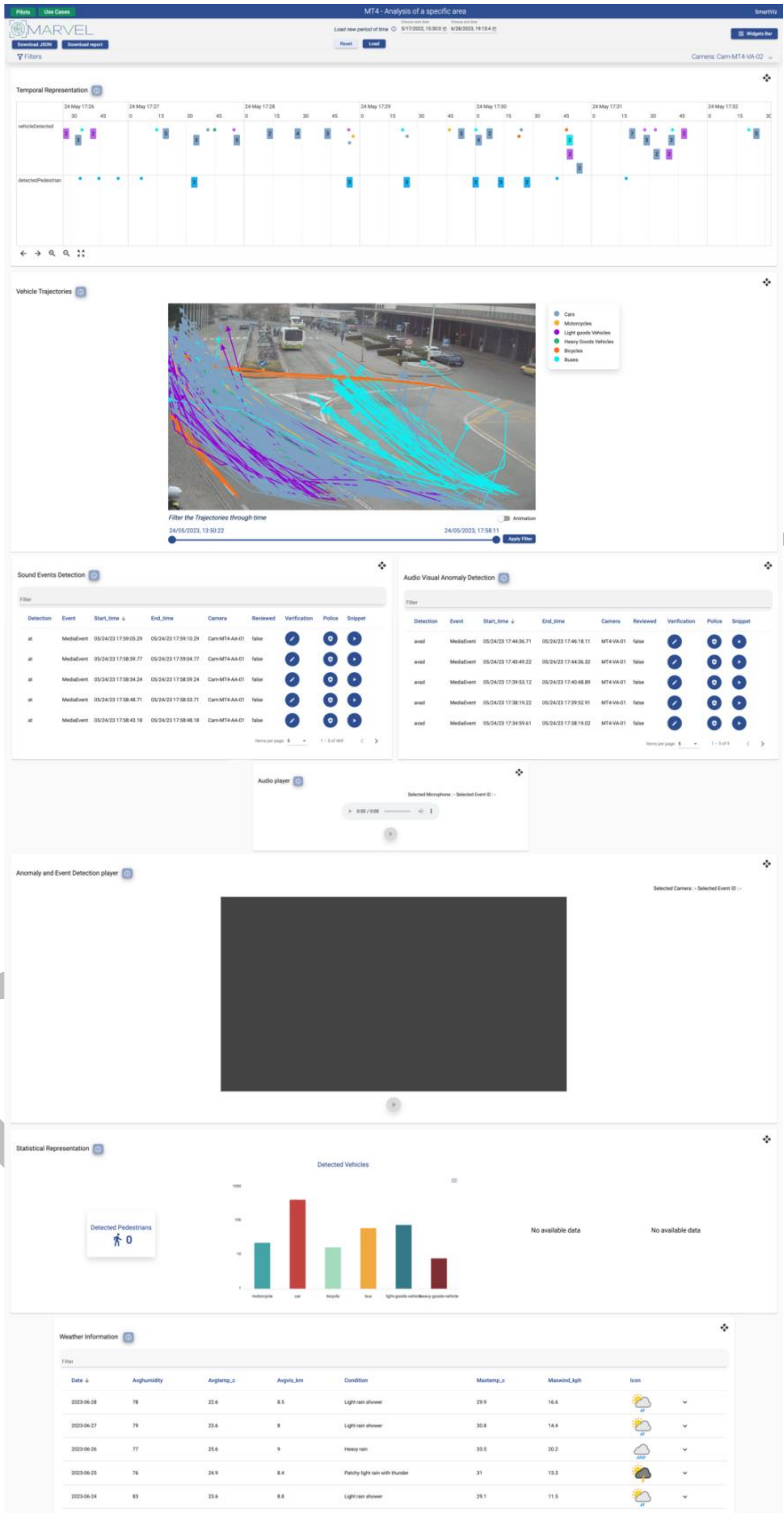


Figure 18: SmartViz dashboard for MT4

The detections are visualised across time in order to be monitored and clustered using the temporal representation widget. To illustrate the trajectories of the detected vehicles and detailed information, we used the Vehicle Trajectories widget. In this widget, the paths of the passing vehicles that are detected by the CATflow component are drawn in the image of the camera feed that is recording them. The paths are grouped and colour-coded depicting different types of vehicles. The users are also able to change the time period and filter by the type of vehicle or lane to further investigate.

In the Statistics widget, available information regarding the AI models detections are visualised to give the users an analysis of the occurrences.

Furthermore, detailed available information about the detected incoming events is represented in a tabular form in the Details Widget. This widget also supports a standalone text filtering capability in order to facilitate a rapid search through the available information.

Within the audio and video player functionality in SmartViz, users can select an event and play the corresponding audio or video snippet. The StreamHandler component retrieves the stream for the relevant time period of the detection, segments it, and plays the snippet in the Audio or Video player widget within SmartViz.

Users have the ability to validate the inference results by marking them as accurate or not. The police intervention functionality is also included in the use case. After reviewing the events, users can mark an event as important and in need of police intervention, facilitating appropriate action.

In addition to rearranging and resizing widgets, the dashboard view offers the functionality to download the visualised data in JSON format. Furthermore, users have the option to save the entire dashboard as a PDF file, allowing for offline access and convenient sharing based on individual preferences.

Finally, the Weather information widget in the use case provides users with a representation of weather-related data, allowing them to view and explore weather information for a selected time period.

10.4 KPI assessment

This section evaluates the use case against a set of functional KPIs (Table 25) and a set of non-functional parameters (Table 26). These tables are sourced from D1.2[2], D6.1[5] and D6.3[6] and the evaluations are specific to the use case. Section 14 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs, Business KPIs and societal goals benchmarks.

10.4.1 Use case specific KPIs

Table 25: Use case specific KPIs for MT4: Analysis of a Specific Area (From D1.2)

MT4					
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**	**Results and Comments**
MT-KPI7: collection of trajectories and snippets of anomalous events related to mobility as well as other	Detection of habits	Current situation	50% of mobility patterns and recurrent dangerous events detected	MT managers / Local Police	
	Evaluation and Comments The local police operations centre currently employs 2 people who monitor 700 surveillance cameras from 7:00 a.m. to 7:00 p.m. The monitoring occurs retrospectively, following a dangerous situation. The operators inspect the images saved by the system, attempting to				

<table>
<tr><td>security-related events</td><td colspan="5">identify the footage of the incident for analysis. In only rare cases, when a demonstration or authorized public event is scheduled in advance, operators dedicate themselves to real-time monitoring of the cameras assigned to monitor the areas affected by the event, seeking to identify potential danger. In these instances, the number of surveillance network cameras is a subset (on average 10-20) monitored by the operations center operators. Even in this scenario, identifying a dangerous situation proves to be particularly challenging and it occurs in the majority of cases retrospectively.

Only in a residual portion of cases, quantifiable at about 5%, can operators identify a dangerous situation if they don't know the location, date, and time of the event (and therefore only retroactively)

On the other hand, the Mobility Department of the Municipality of Trento does not possess real-time data of the traffic situation nor time series to carry out punctual analyses or monitor implemented mobility plans.

This use case makes use of a set of AI models namely, CATFLow (Vision based Object detection), SED (Sound event detection) and AVAD (Audio-Visual Anomaly Detection), to collect anomaly situations and presence of people in the square.

<u>CATFlow (Object Detection and Tracking)</u>

The model mAp (mean Average Precision across all classes) the detector achieved is 0.693, for a confidence threshold of 0.7. The mAp increases to 0.881 for a confidence threshold of 0.5. The mAp considers the overlap of the bounding boxes with the ground truth in the test set. The MOTA (Multi-Object Tracking Accuracy) achieved by the tracker is 54.33%. More importantly for this use case, 84.03% of trajectories averaged over various road scenes were correctly detected and the average count accuracy is 90.82%.

The detailed technical evaluation methodology for the CATFlow model is documented in Section 8.1.2 of D5.5[3]

<u>SED (Sound event detection)</u>

SED component can identify specific sounds within brief time intervals, such as one second. These specific sounds are referred to as characteristic sounds, each of which can be described by a unique label, known as a sound event.

The algorithm achieved F1-score 35.01% (macro).

The detailed technical evaluation methodology for the SED model is documented in Section 8.4.6 of D5.5[3]

<u>AVAD (Audio-Visual Anomaly Detection)</u>

AVAD component uses a combination of audio and video data sequences to analyse a specific time frame for any anomalous events, providing a score which represents the likelihood of an anomaly happening in this time frame.

The detailed technical evaluation methodology for the AVAD model is documented in Section 8.4.2 of D5.5[3]

In parallel, MT staff checked the system stores video snippets of targeted events and that these are properly classified.

Finally, the functionality developed via SmartViz to validate the anomalies recorded by the system enables continuous improvement of the entire framework.</td></tr>
<tr><td rowspan="2">MT-KPI8: Increased efficiency in the urban planning</td><td>Automatic detection of patterns</td><td>Single person observing multiple cameras</td><td>Travel time reduction Decrease in planning time</td><td>MT managers</td><td></td></tr>
<tr><td colspan="5">Evaluation and Comments

MT staff investigated if the system could enable the Local Police to intervene in a shorter time than in the situation where an operator monitors the cameras without the help of the MARVEL framework.</td></tr>
</table>

<table>
<tr><td></td><td>Currently, the intervention time depends on various factors (number of available patrols, distance of the patrol from the scene of the event, etc.), but it occurs after receiving a request for intervention from the operational center and is estimated at around 20 minutes on average.

Thanks to the survey submitted to the Local Police and the Mobility Office of the Municipality of Trento, it emerged that the system could reduce time in planning urban security plans and mobility plans:
• neutral by 8,3% of the voters;
• agree by 58,3 of the voters;
• strongly agree by 33,3% of the voters.

Lastly, the total latency for the MT4 was around 6 seconds. After investigation, the bottleneck seems to be related to the re-streaming of MQTT messages between the two fog servers used to isolate the MARVEL from the MT infrastructure, affecting VAD, as the only inference component running at the edge in this scenario.</td></tr>
</table>

10.4.2 Use case specific non-functional KPIs

Table 26: Use case specific non functional evaluation variables for MT4: Analysis of a Specific Area (from D1.2)

<table>
<tr><th>MT4</th><th></th><th></th><th></th><th></th></tr>
<tr><th>Use case</th><th>Evaluation variable</th><th>How to measure</th><th>Internal evaluators</th><th>External Evaluators</th></tr>
<tr><td rowspan="4">MT4 Analysis of a Specific Area</td><td>End user experience</td><td>Periodic Surveys</td><td>MT staff</td><td>Local Police</td></tr>
<tr><td colspan="4">Evaluation and Comments
The survey confirmed that the MARVEL framework is easy and helpful to use.
The end users judged that the system helps achieve the aims and objectives of anti-social behaviour monitoring, as well as supporting policy-makers and mobility managers in monitoring the number of people, cars, buses and bicycles for efficient traffic management and urban planning:
• very effectively by 25% of the voters;
• effectively by 41,7% of the voters;
• neutral by 8,3% of the voters;
• ineffectively by 8.3% of the voters;
• very ineffectively by 16.7% of the voters.
Moreover, the end users rated, with regard to the Vehicle Trajectories widget, in terms of illustrating vehicle paths, adjusting time periods, and filtering by vehicle type to facilitate detailed investigations in the context of monitoring antisocial behaviors and traffic-related activities:
• very valuable by 50% of the voters;
• valuable by 33,3% of the voters;
• neutral by 16,7% of the voters.</td></tr>
<tr><td>Data protection, privacy preserving</td><td>Data breach reports</td><td>IT managers/infrastructure managers</td><td>DPO</td></tr>
<tr><td colspan="4">Evaluation and Comments
No data breach was reported.
MT staff and FBK staff have complied with all the precepts established by the MT DPO</td></tr>
</table>

	Scalability/Modularity	Extend the solution to other cameras and microphones (other places of city surveillance system)	IT managers/infrastructure managers	Local Police
	Evaluation and Comments The MARVEL framework is highly modular: and scalable adding a sensor is possible without impacting the system. Obviously, adding a sensor implies ensuring that the processing capabilities are available either at the fog or cloud.			

10.4.3 Additional parameters common across all MT use cases from D6.1/D6.2

A subset of parameters suggested in D6.1 are common for all use-cases. Table 17 tabulates these parameters and their assessment, which have been updated since D6.2.

Table 17, in Section 7.4.3, tabulates the parameters that are common for all MT use cases.

11 Use case UNS1: Drone Experiment

11.1 Introduction

This section gathers all the information pertaining to use case UNS1: Drone Experiment, which was implemented during RP1, and following feedback updated during RP2. The first two sections summarise the scope and description of the use case (Section 11.2) and the components, including relevant datasets and the configuration of the framework (Section 11.3). These sections provide a background to the use case evaluations reported in Section 11.4.

11.2 Scope and description of use case

The aim of the UNS1 use case is to assess the potential use of drones in monitoring large public events held in open spaces. Monitoring and surveillance at large public events can pose challenges due to insufficient infrastructure and the occasionally unpredictable nature of crowds. Fixed street cameras may offer frontal views of the crowd, but they are not able to accurately capture finer details. Overcrowded places are potentially dangerous zones because in the event of an accident, emergency services cannot respond quickly, and people can panic. This motivated us to place visual crowd counting in the core of the UNS1 use case. Additionally, there are areas or sections in the vicinity of large public events that remain unmonitored due to the lack of fixed cameras and unpredicted presence of people. The presence of people in such areas could be quickly detected using the IFAG MEMS microphones and VAD component while a drone could be sent for inspection, which is supported using defined architecture.

UNS members had several meetings with the organisers of the EXIT festival, which is one of the largest open-space music events in Europe, held each summer in Novi Sad. Their feedback was used as guidance for upgrading evaluation scenarios, as they are a potential end-user. As a result, crowd counting was identified as the most desirable task to work on and their feedback was used while performing staged recordings.

Experiments within the UNS1 use case were done in the controlled environments and additional goal besides evaluation MARVEL technologies using drones, was to support partners' use cases for the development and validation of the federated learning framework for crowd counting and onboard real-time processing.

Table 27: Evaluation scenario and relevant KPIs for the UNS1

Evaluation Scenario	Target	Relevant KPI
Detection of anomalous events and alerting event organisers about them,	The goal is to achieve a detection accuracy of at least 80% and to incorporate system for alerting.	**UNS-KPI1** **Achieve high precision of anomaly detection using VCC (in terms of overcrowding)**
System performance evaluation against, e.g., accuracy for varying operating system conditions	The goal is to achieve at most MAE = 2 performance degradation comparing to the nominal settings	**UNS-KPI2** **Robustness to different operating conditions (e.g., distance, camera resolution,)**
Monitoring (e.g., by security crew) of streaming data.	The goal is to achieve processing time less than 2 seconds for counting the number of people and rising an alarm and less than 15 seconds	**UNS-KPI3** **Achieve acceptable time delay incurred when identifying the number of people on the scene.**

Evaluation Scenario	Target	Relevant KPI
	to provide automatic visualisation of crowds using heat maps.	
System performance evaluation with presence of both audio and visual modality and modality dropout.	The goal is to achieve 10% improvement comparing to the audio-based only system operation.	**UNS-KPI4 Multimodality: Detection of the presence of people in the scene via multimodal analysis**

Evaluation scenarios from Table 27 remain the same as in the D6.1, but description of relevant KPIs and targets were changed accordingly to the lessons learned and the use case updates reported in D6.2 and D6.3. More details could be found in Section 11.4.1 of D6.4.

11.3 Summary of Technical Setup and Decision Making Toolkit

The UNS inference pipeline consists of 3 layers – Edge, Fog and Cloud, as shown in Figure 19. Data capturing is performed at the edge using the AVDrone component. The sensors consist of a GoPro camera mounted on the drone and ground-based IFAG AudioHub MEMS microphones. The GoPro camera is connected to an Intel NUC, whereas microphones are connected to Raspberry Pi v4 boards on the ground.

Video anonymisation and audio anonymisation are enabled by installing VideoAnony on the same Intel NUC and AudioAnony on the same Raspberry Pi boards, respectively. Data streaming is further enabled via RTSP. AudioAnony is coupled with VAD within the same container, whose purpose is to detect voice activity boundaries, which are further provided to AudioAnony for controlling the activation of the anonymisation process. The outputs of VAD are submitted to MQTT DatAna brokers, also residing at the Raspberry Pi, to be further utilised in the platform (e.g., for audio content inspection by SmartViz, Figure 20). The anonymised data is further transmitted to the fog tier where AI models are executed.

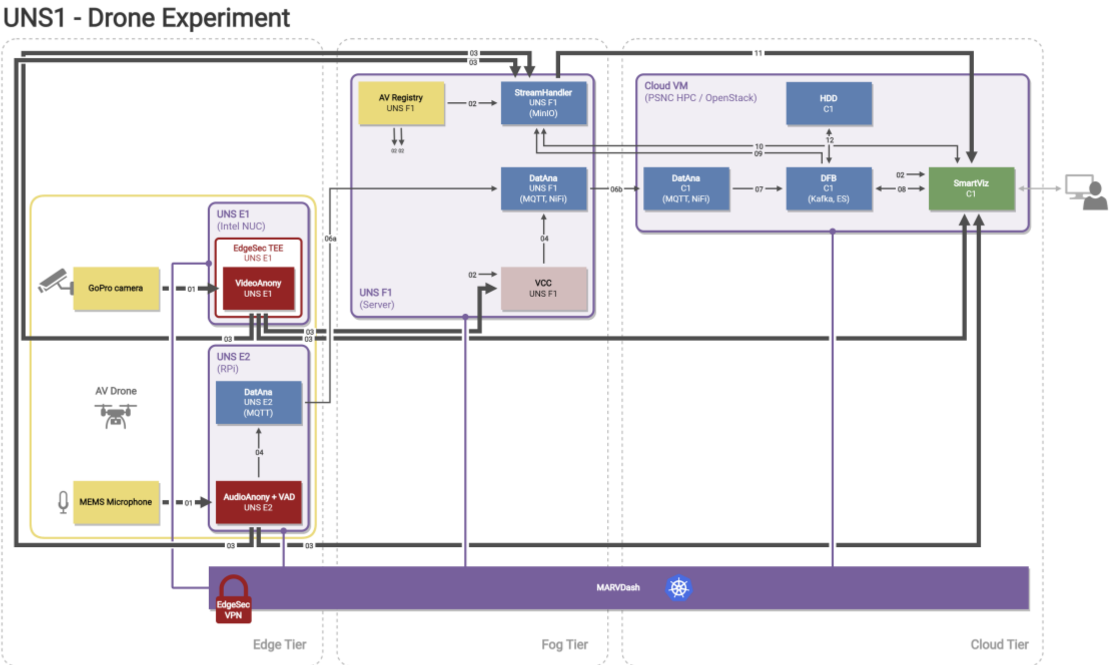


Figure 19: MARVEL RP2 deployment and runtime view of the MARVEL architecture for UNS1 – Drone Experiment (Source: D5.6)

VCC is chosen as one of the AI components of interest within RP1. Besides VCC, FedL-based VCC model training is implemented at the Fog. VCC-FedL client runs on a GPU installed on the UNS Fog server and it also communicates with the VCC-FedL server that runs on the cloud (PSNC HPC via OpenStack). The other components at the fog tier include DatAna Fog (MQTT, NiFi), AV Registry and StreamHandler (MinIO). At the cloud tier the following components are implemented: DatAna Cloud, DFB, HDD, SmartViz, and DataCorpus.

Anonymised data is transferred from the edge tier to the fog tier, to perform inference using VCC and further to inspect streams using SmartViz at the cloud tier. These two components and StreamHandler request metadata details of the AudioAnony and VideoAnony within the initialisation phase from AV Registry, deployed at a server. If the inference output of the VCC component is higher than the set threshold, an alarm is raised so that a human operator can observe the snippet of interest in the stream.

For AI component training, UNS provided the “UNS Drone dataset”. It consists of 55 minutes of raw video snippets, making a total of 11.4GB of video data. Data is recorded using HD resolution, 30fps and H.264 standard. Raw video data is further annotated for visual crowd counting and 819 annotated frames are provided. Annotations were made using the CVAT tool and they are exported in xml format. Besides video data, audio data is recorded using IFAG AudioHub Nano microphones. In RP2, UNS has released audio datasets that complements video frames annotated within UNS Drone dataset. The name of the dataset is “UNS1_VAD_dataset” and it contains about 1 GB of audio data samples. This way, audio-visual dataset is fully prepared. Audio-visual data were recorded within the staged recording process The above-mentioned dataset features snippets capturing variations in several aspects: distances, camera angles and number of people in the scene. The data were acquired during UNS staged recordings in M16 at the Petrovaradin fortress, where UNS staff and students were gathered to simulate crowd movements and relevant crowd events. This dataset contributes to the *UNS drone dataset* described in detail in D2.1 (Section 4.2.6) and also D1.2 (Section 4.6.1) and D8.2. We note that all experimental participants have signed written consent to participate in the staged recordings.

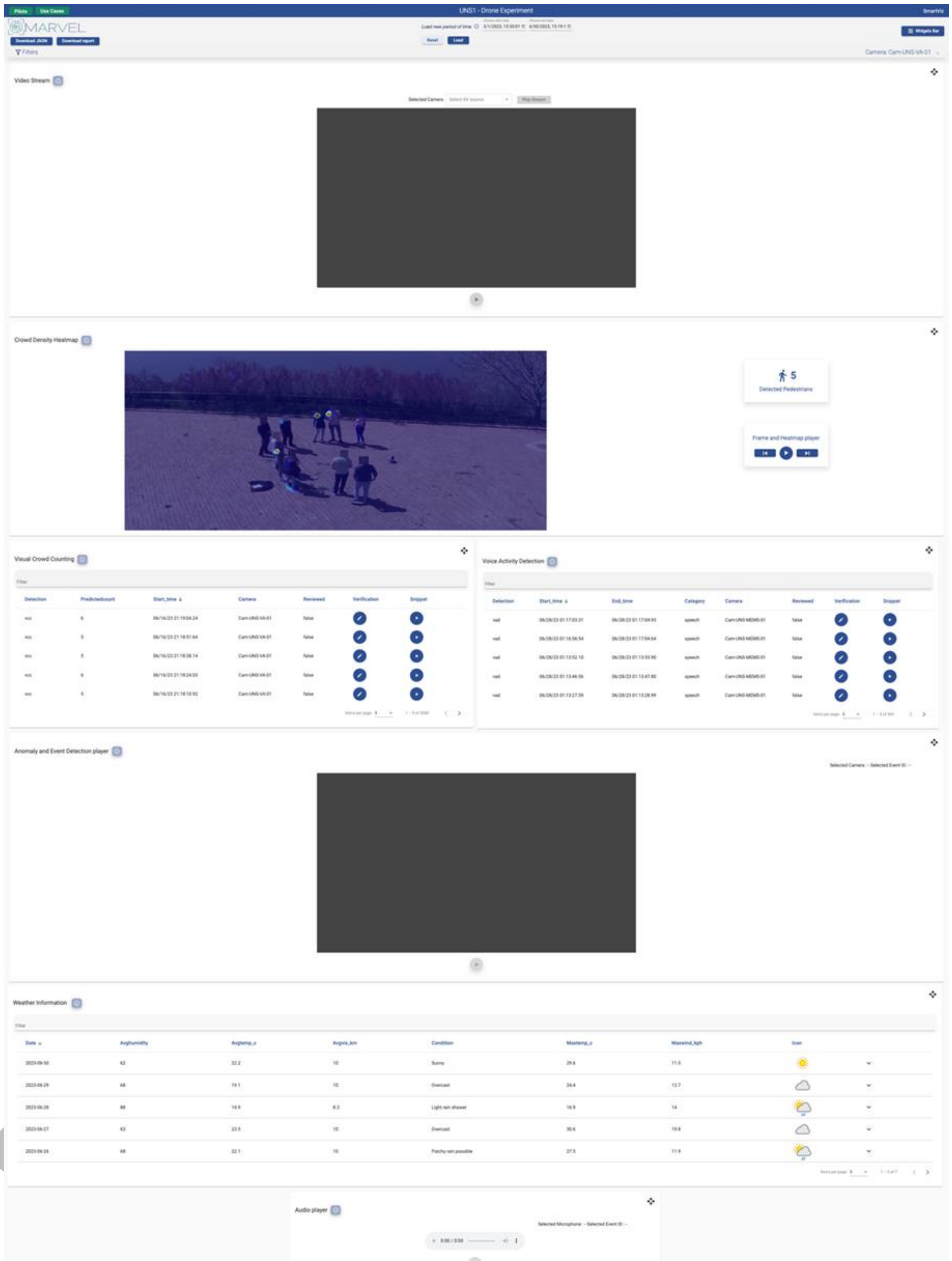


Figure 20: SmartViz dashboard for UNS1 (Source: D4.6)

11.4 KPI assessment

This section evaluates the use case against a set of functional KPIs (Table 28) and a set of non-functional parameters (Table 29). These tables are sourced from D1.2[2], D6.1[5], D6.2[1], and D6.3[6] and the evaluations are specific to the use case. Section 14 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs.

11.4.1 Use case specific KPIs

This section tabulates the progress in evaluating the use case against a set of functional KPIs (described in Table 27 and evaluated in Table 28) and a set of non-functional parameters (Table 29). The performance of VCC and VAD models was evaluated on a set of specific scenarios, as well as processing time required to raise an alert and to visualise heat maps on SmartViz. Here we also redefine functional KPIs in order to better align with the use case measurable objectives, according to the light modifications of the use case throughout the project. The progress of the initial version of the KPIs (defined in D1.2) was reported in D6.2. In this revised version, all ideas of initial KPIs from D1.2 are kept (i.e. to focus on anomaly detection, robustness, processing time, and multimodal analysis), but they adapted to measurable goals of the use case.

Initially, the UNS1 use case was dedicated on defining scenarios of anomalous events in terms of observing problematic behaviour of people within the crowds. After organising meetings with the organisers of the EXIT festival in Novi Sad, the focus was shifted to the detection of overcrowding as anomaly, as this was more interesting to the organisers of large public events. Consequently, in the UNS-KPI1 instead of monitoring the average accuracy for the drone-based audio-visual anomaly detection, we were focused on the anomaly detection in terms of overcrowding, which was monitored using VCC component, that is a vison only case. The audio modality was processed using VAD in order to detect the presence of people and it could not be used in a such form for detecting overcrowding. After recording a dataset within a staged recording process and performing a series of experiments, we achieved an average accuracy value of 91.6% for anomaly detection in terms of overcrowding.

In the UNS-KPI2, among options that we have proposed in the previous phases of the project, we have selected camera resolution and distance as two most relevant robustness measures. Frames per second as an operating condition was dropped out as the VCC is a frame-based model, i.e., it does not infer on a sequence of frames, whereas dropout of audio modality does not lead directly to the change of AI model performance, but to the time required for detecting a suspicious event, for which we need to send a drone for visual inspection. As UNS is a small experimental pilot to test MARVEL technologies in controlled conditions within a staged recording process, it was not appropriate to make an experiment that contains measurements on a large area without audio recording. Performed experiments demonstrated that in the case of camera resolution of 640x360 pixels, the system performance was identical as in the case of nominal settings (HD resolution). In the case of the monitoring from different distances, we have concluded that there exists minor performance degradation in the case of recording from a higher position. We have recorded data from 20m and 30m height, and the average degradation is MAE = 0.72.

In the case of the UNS-KPI3, we had to slightly redefine the metric in the part related to the baseline. Similarly, as in the case of the UNS-KPI2, UNS as a small experimental pilot could easily perform an experiment involving security crew members and measuring time for their detection of observed events. Moreover, such kind of human reactions directly depend on the total number of video streams they observe, meaning that in some kind of a real-world scenario with a big number of screens, their reaction time could be fast. Within the benchmarking process, reported in D5.5, it was measured that the total processing time of the whole MARVEL architecture for the UNS1 use case, including heatmap visualisation is less than 10 seconds, whereas time required to detect number of people in the scene and to potentially rise an alarm is less than 1 second, which is expected to be much faster than continuous human reactions while monitoring several screens.

In the case of the UNS-KPI4, we were focused on the multimodality. Considering initial description of the KPI, we can confirm that it would be satisfied, as we have successfully implemented audio and visual modalities. However, performance comparison in the initial version of the KPI would be difficult to measure, as VAD component leads to the detection of presence of people, and after that to the sending drone for inspection, and not to the detection of number of people in the scene. In the redefined version, we have defined an audio-only system operation as a baseline. The achieved detection accuracy of the presence of people using audio-only system operation was 76%, whereas video modality provided detection accuracy equal to 100% on the observed test set.

Table 28: Use case specific KPIs for UNS1: Drone Experiment

UNS1				
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**
UNS-KPI1 Achieve high precision of anomaly detection using VCC (in terms of overcrowding)	Classification accuracy;	No baseline	Achieve classification accuracy at least 80%	UNS staff
	Evaluation and Comments In the described UNS1 use case, we define anomaly detection as a potential overcrowding scenario, i.e., as a case if a detected number of people in the scene is larger than some threshold. In the test set, the total number of people in the scene can be between 1 and 12. To evaluate the performance of the KPI, we performed an anomaly detection experiment for all possible threshold values (from 1 to 12). As the VCC model performs human detection with a mean absolute error (MAE) equal to 1.27 on the test set (will be further described within performance evaluation of the UNS – KPI2), we have included this value as an upgrade to the anomaly detection setup, defining an anomaly to correspond to the scenario when the output of the VCC is larger or equal to the value of observed threshold reduced by the MAE value. Achieved results are in the range between 76.7% and 100% for a range of possible thresholds, while the average accuracy value is 91.6%.			
UNS-KPI2 Robustness to different operating conditions (e.g., distance, camera resolution,)	Accuracy (distance) Accuracy (camera resolution)	nominal settings	The average MAE degradation not more than 1 comparing to the nominal settings.	UNS staff
	Evaluation and Comments We have performed two experiments. • In the case of measuring accuracy for different camera resolutions, nominal settings include HD resolution and footage from different distances (whole UNS1 video dataset). The achieved performance in such a nominal setting is MAE = 1.27. For comparison purposes, original HD videos are downsampled to the 640x360 pixels resolution. Identical MAE value was obtained in this case, so that there is not any performance degradation. • In the case of measuring accuracy for different distances, nominal settings include footage from a height of 20 meters. In this case, achieved performance is MAE = 1.08. The second measurement is performed taking into account videos recorded from 30 meters height. The achieved performance is MAE = 1.80. Degradation of the performance compared to the nominal settings is equal to 0.72.			
UNS-KPI3 Achieve acceptable time delay incurred when identifying the number of	Processing time	No baseline	Less than 2 seconds for counting the number of people and rising an alarm and less	UNS staff

people on the scene.			than 15 seconds to provide automatic visualisation of crowds using heat maps.	
	Evaluation and Comments SmartViz supports the raising of an automatic alarm if there is a larger number of people in the crowds than a pre-defined threshold. This way, a human operator can focus on anomalies in the crowds, whereas inspection of several data streams at the same time would be much more difficult without the help of a developed framework. The achieved overall global latency within benchmarking experiments is 9.96 seconds, which includes the total processing time to visualise crowds using heatmaps on the screen. Most of the delay was produced while managing heatmap filed in the data stream managed by Kafka via DFB (9.19 seconds). The total processing time required to process data to count people and send an alert without visualisation is 0.77 seconds, as reported in D5.5.			
UNS-KPI4 Detection of the presence of people in the scene via multimodal analysis	Classification accuracy	Audio-only system operation	The goal is to achieve 10% improvement compared to the audio-based only system operation	UNS staff
	Evaluation and Comments In the event of video modality dropout, VAD utilised to detect human presence. The achieved detection of the presence of people in the case of automatic visual inspection using the VCC component on the UNS1 test set is 100%. The achieved detection using audio only system operation as a baseline is 76%.			

11.4.2 Use case specific non-functional KPIs

The detailed evaluations for the UNS1 non-functional KPIs, tabulated in Table 29, are mostly obtained from surveys (Appendix Section 17.3.1). Surveys consisted of two major parts. The first one related to the Use case architecture and functionalities and the second one related to the UI of the Decision making toolkit and user experience.

Respondents emphasised that the presented demonstration of the crowd counting technology within UNS1 seems accurate (80% of respondents selected "agree" or "strongly agree" option, whereas 20% were "neutral"). The same share of the respondents provided positive feedback regarding the usefulness of the incorporated automatic to improve prevention of overcrowding at the covered areas at large public events. Positive feedback was obtained regarding the possible further adaptation of the designed setup by incorporating new equipment, which could be very important for the potential long-term end users. Implementation of audio and video anonymisation, as well as EdgeSec TEE component to support the encryption of the network traffic as well as file encryption in a transparent way, was evaluated positively by 80% of respondents and neutral by 20%, i.e., they agreed that the UNS1 supports secure transmission.

The largest differences were obtained regarding the effectiveness of the implemented multimodal setup, which incorporates VAD and VCC tools. The share of 40% of respondents selected an option that the proposed setup facilitates crowd counting management in an effective way, whereas 40% of respondents think that although the idea is reasonable, implementation could be difficult and detailed planning is needed to cover all areas using VAD. In the end, 20% of respondents selected an option that the idea could have only a limited number of applications as automatic monitoring in different conditions and areas could be challenging and there is a need of engaging security crews.

SmartViz tool was used for visualisation within the UNS1 use case. In general, all provided feedback regarding the SmartViz widgets is positive, with different shares of selected "agree" and "strongly agree" options. The respondents were asked to evaluate heatmap representation widget, the ability to replay the consecutive created heatmaps from a selected time period, and the Video Stream Player widget. More details can be found in the appendix, Section 17.3.1.

Table 29: Use case specific non functional evaluation variables for UNS1: Drone Experiment (6.3 in D1.2)

<table>
<tr><th>UNS1</th><th></th><th></th><th></th></tr>
<tr><th>Evaluation variable</th><th>How to measure</th><th>Internal evaluators</th><th>External Evaluators</th></tr>
<tr><td rowspan="2">Modularity</td><td>Integration of new equipment</td><td>Infrastructure managers</td><td>IT experts of security crews</td></tr>
<tr><td colspan="3">Evaluation and Comments
Modularity of the framework is supported two-fold: in terms of computing devices and in terms of data capturing devices.
With regards to the computing devices, modularity is supported through MARVdash components wherein adding a new device is achieved by installing appropriate Kubernetes tools and by associating the device to the relevant Kubernetes cluster.
Modularity in terms of sensing devices is achieved by updating the AVRegistry with appropriate device metadata and withdrawing this information from AI components that should process the corresponding data streams. The integration process was monitored and evaluated by internal evaluators during the whole length of the project.
Regarding modularity, 80% of external evaluators agreed that the UNS1 setup can seamlessly be adapted for the usage of new equipment for audio-visual monitoring.</td></tr>
<tr><td rowspan="2">Secure transmission</td><td>Periodic evaluations</td><td>UNS staff</td><td>IT experts of security crews</td></tr>
<tr><td colspan="3">Evaluation and Comments
EdgeSec Trusted Execution Environment (EdgeSec TEE) component was incorporated within the UNS1 use case. More details about implementation are provided in subsection 2.3.2 (D4.5), while detailed description of the component is provided in subsection 3.2.3 (D5.5). Besides this component, anonymisation of audio and video streams is incorporated at the edge layer. Such setup was evaluated by 80% of external evaluators as one that supports secure transmission. The integration process was monitored and periodically evaluated by internal evaluators during the whole length of the project.</td></tr>
<tr><td rowspan="2">End-user experience</td><td>Periodic surveys</td><td>UNS staff</td><td>Security crew of public events organisers (e.g., city authorities)</td></tr>
<tr><td colspan="3">Evaluation and Comments
Within the UNS Drone Experiment, SmartViz provides a user-friendly interface. For this use case, SmartViz provides three main functionalities:
• to raise an automatic alarm;
• to play live AV data feed;
• to request historical AV data, to check how the crowd gathered.
An automatic alarm would be triggered if there is a larger number of people in the crowds than a pre-defined threshold. This is supported using VCC component in the backend. Besides VCC, VAD is also supported in SmartViz.
This way, a human operator can focus on anomalies in the crowds. Inspection of several data streams would be much more difficult without the help of developed framework. UNS academic staff was performing evaluation of the results from the beginning of SmartViz integration. All external evaluators provided positive feedback regarding questions related to the demonstrated SmartViz widgets within the evaluation process.</td></tr>
</table>

11.4.3 Additional parameters common across all UNS use cases

This section tabulates (Table 30) the progress in evaluating the use case against a set of general parameters. Most of these criteria are dependent on the use case KPIs (Table 26) and the asset specific KPIs (Table 25).

Table 30: Parameters that are common across all UNS use cases

<table>
<tr><th>Parameter</th><th>How to measure</th><th>Target to be achieved</th></tr>
<tr><td rowspan="2">Efficiency
Related to the efficiency of the system as used in the use case</td><td>Efficiency is largely dependent on how long it takes to detect events. The average time taken to detect and flag the event is measured.</td><td>To achieve the time needed to identify an anomalous event in terms of overcrowding for a short period of time (less than 2 seconds to raise an alarm and less than 15 seconds for visualisation).</td></tr>
<tr><td colspan="2">Evaluation and Comments:
Human operators might not notice relevant events by observing several streams, whereas system performs automatic detection of anomalous events (for RP1 anomalous event is number of people higher than a threshold). The processing time was achieved within the benchmarking process and it was reported in D5.5. The achieved overall global latency within benchmarking experiments is 9.96 seconds, which includes the total processing time to visualise crowds using heatmaps on the screen. Most of the delay was produced while managing the heatmap filed in the data stream managed by Kafka via DFB (9.19 seconds). The total processing time required to process data to count people and send an alert without visualisation is 0.77 seconds. Similarly, the total processing time to visualize the output in the case of the sound event localisation and detection 11.5 seconds, from which 10.2 seconds delay was produces while managing data through DFB. The average time to detect the event, without visualization is 1.3 seconds.</td></tr>
<tr><td rowspan="2">Operability
Related to the ability of the components to keep functioning together</td><td>Add additional multi-modal input sources.</td><td>Different data modalities successfully accommodated in the platform (audio, video, GPS, etc.).</td></tr>
<tr><td colspan="2">Evaluation and Comments:
Audio and video input data sources are successfully implemented and both data sources are supported currently.</td></tr>
<tr><td rowspan="2">Usability
Related to how well the system helps the users to achieve a task in a given use case</td><td>Interview security crew of public event organisers to measure the end-user experience via periodic surveys.</td><td>End-user finds the system easy to use.</td></tr>
<tr><td colspan="2">Evaluation and Comments:
Within UNS Drone Experiment, SmartViz provides a user-friendly interface. For all UNS use cases, SmartViz provides three main functionalities:
to raise an automatic alarm;
to play live audio or video data feed;
to request historical audio or video data.An automatic alarm would be raised if there is a larger number of people in the crowds than a pre-defined threshold (UNS1) or if was detected an anomalous audio event (UNS2). Several widgets are incorporated, including those that visualize crowd density heat maps or city maps with localization of audio events.
This way, a human operator can focus on anomalies in the crowds. Inspection of several data streams would be much more difficult without the help of developed framework.
In the RP2, visualisation was improved by incorporating heatmaps.
Within the external evaluation process of the UI of the UNS use cases, all experts provided feedback the all incorporated widgets (video stream player, audio playback, sound event localisation and heat maps widget), are valuable and effective functionalities and that SmartViz accelerates decision-making in detecting and addressing anomalous events.</td></tr>
<tr><td>Robustness
Related to how robust are the system components</td><td>Change operating conditions (namely distance camera resolution) and measure accuracy.</td><td>MAE degradation not more than 1 compared to the nominal settings.</td></tr>
</table>

<table>
<tr><th>Parameter</th><th>How to measure</th><th>Target to be achieved</th></tr>
<tr><td>during the period of operation</td><td colspan="2">Evaluation and Comments:
We performed experiments related to the robustness of the VCC component, by analysing performance for two different resolutions of the drone camera (HD and downsampled to the 640x360 pixels). The obtained results were identical, i.e., without any degradation. In the case of monitoring from different distances, we have selected data from the UNS Drone dataset recorded from the height of 20 meters and 30 meters. The performance of VCC component was slightly degraded in the case of 30 meters height, compared to the nominal setting (achieve MAE degradation is 0.72).</td></tr>
<tr><td rowspan="2">Performance
Related to how well the system performs the intended task</td><td>Classification accuracy across multimodal sources.</td><td>To increase the average accuracy (10%) for the detection of the presence of people in the scene as compared to baseline (audio only).</td></tr>
<tr><td colspan="2">Evaluation and Comments:
Drone dataset was fully prepared and annotated, as well as corresponding complementary 10-seconds-long audio snippets in the middle of which there is an annotated video frame. The achieved detection accuracy was 76% in the case of audio-only baseline and 100% in the case of video processing.</td></tr>
<tr><td rowspan="2">Transparency
Related to the description of the processes or algorithms that are used to generate system output</td><td>Decision processes are described in a document.</td><td>Document availability.</td></tr>
<tr><td colspan="2">Evaluation and Comments:
All inference results are stored in Elasticsearch (DFB). For possible future inspection in addition end-user is presented with the functionality to verify each inference result during system operation, which is subsequently captured in the relevant field of the Elasticsearch database.</td></tr>
<tr><td rowspan="2">Privacy awareness
Related to the provision of adequate governance mechanisms that ensure privacy in the use of data.</td><td>Periodically evaluate the secure transmission.</td><td>Minimise data breaches.</td></tr>
<tr><td colspan="2">Evaluation and Comments:
Anonymisation is performed at devices that capture raw data (Video anonymisation at drone onboard Intel NUC, Audio anonymisation in the field Raspberry Pi for UNS1, i.e. a laptop for UNS2).
Data transmission security is enabled via EdgeSec TEE. Within the external evaluation process, 80% of experts responded that the proposed UNS1 architecture supports a secure transmission as well as 60% of experts in the case of the UNS2.</td></tr>
</table>

12 Use case UNS2: Localising Audio Events in Crowds

12.1 Introduction

This section gathers all the information pertaining to use case UNS2: Localising Audio Events in Crowds. The first two sections summarise the scope and description of the use case (Section 12.2) and the components, including relevant datasets and the configuration of the framework (Section 12.3). These sections provide a background to the use case evaluations reported in Section 12.4.

12.2 Scope and description of use-case

Ensuring safety in rapidly growing urban city areas and public events is a challenging task, which requires a quick response in the case of anomalous events. Commonly, static cameras are being used for monitoring. In the UNS1, the application of drone cameras and MEMS microphones for monitoring was analysed. However, video monitoring could not help much in the cases of low visibility (for example during the night) or insufficient number of cameras which is likely to happen when cameras are fixed in position. For that reason, in this use case, we analyse the potential of applying microphone array boards for monitoring public events. Microphone arrays can be used for detecting target sound events and finding the direction of the sound propagation, which could help to localise anomalous events in a crowd. Such a system could quickly detect accidents or other kinds of anomalous events.

Table 31: Evaluation scenarios and relevant KPIs for the UNS2

Evaluation Scenario	Target	Relevant KPI
Localisation of target events and alerting event organisers about them	10% relative improvement on metrics in DCASE setup[22] (dataset, cross-validation setup)	**UNS-KPI5: Accuracy Location-dependent error rate and F1-score**
Detection of target events and alerting event organisers about them	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	**UNS-KPI6: Accuracy Classification-dependent localisation error and recall**
Monitoring (e.g., by security crew) of streaming data	The goal is to achieve processing time less than 2 seconds for detecting and localising an anomalous event and less than 15 seconds to provide automatic visualisation using maps on SmartViz.	**UNS-KPI7: Reaction time**

12.3 Summary of Technical Setup and Decision Making Toolkit

The technical setup of the UNS2 use case covers all three layers: Edge, Fog, and Cloud. The IFAG's 8-channel AudioHub Nano microphone boards are used at the edge for data capturing and streaming data to the UNS Edge 3, which is a laptop. The inference is performed using the SELD component, which is deployed at the UNS Edge 3. Before streaming data to the Fog, audio anonymisation is performed alongside the voice activity detection. The UNS Fog server is used for deploying AVRegistry, DatAna Fog, and StreamHandler, whereas the Cloud server

[22] The DCASE setup is described in the Section 3.4.8 of D5.5.

is used for visualisation and SELD training. DataAna Cloud, HDD and DFB are deployed at the Cloud, too. AI inference pipeline of the UNS2 use case is shown in Figure 21.

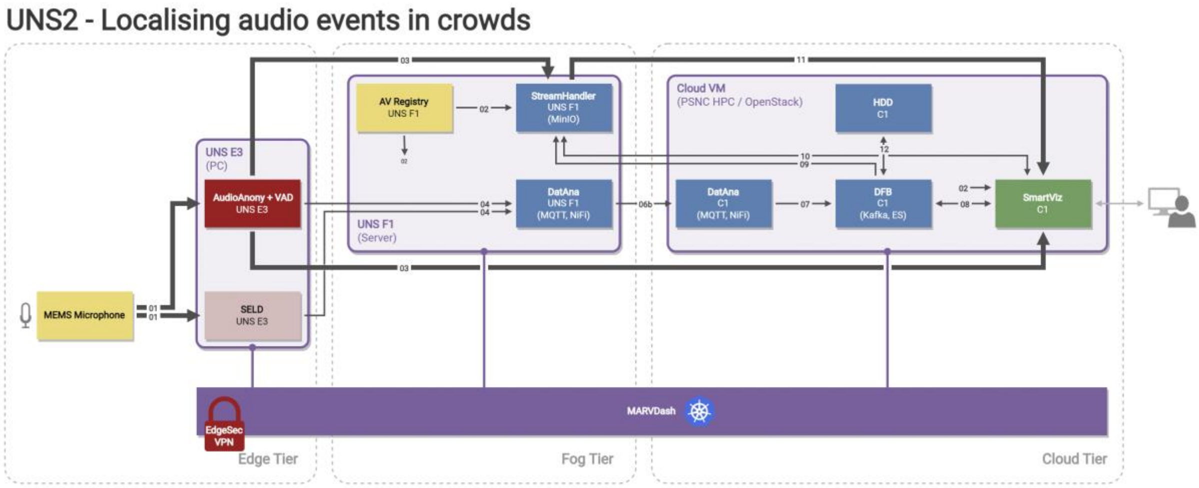


Figure 21: MARVEL RP2 deployment and runtime view of the MARVEL architecture for UNS2 – Localising audio events in crowds (Source: D5.6)

Within the UNS2 use case, we have collected a dataset consisting of 3 different anomalous sound events (gunshot and gunfire, boom and shatter), combined with a chatter class of events. All data are recorded within staged recording process, using samples from the FSD50K dataset, that are further mixed in order to simulate anomalies in crowds. Current state-of-the-art sound event localisation and detection components are trained using data simulated and recorded in laboratory conditions, whereas the UNS2 use case included data recording outdoors, which is much closer to the real scenario. Detailed description of data recording process within UNS2 is provided in D6.3[6] (section 12.2.1).

The objective of the UNS2 use case is to investigate the feasibility of localising and detecting audio events within crowds using audio streams.

The final dashboard for UNS2 includes several key features. One of the prominent features is the Sound Localisation map, which visualises the outputs of the SELD component. Users can interact with this widget to observe consecutive detected events within a selected time period. Each event is represented by an arrow indicating its direction.

Furthermore, the detected events, along with the output of the VAD component, are displayed in a detail widget. Users can request the corresponding audio snippet of an event and play it in the audio player.

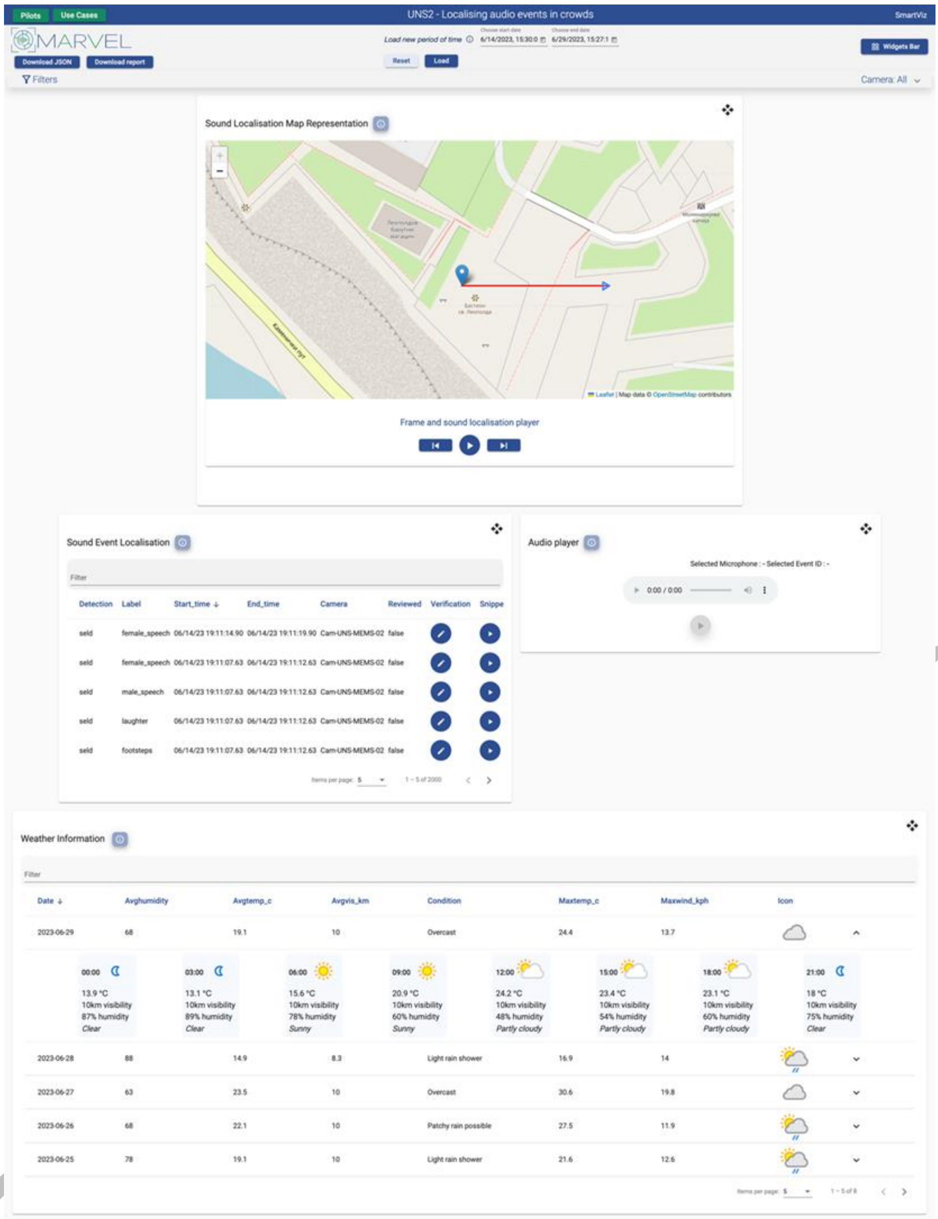


Figure 22: SmartViz dashboard for UNS2

12.4 KPI assessment

This section evaluates the use case against a set of functional KPIs (Table 32) and a set of non-functional parameters (Table 33). These tables are sourced from D1.2[2], D6.1[5] and D6.3[6] and the evaluations are specific to the use case. Section 14 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs.

12.4.1 Use case specific KPIs

This section tabulates the progress in evaluating the use case against a set of functional KPIs (described in Table 31 and evaluated in Table 32) and a set of non-functional parameters (Table 33). The performance of SELD component was evaluated on a set of specific scenarios, as well as processing time required to detect an event and to visualise its direction on SmartViz. Here we also redefine the baseline of the UNS-KPI7, as human reaction strongly depends on the cognitive abilities of the individual operator as well as the number of streams he/she observes.

This means, that initial idea to measure the time needed to a human operator as a baseline is not applicable.

Within the external evaluation process, 60% of experts agreed that the setup facilitates the monitoring of security professionals in an effective way, providing accurate automatic alarms in the case of detected anomalous events which may be a risk to public safety. On the other hand, 40% of experts responded that the idea of incorporating sound event localisation and detection AI tool looks promising but the management and understanding of the provided inference results seem challenging for a large-scale scenario. The achieved results in terms of Location-dependent error rate and F1-score, as well as Classification-dependent localisation error and recall prove the usefulness of the recorded dataset and designed setup as the achieved relative improvements for all observed measures are significant.

The results reported to the reaction time are measured within the benchmarking process and they are previously reported in D5.5. It was demonstrated that the required processing time to automatically detect and localise an event is relatively small (1.3 seconds), whereas the visualisation process of desired event on SmartViz requires slightly larger processing time (11.5 seconds in total).

Table 32: UNS2: Localising Audio Events in Crowds KPIs (source: D6.3)

UNS2: Localising Audio Events in Crowds: functional KPIs				
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**
UNS-KPI5 Accuracy	Location-dependent error rate and F1-score	Baseline of the SELD task at DCASE Challenge	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	UNS staff
	Evaluation and Comments Evaluation was done against data collected within the UNS stage recording process. Error rate (ER) equal to 0.71 is achieved compared to 1.02 ER with baseline. The relative improvement achieved is 31%. In the case of the UNS data, the F-score achieved is 40.0% compared to 2.2% with the baseline, making a relative improvement of 1718%.			
UNS-KPI6 Accuracy	Classification-dependent localisation error and recall	Baseline of the SELD task at DCASE Challenge	10% relative improvement on metrics in DCASE setup (dataset, cross-validation setup)	UNS staff
	Evaluation and Comments Evaluation was done against data collected within the UNS stage recording process. Localisation error (LE) 25.5 was achieved and localisation recall (LR) 51.3, compared with LE 74.0 and LR 10.1 with baseline. The relative improvement achieved is 66% and 408%, respectively.			
UNS-KPI7 Reaction time	Time	NA	Less than 2 seconds for detecting and localising an anomalous event and less than 15 seconds to provide automatic visualisation using	UNS staff

			maps on SmartViz.	
	Evaluation and Comments The achieved overall global latency within benchmarking experiments is 11.5 seconds, as provided in D5.5. Most of the delay was produced while managing heatmap filed in the data stream managed by Kafka via DFB (10.2 seconds). The processing time required for detecting an event and localising it without visualisation is 1.3 seconds.			

12.4.2 Use case specific non-functional KPIs

Table 33 shows the new updated non-functional KPIs for UNS2. Results were mostly obtained from surveys (Appendix section 17.3.1). Surveys consisted of two major parts. The first one related to the Use case architecture and functionalities and the second one related to the UI of the Decision making toolkit and user experience. The respondents emphasised that the end-user experience is very positive and that it supports the overall decision-making process. Scalability and modularity were evaluated without any negative feedback, whereas 20% of respondents raised concerns regarding the challenges in maintaining citizen privacy and data security in the case of integration of sound event localisation and detection systems in urban city areas.

Table 33: UNS2 non-functional KPIs (source: D6.3)

UNS2: Localising Audio Events in Crowds: non-functional KPIs			
Evaluation variable	**How to measure**	**Internal evaluators**	**External Evaluators**
Modularity	Integration of new equipment	UNS staff	IT experts of security crews
	Evaluation and Comments All experts responded that the implementation of new equipment due to continuous long-term upgrades of the UNS2 setup would not be a major problem. Within a dataset recording, UNS staff tested several setup configurations.		
Data protection	Periodic evaluations	UNS staff	IT experts of security crews
	Evaluation and Comments Within the external evaluation process, 60% of respondents selected an option that the integration of sound event localisation and detection systems in urban city areas does not face significant challenges in maintaining citizen privacy and data security.		
End-user experience	Periodic surveys	Infrastructure managers	Potentially public administration or public events organisers
	Evaluation and Comments All collected feedback regarding the user interface was positive. 100% of experts confirm that the sound localisation widget, audio playback option and details widget are very useful and support decision-making process.		
Scalability/Modularity	Extend the solution to the larger number of microphone boards placed in vicinity and exploit a smaller number of microphones per board	UNS staff	IT experts of security crews

	Evaluation and Comments Within the external evaluation process, all respondents provided positive feedback regarding the adaptability of the setup to support several microphone array boards placed in vicinity or to the usage of less than maximal number of audio channels to simplify monitoring. The latter was also experimentally analysed and the results are reported in D5.5 (subsection 8.4.8, RP2 M34 results) in the case of 4 channels, whereas here we report the results of the full setup, i.e., 8 channels. The latter was also experimentally analysed and the results are reported in D5.5 (subsection 8.4.8, RP2 M34 results) in the case of 4 channels, whereas here we report the results of the full setup, i.e., 8 channels. A setup containing a larger number of boards for data recording was assembled and briefly tested in laboratory conditions.

12.4.3 Additional parameters common across all UNS use cases from D6.1/D6.2

Table 30, in Section 11.4.3, tabulates the parameters that are common for all UNS use cases.

13 Early Adopters: The Gozo MVP

13.1 Introduction

This section gathers information pertaining to the Gozo Minimum Viable Product (MVP) which was implemented during RP2 and provided an extra use case as an early adopter of some of the MARVEL framework functions, hence an MVP. The first two sections summarise the scope and description of the use case (Section 13.2) and the components, including relevant datasets and the configuration of the framework (Section 13.3). These sections provide a background to the use case evaluations reported in Section 13.4.

13.2 Scope and description of use case

The Gozo MVP is focused on the requirement of long-term data analytics that shed light on the behaviour of road users (e.g., car drivers, motorcyclists, cyclists, pedestrians, etc.) for the Gozo Ministry. This additional use case, emerged following the second MARVEL info day at M24 that was organised in Malta and is similar to GRN4 but adapted to the specific needs of the Gozo ministry as an early adopter of the MARVEL framework. The Gozo Ministry had a particular interest in gathering long-term traffic data from the cameras already installed to monitor traffic using the Gozo ferry, the only mode to reach the island of Gozo. This data can then be used to find seasonal patterns of vehicles using the ferry and find short-term patterns of the traffic, for example, the time traffic starts accumulating prior to the ferry service. Trajectories of vehicles and pedestrians are also of interest to study the common paths taken by vehicles and pedestrians exiting their vehicles whilst queuing. With this data, the ferry services can potentially be improved which would result in the improvement of quality of life for Gozitans.

From a technical point of view, this use case requires entity detection and its trajectory across the traffic waiting area. Prior to the detection and tracking the video is also anonymised. The descriptive statistics can then be used by the Ministry of Gozo to find traffic trends of interest through the visualisation of the data on SmartViz.

Similar to GRN4, the innovation that is targeted with this use case is the construction of a database that can be used to look up long-term historical data on the trajectories of vehicles and pedestrians. The trajectories and data generated from the CATFlow algorithm are saved on the private MARVEL Data Corpus such that the data can be accessed and processed by the end-user. Patterns of trajectories can be detected through visual inspection of the trajectories, which is a feature of the system.

The users for this system are intended to be transport policy-makers, transport engineers and urban planners in the Gozo Ministry who need data to make informed decisions about infrastructure changes and upkeep. Due to the similarity of the Gozo MVP to GRN 4, the GRN4 evaluation scenarios were used to evaluate the Gozo MVP.

Table 34: Evaluation scenarios and relevant KPIs for the Early Adopters: The Gozo MVP (Similar to Table 9, D6.1)

Evaluation Scenario	Target	Relevant KPI
Detection of the trajectories and the storage of these trajectories to be used later in data-driven decision-making.	The aim is to have a 50% detection rate of the trajectories.	**GRN-KPI6 Availability of historical video samples of pedestrian and vehicle trajectories at two junctions.**

Evaluation Scenario	Target	Relevant KPI
Surveys with relevant traffic experts to determine if this data will help in decreasing the planning time.	The aim is to have confirmation that this data is helpful.	**GRN-KPI7 Increased efficiency in the planning of roads**

13.3 Summary of Technical Setup and Decision Making Toolkit

Figure 23 is a diagram of the Gozo MVP framework configuration which gives information on how the various components interact with each other. In addition to the basic framework, the Gozo MVP makes use of the full CATFlow asset and the VideoAnony component. CATFlow classifies and tracks both vehicles and pedestrians and is internally configured such that it outputs the trajectories of each vehicle as well as the entry and exit points within the camera field of view (FoV). VideoAnony blurs vehicle number plates and pedestrian faces to preserve the privacy of the entities in a video. The video anonymisation was performed on the cloud. In a full MARVEL framework implementation, anonymisation is performed as close to the data source as possible, to minimise any potential data breaches of non-anonymised data. However, the data streams used are inherently available online to the public in their non-anonymised version and coupled with the fact that the Gozo MVP use case is implemented for a limited period of time, anonymisation on the cloud is sufficient. The DatAna component at the cloud receives the results from the AI model instances at the respective layer, transforms them, and sends results to the appropriate Kafka topics of DFB.

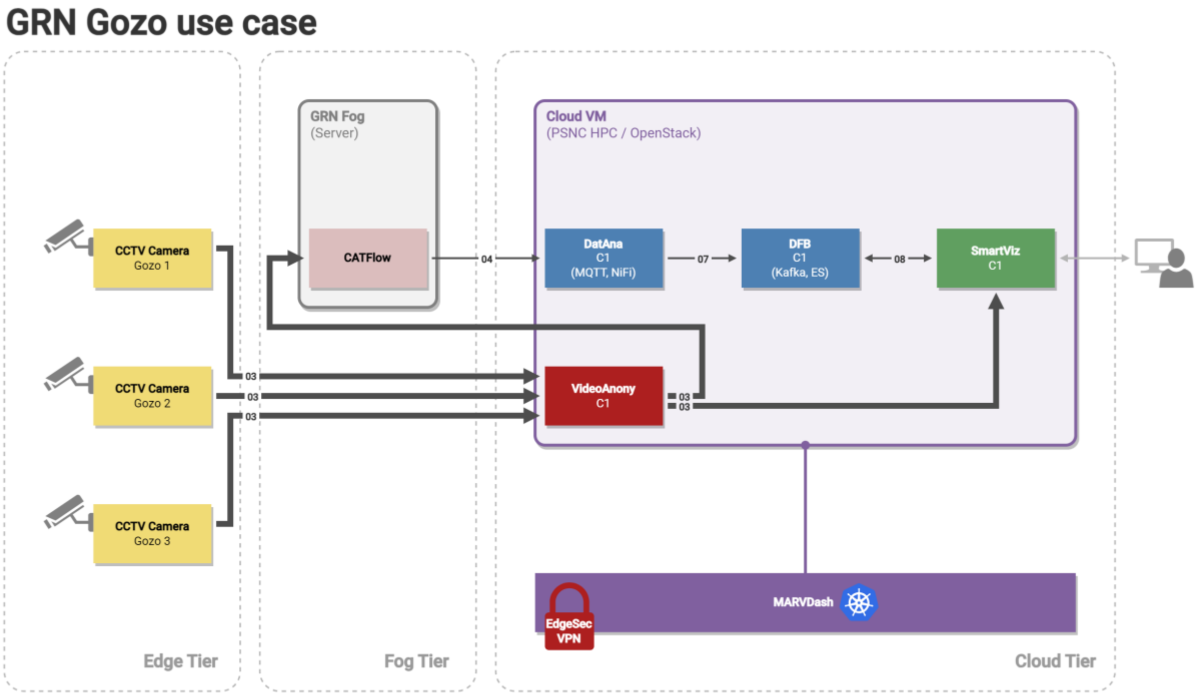


Figure 23: MARVEL RP2 deployment and runtime view of the MARVEL architecture for the Gozo MVP

CATFlow classifies and tracks both vehicles and pedestrians and is internally configured such that it outputs the trajectories of each vehicle as well as the entry and exit points within the camera field of view (FoV). VideoAnony detects vehicle number plates and pedestrian faces and blurs them to preserve the privacy of the entities in a video.

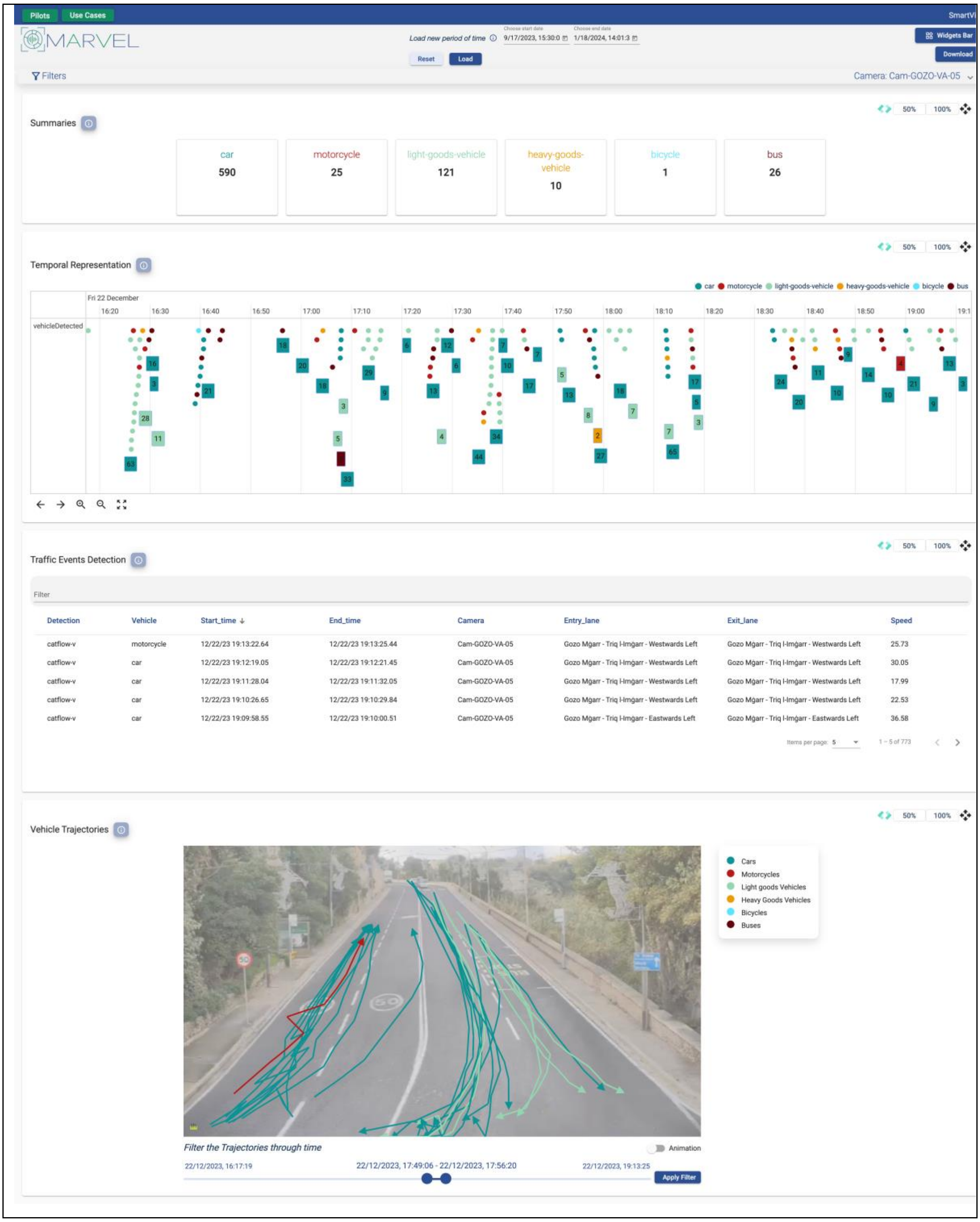


Figure 24: SmartViz dashboard for The Gozo MVP

Figure 24 is an example illustration of the SmartViz dashboard or DMT for the Gozo MVP. The user interface is split into four sections representing the variety of outputs from the AI model. The summaries widget collects the total number of vehicles per category. The Temporal Representation widget visually presents all the vehicles detected in time. The Traffic Events Detection presents incoming events in tabular format and provides a standalone text filtering capability that facilitates a rapid search through the available information. In the Vehicle Trajectories widget, the paths traversed by the vehicles detected component are drawn on the image of the camera feed. The paths are grouped and colour-coded depicting different types of vehicles. All data is collected by the CATFlow component. The users are also able to change the time period and filter by the type of vehicle or lane to investigate further. Moreover, the DMT provides functionalities to rearrange and resize widgets, download visualised data in JSON format, and save the dashboard as a PDF file, thus providing users with greater flexibility and options for data analysis and offline access.

13.4 KPI assessment

This section evaluates the use case against a set of functional KPIs (Table 35) and a set of non-functional parameters (Table 36). These tables are sourced from D1.22, D6.15 and D6.36 and the evaluations are specific to the use case. Section 13 on the other hand deals with KPIs across the project or the pilots, i.e., project KPIs, Business KPIs and societal goals benchmarks.

13.4.1 Use case specific KPIs

The detailed results for the functional KPIs are tabulated in Table 35. The performance of the CATFlow model is evaluated. The system achieves the target of the use case (i.e., correctly detecting trajectories at least 50% of the time). The CATFlow model achieved an averaged object count accuracy of 90.82% and correctly extracted 84.03% of trajectories. It can therefore be concluded that the performance achieved across models satisfies the requirement of the use case. Experts stated that systems like the MARVEL framework bring in the much needed long-term data that is required for long-term studies and planning to address challenges in the urban landscape.

Table 35: Use case specific KPIs for the Gozo MVP

Gozo MVP				
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**
GRN-KPI6 Availability of historical video samples of pedestrian and vehicle trajectories at two junctions.	Automatic detection of patterns	No baseline	80% of the trajectories detected (updated from GRN4	GRN managers
	Evaluation and Comments This use case makes use of CATFLow (Vision based Object detection and tracking) to collect traffic and road junction trajectory data. *The detailed technical evaluation methodology for the CATFlow model is documented in Section 8.1.2 of D5.5*[3] The model mAp (mean Average Precision across all classes) the detector achieved is 0.693, for a confidence threshold of 0.7. The mAp increases to 0.881 for a confidence threshold of 0.5. The mAp considers the overlap of the bounding boxes with the ground truth in the test set. The MOTA (Multi-Object Tracking Accuracy) achieved by the tracker is 54.33%. More importantly, 84.03% of trajectories averaged over various road scenes were correctly detected and the average count accuracy is 90.82%, depending on the camera position and look angle and environmental variables.			
GRN-KPI7 Increased efficiency in the planning of roads	Interviews with the Ministry of Gozo and the Gozo Regional Development Authority	No baseline	potential decrease in time for planning through the availability of data	Urban planners and policy-makers in the Ministry of Gozo and the Gozo Regional Development Authority
	Evaluation and Comments			

Gozo MVP				
KPI	**Metric**	**Baseline**	**Expected result/ Improvement**	**Evaluators**
	The evaluation is carried out from two interviews (Section 17.1.7). The experts in Gozo have highlighted a current lack of long-term data collection before initiating any project planning. All experts emphasised the necessity of long-term data, stating that decisions cannot be made effectively without it, and the MARVEL framework could be a tool for this. One expert pointed out the dependency of the Gozo economy on tourism, which exhibits seasonal patterns not always immediately apparent. The MARVEL framework, according to the expert, could enhance efficiency in planning events historically shown to attract high traffic volumes. Ferry ticket data is a very efficient method to count vehicles boarding the ferry. The MARVEL framework adds temporal queuing data and trajectory data, which is needed for traffic studies in the area. In addition, ministry representatives, including experts, unanimously agreed that the MARVEL framework could be extended to other locations, such as interconnecting roads and village cores, for gathering long-term traffic flow data essential for future urban development and sustainability projects. Addressing challenges faced by regions, an expert noted that data collected by one branch of government may be useful for another but is not often shared. The MARVEL framework's adaptability to various scenarios and the fact that the data can be accessed by multiple users was applauded by experts and would result in more data sharing. They also commended the system's easy-to-understand visualisations, facilitating training of users of different expertise. Regarding system costs, one expert admitted a lack of knowledge about any current data collection, thus no baseline of costs could be set. However, they argued that the costs would be lower when considering large-scale data analytics, aligning with broader EU goals of digitization. Another expert acknowledged that if data is currently collected manually, there would still be an increase in costs upon introducing the system due to the need to retrain personnel working on the system.			

13.4.2 Use case specific non-functional KPIs

The evaluation is carried out from two interviews (section 17.1.7). Overall, interviewees from the Gozo Ministry and Regional Development Authority emphasised their desire to expand the MARVEL Framework to collect long-term data for future projects and policy-making. The SmartViz tool was also commended for the clear and easy-to-use data visualisation which would simplify the training required for personnel using the MARVEL framework. The interviewees commented on the need to carefully introduce the MARVEL framework and highlight its benefits and privacy assurance guarantees to citizens such that it can make the most positive impact on society.

Table 36: Use case specific non functional evaluation variables for The Gozo MVP (Table 6.3 in D1.2)

Gozo MVP			
Evaluation variable	**How to measure**	**Internal evaluators**	**External Evaluators**
End-user experience	Survey	GRN Managers	Gozo policy makers
	Evaluation and Comments The evaluation is carried out from two interviews (Section 17.1.7). All interviewees commented that the data was presented in an easy to use manner. SmartViz easily showed all data collected and made the AI tools accessible for the use of policy-makers.		

Data protection, privacy preservation	Manually evaluate sample of AV data.	GRN DPO	
	Evaluation and Comments See Section 6.4.2		
Secure data transmission and cybersecurity in general	Draw a list of secure data characteristics and evaluate each.	GRN Engineers	
	Evaluation and Comments See Section 6.4.2		

13.4.3 Additional parameters common for all GRN use cases from D6.1/D6.2

Table 4, in Section 3.4.3, tabulates the parameters that are common for all GRN use cases but are also common for the Gozo MVP except for the operability and robustness metrics which are detailed in this section in Table 37.

Table 37: Updated Operability and Robustness for Gozo MVP

Gozo MVP		
Parameter	**How to measure**	**Target to be achieved**
Operability and Robustness *Operability is related to the ability of the components to keep functioning together. This can be quantified by recording downtime for any of the components (assets) along the system pipeline as a percentage of total time.* *Robustness is related to how robust the system components are during the period of operation. This can be quantified by observing whether performance is sustained in various weather conditions (Power source, sensor, and CPU board operation).*	**Evaluation and Comments** The primary AV raw data source in the GRN pilot consists of IP cameras which have no ability to process data at the edge and are limited to transmitting video via an IP connection. The MARVEL cloud infrastructure was used to execute AI models and carry out data anonymisation. The IP camera connection was not always reliable, with some cameras disconnecting occasionally. Typically, the Gozo Municipality was able to get the cameras back online with a few days of alerting them of the issue. The cloud infrastructure was always available.	

14 Project wide KPIs and Evaluation Parameters

This deals with the evaluation of project-wide parameters and KPIs which are presented in three sections; Project scientific and technical objectives (Section 14.1.1) and impact (Section 14.1.2); Societal goals evaluation (Section 14.2) and Business goals evaluation (Section 14.3).

14.1 Project KPIs

The status of the project KPIs is reported in this section.

14.1.1 Project scientific and technical objectives

This section presents the main scientific and technical achievements of the project up to M22 toward the project objectives and is organised around the project objectives as defined in the DoA. For each of the objectives (5 in number), we report in a table the related KPIs, with pointers to the specific part of the DoA (WP, Task), detailed reports on the current achievements and plans until the end of the project.

14.1.1.1 Objective 1: Leverage innovative technologies for data acquisition, management and distribution to develop a privacy-aware engineering solution for revealing valuable and hidden societal knowledge in a smart city environment

Table 38: KPIs status update – Objective 1

KPI-O1-E1-1	**Different kind of resources to be discoverable: ≥ 3**	**Achieved**
This KPI has been addressed within the scope of WP2 and tasks T2.1 and T2.3, where diverse IoT devices were connected to the framework. The KPI has been fulfilled as more than three types of resources i) microphones, ii) cameras, and iii) drones, are used. Related data has been stored in the Data Corpus as well.		
KPI-O1-E1-2	**Increase of data throughput and decrease of access latency by 10%**	**Achieved**
Actions to increase throughput and decrease access latency were taken in the scope of WP4, T4.1. An audio processing board was customised for use with 4 IFAG MEMS (Micro Electro-Mechanical Systems). A custom firmware was written for it which increases the processing speed by x5, thereby reducing the latency significantly. Two AudioHub Nano versions, one with 2 microphones and one with 4 microphones, were released. Both come with a microcontroller that allows connecting the microphones directly via USB sending the data via I2S. This facilitates the connection to a processing edge node. Furthermore, work on an AudioHub version with 8 microphones and a Wi-Fi connection for the data was performed. This 8-microphone board aggregates all 8 audio signals in only one synchronised data channel, eliminating the need for several data channels and decreasing the access latency. This board will enable the audio data processing on the edge, implementing algorithms such as noise reduction or beamforming or sound source location. This board features a PSoC (Programmable System on a Chip) 64 microcontroller that allows for Machine Learning (ML) algorithms to run directly on that microcontroller.		
KPI-O1-E2-1	**Execution time of data management and distribution improved at least 15%**	**Achieved**
Execution time is defined as the time needed between a data action request and the actual results delivered from this action. It can be broken down into different time segments, such as data moving time, data processing time, computation time, inter-node delays, and others. This KPI has been addressed within the scope of WP2 via simulation-based modelling of the Apache Kafka data distribution system adopted by MARVEL and application of combinatorial optimisation solutions tailored to realistic application settings. In D2.2, we reported how we managed to address this KPI by achieving improved data access latency via minimising the Apache Kafka replication latency metric in MARVEL-tailored simulation-based performance evaluation, through the HDD component in T2.2. In		

the same report, we also showed how to achieve improved data loss performance via similarly minimising the Apache Kafka unavailability metric. The methodology has been then enriched with another type of performance evaluation, based on measurements obtained in an Apache Kafka testbed with a variable number of consumers and producers, which allowed us to measure directly the execution time of data management and distribution. The results, which are reported in D2.4, showcase an improvement by at least 15% between resource allocation alternatives in full accomplishment of this KPI.		
KPI-O1-E2-2	**Increase the number of different modality data streams that can be handled by 30%**	**Achieved**
In order to increase the number of different modality streams handled by the MARVEL platform, we have incorporated the ability to handle large binary streams, such as audio-visual data. This was an ability that the components of the project lacked but it was also crucial for the project's success. To that end, we experimented with frameworks that offer multi-cloud storage and retrieval capabilities and can handle such data efficiently. The outcome of this process has been incorporated into StreamHandler and integrated into the project's Data Management Platform, under T2.2. We are monitoring the DMP audio-visual capabilities and continuously provide improvements in terms of stability and performance. In this context, StreamHandler was completely refactored. Particularly, the monolithic approach initially followed was substituted by a microservices-oriented approach. By splitting the functionality of StreamHandler into smaller, more robust microservices we achieved better performance and stability of the component; thus the KPI mentioned is being fully achieved. In the tests performed, StreamHandler was able to efficiently handle multiple streams, both audio and visual, simultaneously, without any degradation in its performance. In addition, the microservice architecture adopted enables horizontal scalability; thus an easy scale-up mechanism. Further information regarding the approach adopted can be found in the deliverable D2.4, while the results of the tests can be found in the deliverables D2.4 and D5.5.		
KPI-O1-E2-3	**Increase the speed of the fusion process by at least 20%**	**Achieved**
The inference results produced by the AI components in the MARVEL framework are eventually fused by being aggregated at the Data Fusion Bus (DFB) component. The operation of the DFB is described in detail in D2.2 and D2.4, while its positioning and function within the overall MARVEL inference pipeline is described in D5.6. The main parts of the DFB are (i) a Kafka messaging system, used for receiving and relaying inference results, (ii) an Elasticsearch data repository, used for the permanent storage of inference results and (iii) an ingestion service (ES-connector), responsible for transferring the inference results from the Kafka messaging system to the Elasticsearch permanent storage. All these constituent parts of the DFB are fully scalable and can operate in cluster mode through the instantiation of multiple service replicas that can distribute the load to the available hardware infrastructure in a horizontal scaling approach and thereby increase the speed of processing (fusing) inference results in order to meet variable demands of addressed use cases. More specifically, concerning the Kafka messaging system technology of the DFB, the Hierarchical Data Distribution (HDD) component further contributes to the optimisation of its performance by recommending an ideal number of brokers and topic partitions for a given load. The operation of the HDD component is elaborated in D2.2 and D2.4. Experiments conducted with the implementation of the HDD technology indicated significant time-related gains. For example, for message sizes of 100kB, the employment of 5 Kafka brokers instead of 3 for serving 100 producer clients led to an improvement of the throughput by 175% and an improvement of the rebalancing time by 71%. Furthermore, in the context of MARVEL activities, a new feature was added to the DFB that is responsible for fusing (merging) multiple distinct inference results from a single origin into unique composite counterparts. This capability has been applied for a selected range of inference results produced by specific AI components (i.e., ViAD, AVAD, SED and AT). In these cases, the original inference results arriving at the DFB are events with a small temporal extent. The DFB fusion mechanism merges consecutive events from the same AI component that share the same event type and refer to the same AV data source according to implemented time-based rules. The initial implementation of this feature relied on a mechanism that accessed raw events that were stored in the DFB Elasticsearch repository to process them. However, this meant that the time required to produce the fused events included the time of transferring the events (inference results) from the Kafka messaging system to the Elasticsearch repository. An updated version of the mechanism that was delivered in the second part of the project was able to access the raw inference results directly from the		

Kafka messaging system, thus eliminating the latency introduced by the ingestion process, thereby increasing the fusion speed. More specifically, this latency for processing 5, 10, and 20 events was found to be improved by 63.87%, 55.27% and 47.20% respectively on average in the updated implementation when compared to the original.		
KPI-O1-E2-4	**Improve data distribution in relevant device resource usage at least 15%**	**Achieved**
Relevant device resource usage considers selected application-oriented resources: network consumption, memory allocation, and number of hardware units. This KPI has been addressed within the scope of WP2 by considering the core data distribution system in MARVEL and investigating resource consumption trade-offs. In this respect, the resources of critical importance that have been optimised are the Apache Kafka topic partitions and brokers through the HDD component in T2.2. Initial results are presented in D2.2 and then updated in D2.4 with systematic experimental results obtained in a testbed through an automatised methodology. The results obtained have shown that HDD can outperform industry best practices by at least 15% in all the scenarios evaluated.		
KPI-O1-E3-1	**Number of incorporated safety mechanisms (e.g., for privacy, voice anonymisation) ≥ 3**	**Achieved**
This KPI was addressed under T3.1 and T4.2. In T3.1, FBK has been working on the development of audio and video anonymisation solutions. A first set of solutions was released and YOLOv5 object detection + blurring was used for video anonymisation (for faces and plates). With regards to audio anonymisation, the current solution is based on shifting the LPC poles and is purely signal processing based which makes it suitable to work in real-time scenarios while being deployed on-premise. In T4.2, AUD has been working on improving its Voice Activity Detection (VAD) model. The model has been upgraded and re-trained with respect to a novel SOTA approach. As real-time inference on edge devices is one of the considered safety measures, the new model has been integrated into devAIce which is conceived to function in real-time on high-end edge devices in order to avoid transferring any sensitive audio data to the successive layers. During the first months of RP1, AUD and FBK have been working on improving their newly developed technologies, and by the end of the integration period, the new VAD module has been integrated with AudioAnony, forming together the audio anonymisation pipeline. This composite component was deployed on edge devices (Raspberry PIs) in the use cases where audio data is present, UNS and MT, and a successful connection and synergy with the rest of the project components (DatAna, DFB, AudioTagging, etc) has been achieved. In T4.3, FORTH has developed, configured and integrated two security-enabling components: EdgeSec VPN, providing end-to-end security across all entities of the MARVEL framework, and EdgeSec TEE, providing security of edge and other devices by ensuring internal device integrity of the code and data. By the end of RP1, the KPI was achieved and overall 6 safety mechanisms were exploited: VAD, AudioAnony, VideoAnony, EdgeSec VPN, EdgeSec TEE and on-premise audio anonymisation. After RP1 and by the end of M24, AUD and FBK kept improving their anonymisation technologies. The reactivity and responsiveness of the VAD model were improved and continuous benchmarking with newly collected data took place. A new approach for voice anonymisation using neural models for voice conversion with pre-trained large-scale models is adopted and a new version of AudioAnony is being developed, same for VideoAnony where a GAN-based face-swapping is also being developed and tested. In addition, shortage and integration bugs of the audio anonymisation pipeline were addressed, and a successful synergy and integration tests of the composite component (VAD-AudioAnony) with the rest of the MARVEL components were achieved. Finally, within T4.1, a new joint innovation was developed, an *on-premise lightweight embedded device for real-time audio streams anonymisation*.		
KPI-O1-E3-2	**The end-to-end data flow from the edge to the cloud, will be 100% encrypted**	**Achieved**
To meet KPI-O1-E3-2, the project's focus is on securing gathered data through contemporary encryption techniques before transmission. This KPI is mainly addressed by the functionality of the security-enabling components developed in WP4 under T4.3, EdgeSec. Upon reception, the data undergoes decryption for processing within the designated layer. End-to-end encrypted communication is facilitated by leveraging EdgeSec VPN, accessible for download from the MARVEL registry. All the exchanged data among the cluster layers (Edge-Fog-Cloud) is transferred through the protected network connection provided by EdgeSec VPN. This latest version of EdgeSec VPN employs the IPSec		

protocol and peer-to-peer encryption. The infrastructure includes a VPN Super Node and VPN Edge Nodes, with all nodes within the MARVEL E2F2C framework functioning as Edge Nodes within EdgeSec VPN. This integration of EdgeSec VPN into the MARVEL framework serves as the foundational layer for incorporating new nodes into the MARVEL Kubernetes cluster.

DatAna developed in WP2 under T2.2, implements encryption protocols to secure the inference messages exchanged among NiFi nodes situated across the edge, fog, and cloud. NiFi facilitates secure SSL data transmission between its nodes by employing certificates, ensuring that only certified nodes can send or receive data from other nodes via the NiFi Site-to-Site (STS) data transfer protocol. Within MARVEL, all NiFi services are set up with secure certificates through trust and key stores, restricting secure communication via STS solely to these certified nodes, thus enhancing the overall security of data transfers.

KPI-O1-E3-3	**Video and voice anonymisation expected to improve by at least 10%**	**Achieved**

As the original description of the KPI did not specify what the 10% improvement refers to, we thus specified it with reasonable interpretation. Considering the different nature when processing video and audio data as well as the different solutions available in the literature, we break down the KPI for audio and video anonymisation:

i. for audio, we refer to a 10% information increase in the audio content compared to complete speech removal;

ii. for video anonymisation, we aim at a 10% reduction in computational complexity as video processing is in general computationally demanding and very robust solutions for face detection and blurring already exist.

Details are also reported in D1.2.

Final formulation: Video and voice anonymisation expected to improve by at least 10%; at least 10% improvement in quality of information preserved in audio in comparison with speech removal; at least 10% computation reduction for video anonymisation.

This KPI is related to enabler E3 and is achieved under WP3 and WP4. Specifically, it was implemented within T3.1 and T4.2, and in relation to T3.5 for the model compression.

For audio, this KPI has been achieved with two different implementations of voice conversion applied to speech segments as identified by the VAD of DevAIce. The first approach implements a signal processing-based voice modification and it is meant to be deployed on edge devices with limited computational complexity. This solution was originally based on the McAdams coefficient and then improved using the Parselmouth Library. The second solution is on a sequence-2-sequence many-to-any voice conversion strategy. With respect to the SOTA, we improved the disentanglement between the speaker and speech content to further increase the quality of the anonymised audio and further remove any speaker-related info. We also investigated the impact of the selection of the target voice. This second approach is meant for offline anonymisation of audio streams for which no real-time and computation requirements are present. To summarise, thanks to the use of VAD and voice conversion techniques the KPI for the audio part is achieved. Unfortunately, results on in-domain data are not available due to the lack of events. However, the evaluation of the single components (as reported in D5.5) confirms the achievement of the KPI.

Regarding video anonymisation, we have already achieved a baseline anonymisation solution with a blurring technique followed by face/car plate detection. We have achieved a GAN-based baseline for anonymisation through face swapping, which has been evaluated on public face-related image datasets. Moreover, we have benchmarked SOTA methods on our new dataset which is built on top of public face-related image datasets but with an emphasis on varying face resolution, which is commonly observed in CCTV (Closed-Circuit Television) videos. Also for video anonymisation, we have implemented two solutions. In the inference pipeline, we have adopted an anonymisation approach with faces/car plate detection followed by a blurring technique. This version has been also optimised for deployment on Jetson Nano devices in case of low fps videos (for example those of MT3 cameras). In addition, as a research-oriented effort, we have developed a GAN-based anonymization method through face swapping. We first addressed the challenges in facial pose preservation and being robust to various face resolutions in surveillance videos. We then reduced the model complexity based on quantisation techniques so that we can avoid training GAN, which is notoriously difficult to converge.

We reported the final result of this option in D5.5. We are able to achieve the KPI with a 73% reduction in model size, and approximately 30% reduction in processing time. Finally, as VideoAnony is foreseen to be deployed at edge devices, we also explored the possibility to perform face-swapping on a microcontroller with more stringent requirements. We could fit the entire pipeline from face detection to GAN-based face swapping on a target platform, a Kendryte K210 with a RISC-V dual core working at 400MHz, achieving an overall 46mJ/frame and 6fps@280mW. While the initial results are promising, there are still many open challenges to tackle, such as the scalability of faces present in the videos. These results are reported in D5.5. To summarize the 10% complexity reduction has been achieved in 2 ways: via quantisation of both blurring and face-swapping approaches and implementing a solution tailored for low-end micro-controllers. The related task T3.1 ended at M24.

14.1.1.2 Objective 2: Deliver AI-based multimodal perception and intelligence for audio-visual scene recognition, event detection and situational awareness in a smart city environment

Table 39: KPIs status update – Objective 2

KPI-O2-E1-1	Standard non-personalised federated learning improvement (performance and speed) at least 10%	Achieved
This KPI relates to WP3, T3.2. Regarding training speed, an adaptive FL protocol was proposed and developed within T3.2 that accounts for communication and node availability statistics thus effectively increasing training speed compared to a standard protocol. For clients that exhibit heterogeneities, the protocol achieves a significant improvement on a synthetic dataset in terms of the training speed over the baseline FedAvg. On the LEAF benchmark (widely adopted and most commonly used benchmark for FL), specifically the MNIST dataset, FedL with NUS strategy reaches an accuracy of 88%, which is only 2.3% (=88/90) lower relative to the accuracy of 90% of the baseline algorithm FedAvg, while at the same time achieving significant savings in the number of communicated data. Specifically, in the same experiments, to reach the indicated accuracy, FedAvg requires transmission of 120Mbs in total, while FedL NUS requires in total less than 80Mb of transmitted model data, achieving savings of more than 33%. Experiments carried out on the MARVEL data for the SASNet visual crowd counting (VCC) model operating within three MARVEL pilots, MT, GRN, and UNS, show that the majority of clients increase the mean absolute error (MAE) accuracy (specifically, for MT with FedL the MAE after 100 epochs is by 30% lower than with FedAvg, while for UNS the MAE is lower by 20%). At the same time, FedL achieves significant communication savings (of 50%, or 28GB) compared to FedAvg.		
KPI-O2-E2-1	**Average accuracy enhancement for audio-visual representations and models at least 20%**	**Achieved**
AU developed a method for Audio-Visual Crowd Counting (AVCC) which provides a 3.06% improvement (in mean absolute error) against SOTA. AU also developed an efficient method for Visual Crowd Counting (VCC) employing Early Exit Branches to reduce the parameter count of the resulting model. The resulting models achieved an improvement of up to 11.11% (in mean absolute error) compared to the competing baseline. Moreover, AU developed methodologies for improving the training of deep neural networks through better initialization of their parameters and by increasing the diversity of representations learned by different neurons in each layer, which improved performance compared to the baselines on visual data classification. Detailed description of these methodologies is provided in D3.1. Furthermore, AU developed a method that increases the accuracy of VCC on heavily compressed images by up to 19.79%. AU developed a method for efficient Visual Crowd Counting (VCC) by lightweight deep neural networks which provides improvements of up to 26.3% (in Mean Absolute Error) against the corresponding baseline. AU developed a method for VCC which improves performance in gigapixel crowd counting by 42.4% (in Mean Absolute Error) against the corresponding state-of-the-art. TAU developed an Automated Audio Captioning (AAC) system and studied approaches to adapt a generic AAC system to a domain-specific task. The transfer learning approach was used to transfer knowledge from other domains, and the resulting model led to a 30% increase (in SPIDEr metric) compared to the competing method. These results have been provided in D3.5.		

KPI-O2-E3-1	Increase the average accuracy for audio-visual event detection by at least 10%	Achieved
TAU developed a Sound Event Detection (SED) method based on a light-weight Convolutional Recurrent Neural Network (CRNN) obtained by replacing Convolutional Neural Network blocks with depth-wise separable convolutions, and replacing recurrent neural network blocks with dilated convolutions, and by utilizing data augmentation techniques to diversify data. The resulting model led to a 56% increase (in macro-averaged F1-score) compared to the competing method. TAU developed the SED method based on audio tagging approach, where tagging is applied inside consecutive segments to get sound event detection output. The system was based on Convolutional Neural Network (CNN) architecture, and the resulting model led to a 76% increase (in macro-averaged F1-score) compared to the competing method. Detailed description of these methodologies is provided in D3.1. In the DCASE2017 Challenge sound event detection setup, the developed SED method achieved a relative increase of 15% in F-score (macro-averaged) and an 11% decrease in error rate with respect to the DCASE2017 baseline. These results have been provided in D6.2.		
KPI-O2-E3-2	**Increase the average accuracy for unsupervised audio-visual event detection at the edge by at least 10%**	**Achieved**
AU developed an Unsupervised Visual Anomaly Detection (ViAD) method based on structured parameter pruning on Memory-augmented Deep Autoencoder which improves performance by 1.2% while reducing the model size and its number of floating-point operations (FLOPs) by 71.78% and 21%, respectively. Detailed description of this methodology is provided in D3.1. The intention of using unsupervised audio-visual event detection was connected to the absence of public audio-visual event detection datasets. However, since the performance of unsupervised methods is considerably lower compared to that of supervised and weakly-supervised methods, a new audio-visual anomaly detection dataset was collected and made publicly available during Y3 of the project and a weakly-supervised AVAD method (as well as the corresponding ViAD method) was developed. Moreover, the audio-visual supervised event detection method YOLO-SED was developed, which is able to run on the embedded device (NVIDIA Jetson Xavier NX) used at the edge. These are described in D3.5.		
KPI-O2-E3-3	**Decrease the time needed to identify an event by at least 30% of current time**	**Achieved**
FBK developed edge solutions for SED based on compact neural models with low parameters and MOPS counts. These solutions are suitable for low-resource pervasive devices. Being able to detect sound events very close to the microphones they eliminate the latency due to data transfer to the cloud. TAU developed a Sound Event Localisation and Detection (SELD) method based on self-attention on learned audio features. The resulting models have almost double the number of parameters compared to the competing method. However, they were benchmarked to be 2.5 times faster than the competition during the inference due to parallelisation achieved with the self-attention blocks, i.e., they led to a decrease in the inference time of 40%. AU developed an efficient method for Visual Anomaly Detection (ViAD) which employs structured pruning for reducing the parameters of a Memory-Augmented Deep Autoencoder network. The resulting model led to a reduction of the parameter count equal to 71.78% and a decrease of the FLOPs equal to 21%, while slightly improving accuracy (AUC). AU and UNS developed a Dynamic Split Computing method which determines the optimal neural network split for achieving the fastest execution time considering the structure of the neural network, the data batch size, and the transmission channel data rate. Detailed description of these methodologies is provided in D3.1. AU implemented FFmpeg and GStreamer audio and visual RTSP readers and integrated them into the AI pipeline. Combined with new preprocessing procedure, the pipeline achieves 93.6% improvement in the AI components latency, as reported in D5.7.		

14.1.1.3 Objective 3: Break technological silos, converge very diverse and novel engineering paradigms and establish a distributed and secure Edge-to-Fog-to-Cloud (E2F2C) ubiquitous computing framework in the big data value chain

Table 40: KPIs status update – Objective 3

KPI-O3-E1-1	Number of novel algorithms and tools utilised from diverse multi-domain technological areas ≥ 3	Achieved

The KPI has been achieved. The MARVEL project takes advantage of a significant number of novel algorithms and tools which allow for the adoption of the Computing Continuum paradigm with a fluid integration of resources at all layers of E2F2C and along the entire data path. Utilised algorithms and tools come from diverse multi-domain technological areas corresponding to the individual layers of MARVEL architecture.

Data Management and Distribution Subsystem

- BroMin and BroMax are the two heuristic algorithms that Hierarchical Data Distribution (HDD) currently offers for Apache Kafka topic partitioning. Published in the proceedings of IEEE CITS 2022 conference[23].
- DatAna tool repurposes the NiFi ecosystem to enable E2F2C data management in terms of communication, moving, and transforming.
- Data Fusion Bus (DFB) tool is a customisable product for data fusion for multiple modality data streams between several connected components and permanent storage. The tool is under development.
- PSNC, an IT technology partner provides data infrastructure within the cloud and services for HPC infrastructure. In addition, the raw data and project results are stored at the data repositories located at the Polish data centre (BST, Poznan, PL).

Optimised E2F2C Processing and Deployment Subsystem

- DynHP uses an incremental pruning method for convolutional neural networks.
- FedL uses a novel non-uniform sampling FL strategy, allowing for flaky updates from federated clients. This strategy is implemented and offered as a custom strategy component for the Python flower federated learning framework.

E2F2C infrastructure

- PSNC offers novel HPC infrastructure (classified in the TOP500 list) and virtualised private cloud with an OpenStack web interface tool allowing for a flexible allocation of computing resources and various class storage services connected with the HPC system.

Audio, Visual and Multimodal AI subsystems

- Text Anomaly Detection (TAD) is a novel anomaly detection tool under development.
- Visual Anomaly Detection (ViAD) is a novel object-based visual anomaly detection tool.
- Visual Crowd Counting (VCC) implements a novel method published in the proceedings of the BMVC 2021 conference[24].
- Audio-Visual Crowd Counting (AVCC) implements a novel method for integrating audio and visual features in early exits. Published in Neural Networks journal[25].

KPI-O3-E2-1	Model compression algorithms to achieve 70% compression rates, without a noticeable degradation of accuracy	Achieved

The approach followed to fulfil this KPI was the development of a Model Compression Procedure that, given a Deep Network Model to be trained on a given dataset, it reduces the number of active parameters during the training in order to save resources both at training and inference time. This methodology was an original output of the project, and it has been tested on several DNN models developed under WP3, Task 3.5. The goal is to reduce the model's size while keeping the accuracy almost unchanged.

The procedure followed was first to explore how to remove as many active parameters as possible. However, since depending on the specific learning Task and DNN at hand, removing a fixed quantity

[23] T. P. Raptis, A. Passarella, "On Efficiently Partitioning a Topic in Apache Kafka", 2022 International Conference on Computer, Information, and Telecommunication Systems (CITS2022), Athens, Greece.

[24] A. Bakhtiarnia, Q. Zhang, and A. Iosifidis, "Multi-exit vision transformer for dynamic inference", British Machine Vision Conference (BMVC), 2021.

[25] A. Bakhtiarnia, Q. Zhang, and A. Iosifidis. Single-layer vision transformers for more accurate early exits with less overhead, Neural Networks, 2022.

of parameters might deeply harm the final performance, we combined it with other techniques, such as low-bit representation. The performance of the methodology has been tested on standard benchmarks, achieving up to 88% compression with a negligible drop in accuracy. These results are published in a paper that appeared in the Elsevier Journal of Network and Computer Applications (JNCA). Such a methodology has been further investigated and tested on one of the MARVEL models (i.e., CSRNet) used in the crowd counting task of the GRN4 use case where the compression obtained was ~66% with less than 3% accuracy drop as reported in D3.2 and D3.6.		
KPI-O3-E2-2	**Optimise performance (prediction accuracy, time-to-decision) of DL deployment by 20%**	**Achieved**
This KPI is associated with the decentralized execution of DL tasks implemented in WP3 under T3.4. To achieve this, the implemented MARVEL E2F2C framework, comprising the Kubernetes cluster and MARVdash dashboard, facilitates the distributed execution of DL tasks while prioritizing the efficient utilization of execution resources. MARVdash plays a crucial role in aligning task resource needs (specifically GPU availability) with the diverse execution sites within the MARVEL distributed environment. The selection of hardware significantly impacts decision-making speed. GPUs, known for expediting computations in DL, accelerate decision times compared to relying solely on CPUs. As a result, enhancements in performance, particularly in decision-making speed, and in the complexity of deployed DL models become achievable, thereby augmenting prediction accuracy. The optimization objective of this KPI is attained through the deployment methodology provided by the updated MARVdash leveraging the capabilities of the personalised Federated Learning approach implemented in T3.2. The revised MARVdash aims to automate deployment target selection within the MARVEL E2F2C framework by leveraging resource availability across its layers. This updated version streamlines YAML file creation by removing the necessity to specify deployment node information. Users now have the flexibility to choose specific nodes based on deployment layers and component resource needs using the MARVdash UI (e.g., GPU). This selection dictates where the component gets deployed. Alternatively, users can opt to select only the desired deployment layer without node specification, leading to deployment on any node within that layer meeting the resource criteria. MARVdash enriches provided YAMLs with corresponding tolerations, affinity, and resource fields, enabling Kubernetes' scheduler to decide the execution environment based on component requirements and the real-time availability of devices, network resources, and overall resource utilization within the E2F2C framework. Furthermore, by utilizing the E2F2C framework's exit points, MARVEL users acquire decision-making capabilities spanning various stages of data processing. This distributed approach strikes a balance between local autonomy and centralized control, optimizing operations, enhancing efficiency, and elevating overall decision-making capabilities. To achieve this, the ability to instantiate or remove the decision-making service at various layers is crucial. MARVdash offers a REST API enabling external systems to manage these services. SmartViz utilizes this functionality to communicate with MARVdash, accessing information about deployed services. Users can then start or stop services at specific exit points (edge, fog, cloud). For instance, a user can terminate a component at the edge and activate it at an alternative exit point like fog or cloud. This interaction empowers end-users to control and manage services, facilitating an efficient distributed E2F2C DL deployment that enables real-time decision-making across all framework layers. In another direction, UNS performed federated learning with non-uniform sampling for visual crowd counting over the three pilot sites. The crowd counting accuracy with the federated learning model is improved by at least 20% with respect to the baseline, being the worst-performing site working in isolation.		
KPI-O3-E2-3	**Increase accuracy levels of real-time observations at the edge devices by 20%**	**Achieved**
This KPI refers to improving the accuracy of observations by 20%. In order to achieve this target, we worked on reducing the time needed during inference by an AI model because. The rationale is that when running an AI model on a resource-constrained device it is paramount to reduce the number of calculations in order to increase the number of queries that it can process. Therefore, provided a correct model, i.e., which provides correct answers, the achievement of this target translates into reducing the		

inference time (since the accuracy is constant) such that the AI model can process a higher number of inferences without missing important events happening in the context under observation. We achieved this target through the compression of the AI models. To showcase it, we measured the inference time needed by an uncompressed model and the one needed by a compressed one (obtained through the DynHP methodology). The results reported in D3.6 show that we had a 4.5x improvement which exceeds by far the 20% targeted by the KPI. The results have been obtained on real edge devices, i.e., Apollo Kit from Smartcow Co. Finally, we benchmarked the deployment of the CRSNet in the AVCC component applied to the GRN4 use case, reporting a speed up of 1.8x which, also in this case, achieves the KPI target.		
KPI-O3-E3-1	**Realise a secure computing framework at all the processing layers**	**Achieved**
This KPI is addressed in the work carried out in WP3 and WP4 under T3.4 and T4.3 respectively. EdgeSec combines two functionalities providing essential security features to the E2F2C framework. It ensures the security of data flow across MARVEL components, securing a substantial portion of network communications between the Edge, Fog, and Cloud layers and associated components (EdgeSec VPN). In addition, it secures application execution for processing sensitive data (EdgeSec TEE). The achievement of goals outlined in KPI-O3-E3-1, specifically related to EdgeSec VPN, involves encrypting collected data and establishing secure communication channels across processing layers. The collaborative efforts of EdgeSec TEE and EdgeSec VPN contribute to secure computing, with both components uploaded to the MARVEL registry. EdgeSec VPN currently facilitates an encrypted tunnel, ensuring data confidentiality and integrity between participating nodes. Notably, it eliminates the need to trust intermediaries like routers and ISPs, mitigating threats such as ISP snooping, attacks on insecure wireless networks, and compromised networking equipment. The comprehensive security measures implemented by EdgeSec VPN significantly contribute to averting potential risks associated with data transmission within the project framework. Moreover, we have implemented HTTPS in our Kubernetes cluster's Ingress which is available through MARVdash, realizing a secure computing framework at all processing layers. By encrypting communication between clients and services, HTTPS safeguards sensitive data during transit, mitigating the risk of unauthorized access or data interception. This security measure ensures end-to-end protection, enhancing the overall resilience of your computing infrastructure and fostering a secure environment across various layers of data processing within our Kubernetes cluster. Finally, authentication offered by MARVdash adds an additional layer of protection by ensuring that only authorized users or systems can access the services.		
KPI-O3-E4-1	**Detailed insights to more than 5 hidden correlations**	**Achieved**
This KPI is linked with the design and development of the Decision-Making Toolkit (DMT) and the visualisations depicted there. It is naturally linked to Enabler 4 and is a direct output of T4.4. However, T4.2, T3.3 as well as the whole WP6 where the real-life experiments are being performed are of key importance since they feed the visualisations (T3.3 and T4.2) and evaluate them (WP6) respectively. This KPI is being closely monitored since the MVP release, where end users had the opportunity to experiment with DMT, and its monitoring continued within RP1. Since then, the DMT was developed in a way to facilitate finding hidden correlations in MARVEL's data. In the final release, in collaboration with the pilots, areas of exploration have been identified to seek deeper insights into hidden correlations. The ones identified are the following: 1. Extend the investigation into the impact of time on the occurrence of traffic anomalies and unusual events in public safety-related use cases. 2. Delve into analysing simultaneous events occurring in different parts of the city during the same time period based on historical data. The implementation of split screens fuelled by distinct data batches is a promising technique and we experimented with it and applied it into two use cases. 3. Investigate the correlations between the number of vehicles on a specific road or junction and the potential requirement for maintenance or repair alerts. 4. Explore the potential correlation between environmental parameters and traffic patterns: This hypothesis aims to examine the intricate relationship between environmental conditions, such as air quality, temperature, humidity, and even seasonal changes, with traffic patterns. The		

incorporation of the weather widget in the DMT is the preliminary stage of this exploration. For instance, it is plausible that increased smog or high temperatures could discourage motorists from traveling, thereby impacting the traffic volume. Similarly, sudden changes in weather, such as rain or snow, can significantly affect traffic speed and density. By understanding these relationships, we can predict traffic fluctuations based on environmental forecasts and implement proactive traffic-management strategies. 5. Assess the correlation between time-of-day, week, or month and the frequency of public safety incidents: Different periods might exhibit different frequencies and types of public safety incidents. For example, late night hours might increase the rate of certain types of crimes or accidents. Similarly, specific days of the week or even times of the year (e.g. holidays and festivities) may show a spike in public safety incidents. By understanding these correlations, we can better predict when and where these incidents are more likely to occur, leading to a more efficient allocation of resources, such as patrol units and emergency services, to prevent or quickly respond to these incidents. This type of temporal analysis offers significant insights for strategic planning and proactive public safety management. 6. By integrating the public transport schedule into the dashboards for the specific monitored areas, the toolkit enables the verification of adherence to the schedule. The detection of vehicles further contributes to this verification process, providing additional evidence on whether scheduling is being followed. This information is particularly valuable for public transport managers, as it helps them assess the effectiveness and efficiency of transportation in these specific areas. It allows them to identify any deviations or irregularities in the transportation system, enabling timely interventions and adjustments as needed. Each hypothesis not only has the potential to reveal more than one hidden correlation, but may also generate additional hypotheses for further exploration once the fully integrated version of the MARVEL framework is released, and users can interact with it extensively.

14.1.1.4 Objective 4: Realise societal opportunities in a smart city environment by validating tools and techniques in real- world settings

Table 41: KPIs status update – Objective 4

KPI-O4-E1-1	More than 10 trial cases to showcase framework's capabilities	Achieved
This KPI is related to enablers E1 and E3. Within T1.3 of WP1, 10 use cases were identified, based on the pilots' functional and non-functional requirements. The use cases are described in detail in D1.2 and cover multiple trial cases as described in the user stories in D6.1 and further updated in D6.3. Five of these use cases have been showcased in the RP1 prototype while the remaining trial cases have been realised during RP2 for the final prototype. One more use case was developed for the city of Gozo acting as an early adapter of the MARVEL Framework.		
KPI-O4-E2-1	**Identify at least 20 dependent and independent verification and validation variables for the system**	**Achieved**
The experimental protocol, described in D1.2 as a result of the efforts under T1.3, includes the definition of the experimental variables used to verify the progress of the project and validate the final framework. Tables in Section 6 of D1.2 list both the dependent variables related to the use case KPIs, as well as the independent KPIs used to assess the achievements of each MARVEL asset. Table 6.12 in D1.2 summarises the experimental parameters, their variables, and how they are measured. These have been defined starting from the pilot requirements as well as accounting for the expected asset innovation reported in the DoA, also in terms of Technology Readiness Level (TRL). The evaluation procedure has been updated in WP5 (benchmarks) and WP6.		
KPI-O4-E3-1	**Execute the trial cases in at least 2 real life smart city environments**	**Achieved**
The trial cases are field-implemented experiments, which are combinations of the use cases defined in Section 1.3.4.5 of the DoA, following the output from T1.3 and the technology developed during the project. This KPI is related to WP6, T6.2. The activities started in M12 and concluded in M6. They are connected to several other tasks and WPs (WP2-T2.1 and T2.2, WP3-T3.2, T3.3, T3.4, and T3.5, WP4-		

<table>
<tr><td>T4.1 and T4.2, WP5-T5.1, and WP6-T6.3). MT, GRN, and UNS led the realization of 10 use cases defined in D1.2 (MT1, MT2, MT3, MT4, GRN1, GRN2, GRN3, GRN4, UNS1, and UNS2). In the second half of the project, the use case UNS2 was modified until it reached its current configuration. The collected data was anonymised, partially annotated for AI algorithm training, provided, and shared with the MARVEL consortium. MT, GRN, and UNS, to successfully fulfil this KPI, organized the pilot-focused meetings with technical partners to share information and discuss issues related to data acquisition (WP4), data storage (WP2, WP3, and WP4), analytics/IA/ML (WP3), infrastructure (WP5), integration (WP5), visualization (WP2 and WP4), demonstrators (WP6), and evaluation (WP6). The MARVEL's experimental protocol (D1.2) and the MARVEL MVP (D5.1) served as the starting point for aligning all data providers to ensure smooth and adequate execution of the experiments. As a summary, 8 use cases were realised in Malta and Trento in two real-life smart city environments and 2 use cases in Novi Sad, Serbia in a simulated smart city environment.</td></tr>
</table>

14.1.1.5 Objective 5: Foster the European Data Economy vision and create new scientific and business opportunities by offering the MARVEL Data Corpus as a free service and contributing to BDVA standards

Table 42: KPIs status update – Objective 5

<table>
<tr><th>KPI-O5-E1-1</th><th>More than 3.3PB of data made available through a Corpus-as-a-Service</th><th>Partially achieved</th></tr>
<tr><td colspan="3">In the first phase of the project, the main components and their interconnections were developed. Then, in the second phase and in order to achieve more than 3.3PB of data to be available through the Corpus, a continuous augmentation of the data that are ingested into the Corpus had to be performed. Therefore, under T2.3 in WP2 an augmentation strategy was developed including SOTA techniques in terms of augmenting audio/video data. Furthermore, a series of scripts were developed in order to automatically inject to the Corpus data that are received from the data pilots. However, due to the limited data availability in terms of volume and velocity that we faced, a portion of data has been injected and augmented into the Corpus.

The current size of the Corpus is around 1.1PBs. The main restricting factor for a higher size was the fact that we had to process video of high compression and low fps to reach processing requirements for real-time operation of the AI/ML components. Although the size is lower than the targeted level, the produced datasets were quite valuable and were appreciated by the community of external users that was established. A detailed justification for the Corpus size is documented under the deliverable D2.5.</td></tr>
<tr><th>KPI-O5-E2-1</th><th>Release SLAs and consider all the relevant aspects, namely accessibility, operability, managing streaming and network, legal considerations, security, privacy and technical concerns</th><th>Achieved</th></tr>
<tr><td colspan="3">In order to facilitate the latter KPI reflected under T2.4 in WP2, a monitoring mechanism of Data Corpus was deployed for continuous observation of the latter in terms of security, privacy, and technical aspects. A Security Information and Event Management (SIEM) has been deployed (using Elasticsearch) as a mechanism offering a holistic view of the Corpus information security. Additional actions that were taken during Y3 included: a) an SLA for the availability of the Corpus infrastructure by PSNC, and b) Letters of Intent (LOIs) by technical partners (STS, FORTH) and the pilots (MT, GRN, UNS) that they will continue supporting the service's operation and accessibility to data, respectively. More details can be found in D2.5.</td></tr>
<tr><th>KPI-O5-E3-1</th><th>More than 5 SMEs used the Corpus</th><th>Achieved</th></tr>
<tr><td colspan="3">The actions performed to address this KPI that is connected to WP2 (T2.3, T2.4) were focused mainly on creating a robust set of functionalities in order to make it possible for MARVEL to share its own Corpus in a secure environment where innovative applications can be built by third parties. The approach followed was to develop, using SOTA open-source technologies, a secure environment for delivering the Corpus as a public, obtained free of charge and as-a-service, making it possible for SMEs and start-ups to build on top of these data assets and create new business by exploring extreme-scale multimodal analytics. Towards this aspect, the strategy was to ensure and enhance the stability of the Corpus in terms of security and reliability, and disseminate the Corpus functionalities through a variety of social events and dissemination activities for attracting potential SMEs for its usage. We performed</td></tr>
</table>

a security assessment with STS's Assurance Platform and a privacy and GDPR-compliance analysis with the SENTINEL Platform. Performance and benchmarking activities were performed in collaboration with WP5 activities. Moreover, three user evaluation trials were performed (two by internal and one by external users). Thereupon, through several info days and dissemination/communication events, a community of around 27 organisations has been established that seems willing to utilise the work done. This includes academic partners from universities, colleges, and research institutes, as well as industries and 8 SMEs. A detailed presentation of these achievements is documented under the deliverable D2.5.		
KPI-O5-E4-1	**Maintain the corpus for at least one year after the end of the project**	**Achieved**
This KPI is addressed within the scope of WP2 and T2.4 by delivering at the end of the project the Data Corpus-as-a-Service, including all the necessary maintenance procedures. An investigation was performed to locate the mandatory objectives of maintenance management of the Corpus such as the optimisation of the reliability of the equipment and infrastructure that will host the Corpus, improvements and adjustments in the operational safety of the Corpus. PSNC, which provides the infrastructure that is used by the Corpus, is committed to supporting this KPI after the end of the project and continuing the Corpus operation.		

14.1.2 Impact

One of the main goals of MARVEL apart from the scientific and technical advancements to meet the project objectives is also achieving the expected impacts. To do so, the MARVEL consortium relies on an impact maximisation strategy that is based on four fundamental axes:

- **Axis 1 - Mission-oriented approach:** cracking specific challenges in the Big Data Value chain addressing rising challenges of the European Data Economy (WP1-WP7).
- **Axis 2 – Openness:** open-access sharing of knowledge and cross-fertilisation with other relevant EU-funded programmes and communities for multimodal event detection (WP1-T1.1, T1.2, T1.3, T1.4, WP7- T7.2, T7.3).
- **Axis 3 – Sustainability**: invest in research and innovation to produce new knowledge and advance existing one, ensuring sustainable growth for the technological advancements (WP6-T6.3, WP7-T7.2, T7.3).
- **Axis 4 - Public engagement**: engage citizens and secure their support to promote breakthrough innovation (WP7-T7.2, T7.3).

The expected impact of the MARVEL project is related to (i) the work programme, (ii) the innovation capacity, competitiveness, and growth, and (iii) standards. For each impact category, well-defined KPIs have been identified by the MARVEL consortium aiming to monitor the progress within the course of the project and measure the respective achievements. Below, for each impact iKPI, we provide a summary of how each KPI has been achieved.

14.1.2.1 Impact related to the work programme

Table 43: KPIs status update – Impact related to the work programme

iKPI-1.1	**At least three (3) tools for complex/federated/distributed systems handling extremely large volumes and streams of data**	**Achieved**
As a result of the work carried out in T3.2, UNS has delivered FedL component implementing the non-uniform client sampling strategy that is particularly well-suited for federated training with a large number of participating clients with arbitrary data modalities, including multimodal data streams; see Sections 2 and 3 in D3.4 for detailed descriptions of FedL. Second, within T3.2 a novel methodology for client clustering within FL with a matching algorithm is developed that finds hidden cluster		

<table>
<tr><td colspan="3">structures in an arbitrarily large pool of clients. The relevance of this methodology is to facilitate model search for a newly arriving client by checking only a small number of model candidates, which are provided by the tool, and hence significantly reducing the complexity of the overall process when a very large number of clients/datasets are participating in the training; see Sections 4 and 5 in D3.4 for details. Third, a tool for Dynamic split computing for distributing DL architectures to two consecutive tiers of the E2F2C architecture is developed within T3.3 (AU, UNS). The tool can operate on DL models of large sizes (large number of layers) and is suitable for data of large volumes (such as video); this work is reported in D3.1. Further, within T3.2 a tool is proposed for optimising data exchange protocols in fully distributed systems with complexity exhibited both in topology (e.g., generic topology) and inherent system randomness (link failures, node failures). For such systems, we develop a tool for adaptively optimising communication frequencies of nodes' interactions/data transmission and observe important communication savings can be achieved with respect to a time-constant strategy, with negligible performance deterioration; the details can be found in Section 10 of D3.4. Finally, a contribution to this KPI is the novel analytical framework based on the theory of large deviations for stochastic gradient descent-based learning algorithms, including federated learning, and the accompanying design metrics that show advantages over the standard mean-square error metrics, being able to analytically capture the geometrical interplay between the noise distribution and loss functions, higher-order noise moments, noise skew and other parameters that impact learning performance; details are provided in Section 9 of D3.4.</td></tr>
<tr><td>iKPI-2.1</td><td>Reduced time-to-market (TTM) by at least 15% for the development of analytics in the MARVEL's decision-making toolkit based on the E2F2C MARVEL architecture</td><td>Achieved</td></tr>
<tr><td colspan="3">The development of MARVEL's Decision-Making Toolkit (DMT) has been carried out as part of the T4.4 activities. A questionnaire has been prepared by CNR, with the help of ZELUS, and distributed to all the partners who developed the DMT and related software. Most of the experts (75%) participating in the questionnaire reported that, based on their experience and expertise, the Time-To-Market (TTM) for the development of analytics with DMT based on the E2F2C MARVEL architecture can be reduced by at least 50%, which fulfils the KPI measurable objective.</td></tr>
<tr><td>iKPI-2.2</td><td>At least three (3) different cyber threats avoided due to E2F2C</td><td>Achieved</td></tr>
<tr><td colspan="3">In regards to iKPI-2.2, which is implemented in WP4 under T4.3, the system is protected against at least three potential attacks, including scenarios such as man-in-the-middle, eavesdropping, and impersonation. However, the system's security measures extend beyond these specific attacks. FORTH additionally to EdgeSec VPN which ensures secure network communications (encrypting network traffic exchanged among MARVEL components), offers EdgeSec TEE. EdgeSec TEE provides a trusted and protected execution environment for processing sensitive data in untrusted settings. It is based on the SCONE confidential computing technology and is appropriately configured for Python applications to run inside Intel SGX enclaves. SGX is a technology developed by Intel which provides hardware-based security measures to protect against cyber threats by creating secure enclaves within the CPU. These enclaves isolate sensitive computations and data, shielding them from unauthorised access even in the event of a compromised system.</td></tr>
<tr><td>iKPI-3.1</td><td>Accuracy ratio (>95%) of training of deep learning models with respect to full dataset training</td><td>Achieved</td></tr>
<tr><td colspan="3">On the LEAF benchmark (widely adopted and most commonly used benchmark for FL), specifically MNIST dataset, FedL with NUS strategy reaches an accuracy of 88%, which is only 2.3% (=88/90) lower relative to the accuracy of 90% of the baseline algorithm FedAvg, while at the same time achieving significant savings in the number of communicated data. Specifically, in the same experiments, to reach the indicated accuracy, FedAvg requires transmission of 120Mbs in total, while FedL NUS requires in total less than 80Mb of transmitted model data, achieving savings of more than 33%. We consider here the full dataset training to be the method that adopts a bucket sampling across the users data, that is, we consider SGD where a mini-batch is formed by sampling data from all users. Compared with this benchmark, we approach its accuracy to within 95%, 90% versus 88% achieved by NUS. We also evaluated the performance on MARVEL data for VCC training task using the SASNet VCC model. The obtained savings due to the employed protocol are significant: for 100 training epochs, FedAvg requires transferring >50% more (model) data (46GB vs 28GB). The FedL-NUS training process results in higher accuracy for 2 clients that exhibit similarity in their data (MT and UNS), showing clear benefits of federated training (see Table 3 in D3.4 for details).</td></tr>
</table>

iKPI-3.2	At least 20% reduction in LOC (lines of code) required due to new deep learning models	Achieved
All developed software implements easy-to-use implementations and interfaces based on widely adopted ML and DL libraries to reduce the number of lines needed for model training and for deploying the resulting models.		
iKPI-3.3	**At least three (3) approaches tested for ML training algorithms**	**Achieved**
FBK investigated the use of Knowledge Distillation for training edge neural models. TAU employed data augmentation techniques specAugment and mix-up to improve the training of the Sound Event Detection (SED) models. TAU also proposed a continual learning methodology for training Automated Audio Captioning (AAC) models, an active learning method for unsupervised training of deep learning models, and a cross-modal contrastive learning method for improving performance in audio data analysis using other data modalities. AU employed curriculum learning and Copycat finetuning for training deep learning models with early exit branches, and unsupervised learning of Convolutional Autoencoders. AU also proposed a methodology for creating visual data augmentations based on textual descriptions or example visual data.		
iKPI-4.1	**At least 15% sound/video data throughput improvement benchmarked against inference in the cloud (standard industry-validated process)**	**Achieved**
Besides specific component benchmarks, T5.4 carried out a system-wide performance assessment of the MARVEL framework for each of the 10 use cases of the 3 pilots. Based mainly on the timestamps of intermediate steps of the inference pipeline, and partially in the monitorisation of the system, T5.4 carried out the benchmarking of the RP2 (M30) use cases and components in relation to performance (latency and throughput). The results have been reported in M34 in D5.5. Most of the individual components report significant improvements in terms of performance beyond the SOTA. Although it is difficult to ensure that overall the system performs more than 15% in throughput, due to the fact that the MARVEL framework is composed of many components running in different layers (edge, fog and cloud), the overall system pipeline performance offers a relatively low latency valid for the purpose of the pilots. As a matter of example, the overall latency in GRN1 use case is around 1 second, while GRN2 (a more complex case with many components in several layers) is around 3 seconds.		
iKPI-4.2	**At least 15% sound/video data access time delays reduction benchmarked against inference in the cloud (standard industry-validated process)**	**Achieved**
Due to the high variety of cameras and audio devices and to ensure privacy, all MARVEL pipelines start with a process of standardising the AV signals into RTSP streams as well as in a process of anonymisation carried out by VideoAnony and AudioAnony components. This process occurs once, typically at the edge, so the data to be analysed by the inference components in all layers is ready to be used. Components then subscribe to the same RTSP streams and proceed with the inference. This decision impacts in the time to access to the AV data by the inference components, but it is a necessary step that ideally should be carried out by any system dealing with heterogenous and potentially sensitive data. Each of the component-owners carried out test cases/benchmarks for their components. These measurements have been reported partially in their respective WPs (WP2, WP3 -focused on inference components- and WP4 deliverables), and in deliverables D5.2 (over MVP results in M14) and D5.5 (over RP2 release in M34) under their respective component benchmarking section. Most of the inference components introduce very low latency to the system. In particular D5.5 reports on the benchmarking from the point of view of the use cases, as explained in iKPI-4.1. The overall latency of the system can be considered good for the type of use cases carried out in MARVEL.		
iKPI-5.1	**At least two (2) different trial cases for experimentation**	**Achieved**
The data providers and pilot owners GRN, UNS, and MT have planned and implemented ten use cases spanning societal challenges in the areas of transport, personal safety and security, and crowd monitoring. Details on the pilot experiments can be found in deliverables D6.1 and D6.3. All GRN and MT use cases executed in real-world settings using real-world data. The two UNS use cases were realised in experimental settings with the recruitment of volunteers. During the planning stages, GRN		

and MT had talks with authorities/private transport consultants and/or law enforcement agencies and /or legal bodies to ensure that the use cases indeed address societal challenges and satisfy legal constraints.

The main enabler task for the implementations of the three trial cases was WP6. The main effort was led by the pilot owners. Each pilot owner chaired a biweekly meeting to discuss and plan all the steps that were necessary to implement and execute the use cases. This started with outlining the user stories, discussing the AI models required and the data that needed to be collected and annotated, the hardware requirements and the data flow through the system. Each necessity was thoroughly discussed with the involved partners to ensure its feasibility and action items were created and monitored by the pilot owners to ensure that the trial execution progressed as planned. Occasionally meetings were held with individual partners to discuss specific topics as required. Partners were then responsible for the implementation and testing of the individual components for the trial execution. The progress of individual components was discussed in the corresponding work packages, (WP2, WP3, WP4, WP5).

WP5 was another essential enabler to ensure the integration of all components into the MARVEL framework and thus execute the trial use cases. To ensure the successful execution of trials weekly integration meetings were organised to discuss and solve any issues with hardware and software components required for the implementation of the use cases. In addition, the issues were logged and tracked on git-hub.

iKPI-5.2	At least six (6) stakeholders engaged by the end of the project to further adopt MARVEL extreme-scale analytics tools and broader technology ecosystem	Achieved

Throughout the project's duration, concerted efforts have been made by all consortium partners to raise awareness of the MARVEL tools and actively engage with potential stakeholders. These efforts have been multifaceted, leveraging a variety of platforms and events to showcase the capabilities and potential applications of MARVEL and its Technologies, tools, etc.

Achievements and Actions:

Event Participation and Workshops: The consortium has actively participated in numerous events, including the Smart Cities Expo World Conference, DCASE 2022 Workshop, and HiPEAC2023. These events served as platforms to demonstrate MARVEL's innovative solutions and engage with a broad audience, including business-oriented stakeholders. The full-day workshop at the IEEE International Smart Cities Conference 2022 was particularly notable for presenting the MARVEL prototype to potential adopters.

Info Days: The MARVEL Info Days have been a cornerstone of our engagement strategy. The first Info Day attracted 96 attendees, many from municipalities interested in the project's progress. On the second Info Day, they had 86 attendees, including academic experts and industry professionals, leading to several dedicated meetings and focused discussions. On the 3rd Info Day, we were focused more on engaging Smart Cities representative and understanding their needs.

Direct Engagements: Beyond events, candid discussions were initiated with smart cities in Greece, such as Trikala and Athens, and with the Urban Technology Alliance to explore the implementation of MARVEL in their testbeds.

Dissemination Activities: The consortium has intensified dissemination efforts, targeting business-oriented stakeholders through news articles and participation in workshops and conferences, as detailed in the provided links. These activities have been crucial in showcasing the practical applications of MARVEL's technologies in various domains, from road safety to autonomous systems.

Expo Participation: MARVEL's presence at the Smart City Expo World Conference, with both physical and e-presence, was a significant milestone. The booth served as a hub for demonstrating MARVEL technologies and engaging with representatives from smart cities.

Results:

The project has met its initial target of engaging six stakeholders, and we are happy to announce that we have successfully engaged multiple municipalities and institutions. These engagements have sparked ongoing discussions and plans for implementation, demonstrating the interest and potential impact of the MARVEL tools.

14.1.2.2 Impact on innovation capacity, competitiveness and growth

Table 44: KPIs status update – Impact on innovation capacity, competitiveness and growth

iKPI-10.1	**Integration of benchmarks, e.g., HiBench and SparkBench for Big Data; LEAF benchmark for federated learning for AI and machine learning; the DeFog benchmark for edge-fog.**	**Achieved**
Based on the monitorisation of the benchmarks, T5.4 partners cross-checked the potential usage of the benchmarks cited, as well as the ones mentioned in D1.2 (or emerging ones), especially for component benchmarking rather than for the system-wide benchmarks (based on timestamps). Some of the benchmarks listed have been used so far, such as LEAF for the evaluation of the Federated Learning component, or the use of timestamps introduced in the project for the edge-fog-cloud performance assessment. Each of the components of the MARVEL framework uses specific benchmarks for its own assessment, and carried out the benchmarking activity for the final release of the components Most of the general-purpose benchmarks listed in the DoA have not been used, as they are not focused on AV data assessment and therefore are not the best match for MARVEL components. D5.5 reported on the benchmarks used for each of the components of the framework		
iKPI-11.1	**Addressing at least seven aspects for efficient sharing of heterogeneous data pools; (i) accessibility, (ii) operability, (iii) streaming and network management, (iv) legal considerations, (v) security, (vi) privacy and (vii) technical concerns.**	**Achieved**
For addressing the technical aspects, Data Corpus development has been based on stable and tested technologies, such as the HDFS and HBase. The flexibility that the latter technologies deliver along with the combination of accessing the Corpus through extendable REST APIs ensures the efficient sharing of the Corpus data pool while addressing the accessibility and operability aspects (T2.3 and T2.4). Moreover, the incorporation in the MARVEL platform of components, like StreamHandler and DatAna, enables streaming and network management. As the Data Corpus is provided as a Service to external communities, a monitoring system has been developed among the Corpus based on STS Security and Privacy Assurance Platform. Moreover, all data that are stored in the Corpus have been anonymized in advance (i.e., VideoAnony and AudioAnony components). The infrastructure is provided via an SLA and the data are disseminated under a Creative Commons license. Based on this, the Corpus is monitored, not only in terms of usage resources (CPU, network, file disk, etc.) but also evaluated in terms of privacy, addressing aspects regarding legal obligations and security (T2.4).		
iKPI-12.1	**Reduced training time by at least 10% compared to standard approach**	**Achieved**
An adaptive FL protocol was proposed and developed within T3.2 that accounts for communication and node availability statistics thus effectively reducing training time compared to a standard protocol. For clients that exhibit heterogeneities, the protocol achieves an improvement of approximately 30% on the standard benchmark dataset (LEAF) and of approximately 50% on MARVEL data in terms of the amount of data transfers and effective training speed over the baseline FedAvg.		
iKPI-12.2	**Increased performance in terms of response time, throughput and reliability compared to standard approach**	**Achieved**
The MARVEL performance in terms of throughput and response time has been analyzed in detail in RP2 in the context of D5.5. Therein, throughput and latency have been rigorously evaluated based on an end-to-end AI inference pipeline application benchmarking strategy, across all projects' use cases. In more detail, latency was defined as the addition of the time consumed by different operations/tasks from the AI inference pipeline. In a pipeline, some operations involve a single component, while in other cases they involve several components executed in parallel. In the latter, the latency is determined by the slowest operation performed by a single task in a concrete time cycle. Throughput was defined as the number of units of information that the MARVEL AI inference pipeline can process for a given amount of time. The measurement methodology used is an adaptation of the pipelining methodology proposed by the M.I.T. Department of Electrical Engineering and Computer Science (https://computationstructures.org/lectures/performance/performance.html). The achieved end-to-end latency and throughput results are reported in detail in D5.5 and are satisfactory in all cases. Note that the end-to-end AI inference pipeline benchmarking strategy implicitly accounts for per-pipeline throughput and latency metrics, i.e., per component groups working in parallel, hence also quantifying		

the relevant isolated component metrics. However, when throughput and latency are also the metrics of interest for certain components working in isolation, detailed isolated component measurements and comparisons with relevant per-component baselines have been carried out. Specifically, latency has been evaluated for multiple components including DatAna, StreamHandler, HDD, DFB, GPURegex. In addition, throughput has been evaluated in detail for several components such as DFB, DatAna, StreamHandler, and GPUregex. Regarding reliability in MARVEL, we adopted the approach of evaluating reliability for various MARVEL infrastructure elements, including PSNC cloud infrastructure, as well as edge infrastructure and fog insfrastructure of the project's use cases. The results achieved were satisfactory, confirming the MARVEL infrastructure reliability.		
iKPI-13.1	**At least 3 industrial companies boosted exploiting MARVEL framework capabilities**	**Achieved**
The MARVEL project stands as a successful catalyst for industrial companies, significantly enhancing their capabilities and offerings through the innovative MARVEL Framework. This progress builds on the significant headway including the full integration of the MARVEL Framework and the recognition of disruptive innovations by the EU Innovation Radar. The project strategically harnessed its capabilities to empower industrial entities, resulting in tangible impacts. Notably, within the project consortium, industry partners such as IFAG, INTRA, and ATOS have experienced substantial advancements in their technologies, serving as compelling evidence of how the MARVEL Framework fosters innovation and technological progress. Additionally, the project actively engaged with external industrial companies during major events like the European Big Data Value Forum (EBDVF) in Valencia and the Smart City Expo in Barcelona, both held in 2023. These engagements have played a pivotal role in fully achieving this project's impact KPI, providing a platform for showcasing the MARVEL Framework's capabilities and its potential to transform the industrial landscape. This comprehensive approach underscores the MARVEL project's commitment to enhancing industrial companies' capabilities and offerings, reinforcing its success in promoting innovation and empowering the industry.		
iKPI-13.2	**At least 3 smart cities directly engaged to MARVEL by the end of the project**	**Achieved**
The MARVEL project is engaged in an ongoing iterative process to establish a network of European smart cities that are kept informed of the project's findings and outcomes, thereby ensuring greater exposure and sustainability. During the project's lifespan, consortium partners have established advanced communication channels with multiple European smart cities regarding MARVEL. To date, seven European smart cities have been approached in order to initiate knowledge transfer activities and communicate the project. Thus far, a network of smart cities has been established, including e-Trikala, Athens (DAEM), Lisbon (Emel), Helmond, Groningen, Rimini, Civinet GR/CY, and Civitas initiative. e-Trikala and Civinet have already expressed interest in MARVEL, and Athens (DAEM), Groningen, and Lisbon (Emel) were invited to the Malta MARVEL Info Day on November 28, 2022, which they attended remotely. In addition, INTRA contacted two of the aforementioned smart cities, GREEK CIVINET and RIMINI, as well as two projects named GREENLOG and SOTERIA and their constituent cities, via email. INTRA requested that the aforementioned smart cities and projects familiarize themselves with the MARVEL components through videos in order to explore potential opportunities for cooperation and demonstration. Furthermore, INTRA organized two dedicated sessions within SOTERIA's Plenary Meeting on May 9 and GREENLOG's technical meeting on July 4 in order to have the chance to present MARVEL's offerings in person and, as a result, engage more smart cities. INTRA received positive initial responses from the parties. Consequently, within September 2023, INTRA and ITML started to organise a dedicated session, the MARVEL 3rd Info Day, to increase awareness of the MARVEL project among EU Smart Cities showcasing its capabilities and potential applications. The 3rd MARVEL Info Day, held on December 4th, 2023, brought together 47 Smart Cities professionals and enthusiasts. During this engaging online event, distinguished experts in the field showcased MARVEL's main achievements, diving into the benefits of the MARVEL project, and showcasing real-world applications. Representatives from four forward-thinking cities in Belgium, Romania, and Greece actively participated, exchanging ideas and networking. Feedback gathered through a nine-question survey provided valuable insights on user acceptance, value proposition, usability, and market potential. A total of 19 responses were received		

from the participants of the 3rd Info Day. This event covered in more details in deliverables D7.6 and D7.7.

iKPI-14.1	**>8 SMEs providing or using data and/or data-related services (e.g. smart services) engaged by the end of the project**	**Achieved**

Throughout the project, we have engaged with a significant number of SMEs, surpassing the target set by iKPI-14.1. These engagements were made possible through a range of strategic dissemination activities, such as targeted events, workshops, and the MARVEL Info Days. Our efforts have helped to raise awareness among SMEs about the potential of MARVEL's data and services, and we are proud of the positive impact we have made.

Info Days and Workshops: The MARVEL Info Days have attracted many industry professionals, innovation hubs, and enterprise representatives. These events have served to inform and create a network of interested parties that have been kept up-to-date with the project's progress.

Showcasing MARVEL Capabilities: The consortium partners have showcased MARVEL's Framework and broader capabilities at various events and conferences. This has allowed SMEs to explore how MARVEL's services can enhance their portfolios and drive innovation in their businesses.

Direct Outreach: Beyond events, targeted outreach efforts have been made to engage SMEs in discussions about the integration and utilization of MARVEL services in their operations.

Results:

The direct result of these efforts is the active engagement of SMEs, with several expressing a strong interest in adopting MARVEL's data-related services. These SMEs span a range of industries, demonstrating the versatility and broad appeal of the MARVEL project's outputs.

Connection to Enablers, Work Packages (WP), and Tasks:

The success of iKPI-14.1 is closely linked to the enablers outlined in the project:

Data Corpus Accessibility: The availability of the MARVEL Data Corpus has been a critical enabler, allowing SMEs to access a rich dataset for developing and enhancing smart services.

Dissemination and Communication (WP7): The dissemination activities under WP7 have been critical in reaching out to and engaging with SMEs. These activities have been designed to highlight the benefits and applications of MARVEL's data and services.

Innovation and Business Development (WP7, WP8): The efforts to engage SMEs also tie into WP7 and WP8, which focus on exploiting the project's results and fostering business opportunities.

iKPI-14.2	**>15% market share increment in MARVEL consortium technology providers**	**Partially Achieved**

This ambitious KPI reflects a long-term vision of the MARVEL partnership and by nature cannot be completed within the timeframe of the project. Acknowledging that MARVEL is inherently a research-oriented project (RIA) and that MARVEL's commercialization is yet to be realized through a solid and well-defined strategy (see deliverable D7.6), the technology providers within the project recognize the imperative to mature results to TRL 8/9. Nevertheless, partners have made strategic advancements throughout the project which comprise positive steps toward the achievement of the >15% market share increment.

To systematically monitor the project's impact on the market share, ITML utilized regular monitoring initiatives established at the beginning of the project. In the first year, a questionnaire addressing individual exploitation plans was distributed. Subsequently, in the second year, an enhanced questionnaire/form - rooted in Innovation Radar and aimed at identifying components/offerings, was circulated within the MARVEL consortium. In this way, the status and progress of targeted markets was effectively tracked. A comprehensive follow-up was conducted in the 3rd year to validate the achievement of the key performance indicator (KPI).

At this phase, the following partners have reported early business successes associated with MARVEL competencies:

- **IFAG (MEMS Microphone Market):** Achieved a remarkable 5% increase in sales in the targeted market, which corresponds to 7.5% revenue growth from 2021 to 2023.

- **PN (Smart Cities Domain):** Secured new business with a contract value of €50K slated for 2024, signalling growth in the domain of smart cities.
- **ZELUS (MARVEL Services):** Experienced a substantial 59.18% growth in turnover related to MARVEL services from 2021 to 2022.

In addition, during the project period, other significant milestones have been accomplished contributing to broadening the market share. In particular,

- **D7.6 Exploitation Strategy:** A comprehensive exploitation strategy integrating building blocks and the MARVEL framework has been completed, laying a solid foundation for future market penetration.
- **Collaboration Agreement Draft:** Partners have proactively drafted an exploitation [TT1] agreement, showcasing a commitment to joint efforts and mutual success.
- **Financial Projection for Joint Venture:** A financial projection has been prepared, outlining the potential for a joint venture among core partners. This will facilitate decision-making in future investments and planning ahead. The projection targets a positive cash flow within 1–3 years with limited investments.
- **Validation through Horizon Result Booster:** The business strategy has been validated by experts from the Horizon Result Booster service, providing external affirmation of its viability.

iKPI-15.1	>5% increment of penetration of IoT applications in MEMS microphone market	Achieved
Reports about MEMS microphones show that the current trend is that more and more revenue has been made with MEMS microphones in the recent past. This trend is expected to continue in the future. The highest increase in revenue was seen in the consumer electronic sector, with an expected increase from 168M to 264M USD in the years from 2020 to 2025. Translated to microphone volume, in this period it is expected to grow from 870M to 1.9BN units. When looking into IoT smart infrastructures market, the expected growth is from 3.8M to 6.5M units in the 2020-2025 period. IFAG contributes to that increase by presenting the boards developed during MARVEL to the different actors in the market (i.e., public authorities, customers, technology developers,...). Those boards present the advantages of introducing high-quality MEMS microphones together with processing technology that provides further algorithm implementation capabilities (i.e., ML algorithms).		

14.1.2.3 Impact on standards

Table 45: KPIs status update – Impact on standards

iKPI-16.1	At least four (4) formal specifications of standards and benchmarks that will be used and influenced	Achieved
T1.3 revised the SOTA on benchmarking related to the key technologies of the project candidate to produce solutions beyond the SOTA. T5.4 is also revising the proposed benchmarks and monitor their evolution and potential new benchmarks that might appear. The different components of the MARVEL framework are using different benchmarks for assessing their functionality (reported in D5.2). Many of the components have benchmarked their results from RP1 and RP2 against specific benchmarks and against the state of the art in their respective fields (e.g., the inference models). In this sense, more than four benchmarks or shared datasets for assessing performance have been used for that purpose so far (e.g., LEAF, UrbanSound8k, MOT, CelebA, etc.), or via the usage of challenges (e.g., DCASE), or existing datasets (e.g., DISCO).		

14.2 Societal goals evaluation and Impact

Table 46 lists the societal goals that have been identified by the GRN and MT pilots as reported in D1.2-Table 7.2. It should be noted that most of these goals are difficult to evaluate during the lifetime of this project for two reasons: (a) the pilot implementations are limited in space and time and any evaluation would be limited to such implementations; (b) a longer time period beyond the end of the project together with widespread deployment is necessary to evaluate whether the framework addresses some of the societal goals in the longer term. For these reasons, most of the evaluations will rely on surveys that study the perceived impact of the project on these goals. Table 46 summarises the evaluations which the pilot owners collected till the end of the project.

Table 46: Summary of the societal goals addressed by the pilots and the related measurement strategies. * indicates metrics whose evaluation would require a time span longer than the project. The tables describe the perceived impact each use case has on each societal objective.

Pilot	Societal Objectives	Measurements
GRN	**Increased uptake of active modes of travel, such as walking and cycling**	**Perceived changes in habits through targeted surveys***
	Evaluation and Comments The evaluations are carried out on the basis of the responses obtained via the surveys (Section 17.1.1) and interviews (section 17.1.7). The evaluations are organised in terms of the impact of each use case on the societal objectives GRN1: Safer Roads 58.5% of respondents from the general public are confident that more people will take up cycling if car drivers are alerted and reminded of nearby cyclists on the road. Furthermore, 78.6% of the general public would consider using a bicycle if roads are safer for Vulnerable Road Users (VRUs). GRN2: Road User Behaviour There is general agreement that law enforcement and better driving behaviour will result in an increased uptake of active modes. Educational campaigns will contribute very little to the increase in uptake of active modes, unless carried out in parallel with increased law enforcement. 75% of respondents from the general public are in favour of using AI models in detecting cars that overspeed and 49.3% are in favour of using AI models for ubiquitous law enforcement. Experts are of the opinion that AI models are required for law enforcement, with the human responsible for the final review and decision. GRN3: Traffic Conditions and Anomalous Events The automatic computation of traffic conditions may influence the uptake of active modes. On the other hand, the real-time detection of anomalous events does not seem to have an impact on active modes uptake. GRN4: Junction Traffic and Trajectory Collection More than 85% of the respondents from the general public are of the opinion that the collection and use of long-term traffic data in transport planning can have a significant impact on the uptake of active modes of travel, such as walking and cycling, since such data can help in the planning/designing for such modes. Experts are of the opinion that widespread long-term data can have a significant impact on the uptake of active modes, but only if the data is used effectively in the planning process.	
	Increase of safety for vulnerable road users	**Perceived safety through targeted surveys***
	Evaluation and Comments The evaluations are carried out on the basis of the responses obtained via the surveys (Section 17.1.1) and interviews (Section 17.1.7). The evaluations are organised in terms of the impact of each use case on the societal objectives. GRN1: Safer Roads 77.1% of the public is confident that if the GRN1: Safer Roads system is deployed over a wide area cyclists and pedestrians will feel safer on the road.	

Pilot	Societal Objectives	Measurements
	Experts agree only if law enforcement is active. Without law enforcement there is always a percentage of drivers who will ignore the alerts and this will get worse as drivers realise it. GRN2: Road User Behaviour Both the general public and the experts agree that Law enforcement (over speeding highest on the list) has a high impact on increasing safety for VRUs. In addition, the promotion of road etiquette and the detection of bad behaviour feeding into educational campaigns and also law enforcement (e.g. detecting vehicles on paths intended for cyclists) will increase safety for VRUs. GRN3: Traffic Conditions and Anomalous Events Detecting obstructions on the road (anomalous event detection) would increase safety for VRUs. This follows the fact that VRUs have to get off the path in the presence of obstructions. GRN4: Junction Traffic and Trajectory Collection There is strong agreement among experts that the long-term collection of data regarding traffic conditions and trajectories can help increase the safety for vulnerable road users (i.e., cyclists and pedestrians) in various ways; (a) Trajectories especially for pedestrians and cyclists are useful to determine desired paths or critical crossing points; (b) Speed data is important as it will highlight and frequency count over speeding issues; (c) Data may be used to discover which parts of the infrastructure are friendly to active modes of travel and which parts are not.	
	Better education of road users.	**Perceived usefulness through a survey***
	Evaluation and Comments The evaluations are carried out on the basis of the responses obtained via the surveys (Section 17.1.1) and interviews (Section 17.1.7). The evaluations are organised in terms of the impact of each use case on the societal objectives. GRN1: Safer Roads Experts are confident that notifying car drivers of the presence of cyclists and pedestrians is in itself educational and will result in better drivers in the long term. This idea can be extended to other instances of road etiquette. GRN2: Road User Behaviour 60% of the respondents from the general public believe that educational campaigns can be effective at reducing certain types of behaviour. However, 80% are confident that law enforcement has a larger impact on educating road users and this includes educating pedestrians and cyclists. Experts say that educational campaigns are mostly needed when the law changes or the population has developed bad habits over a long time. Otherwise, law enforcement is the better teacher, however, there are exceptions, and some examples were given. GRN3: Traffic Conditions and Anomalous Events The detection of anomalous events related to specific events can be fed to the education of road users to increase awareness of bad driving habits. Such a system can be linked to VMS (variable message signs) which are spread across the country (Malta) but do not currently offer any real-time information. This is similar to ISA and road-side warning systems. GRN4: Junction Traffic and Trajectory Collection Experts agree that long-term collection of data regarding traffic conditions and trajectories can help in the education of road users. Speed data needs to be collected and then drivers are alerted of their excessive driving speeds since speeding is a key contributor to collisions and injuries.	
	More efficient use of the physical transport resources	**Perceived usefulness through a survey***
	Evaluation and Comments	

<table>
<tr><th>Pilot</th><th>Societal Objectives</th><th>Measurements</th></tr>
<tr><td rowspan="4"></td><td colspan="2">The evaluations are carried out on the basis of the responses obtained via the surveys (Section 17.1.1) and interviews (Section 17.1.7). The evaluations are organised in terms of the impact of each use case on the societal objectives

<u>GRN1: Safer Roads</u>

Experts mentioned that, amongst other measures, significant increase in the uptake of active modes should reduce demand for more vehicle infrastructure.

<u>GRN2: Road User Behaviour</u>

Experts believed that improving certain road user behaviour should result in better traffic flows.

<u>GRN3: Traffic Conditions and Anomalous Events</u>

88.5% of the general public believe that the detection of anomalous events such as road accidents, obstructions on the road and irresponsible driving could result in less congestion.

61.5% of the general public believe that predicting journey times (from measured traffic conditions) on main roads in advance allows drivers to consider other routes. However, 21.4% are against re-routing (more efficient use of the physical transport resources) since such a system will make use of residential roads.

<u>GRN4: Junction Traffic and Trajectory Collection</u>

88.4% of the general public believes that the use of long-term data in the analysis of road use would result in a more efficient allocation of physical resources for road transport. All experts agree that the collection of short and long-term data is necessary for maximising the efficient use of the physical transport resources.</td></tr>
<tr><td>Better maintenance of the physical transport resources, for example, estimation of the volume of road traffic informs the kind and quantity of material to be used in road building.</td><td>Perceived usefulness through a survey

Mobility*</td></tr>
<tr><td colspan="2">Evaluation and Comments

The evaluations are carried out on the basis of the responses obtained via the surveys (Section 17.1.1) and interviews (Section 17.1.7). The evaluations are organised in terms of the impact of each use case on the societal objectives.

<u>GRN1: Safer Roads</u>

Data collected by bicycle detectors can be used for the design of segregated cycle paths or improvement of cycling infrastructure (note: overlaps with GRN4 data collection capacity).

<u>GRN2: Road User Behaviour</u>

Law enforcement, accident data and anomalous trajectories can be used to indicate portions of the infrastructure whose design may require revisions. For example, the detection of a collision could automatically send a message to the road maintenance department to visit the site and fix the damage, once it has been reviewed by court experts or road safety auditors.

<u>GRN3: Traffic Conditions and Anomalous Events</u>

Data originating from the detection of traffic conditions over time can be used for the scheduling of physical transport resources maintenance jobs.

<u>GRN4: Junction Traffic and Trajectory Collection</u>

The summarised counts of various vehicle types can be used in the forecasting and planning of maintenance. In addition, trajectories at a junction may indicate that its design may need to be reviewed.</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>MT</td><td>Better and quicker detection of potentially dangerous situations</td><td>Surveys for Local Police</td></tr>
</table>

Pilot	Societal Objectives	Measurements
	Evaluation and Comments The evaluations are carried out on the basis of the responses obtained via the surveys (Section 17.2.1) and interviews (section 17.2.6). The evaluations are organised in terms of the impact of each use case on the societal objectives <u>MT1: Monitoring if Crowded Areas</u> The system serves as a valuable support for the daily operations of the Local Police (81% of the interviewees), which currently lacks any support to monitor more than 700 cameras simultaneously. The alert system enables real-time monitoring, unlike now when such control can only be done retrospectively. The survey conducted, the interviews carried out, and the checks performed by the personnel of the Municipality of Trento corroborate this conclusion. <u>MT2: Detecting Criminal and Anti-Social Behaviours</u> The system serves as a valuable support for the daily operations of the Local Police (82% of the interviewees), which currently lacks any support to monitor more than 700 cameras simultaneously. The alert system enables real-time monitoring, unlike now when such control can only be done retrospectively. The survey conducted, the interviews carried out, and the checks performed by the personnel of the Municipality of Trento corroborate this conclusion. <u>MT3: Monitoring of Parking Places</u> The system serves as a valuable support for the daily operations of the Local Police (73% of the interviewees), which currently lacks any support to monitor more than 700 cameras simultaneously. The alert system enables real-time monitoring, unlike now when such control can only be done retrospectively. The survey conducted, the interviews carried out, and the checks performed by the personnel of the Municipality of Trento corroborate this conclusion. <u>MT4: Analysis of a Specific Area</u> The system serves as a valuable support for the daily operations of the Local Police (66% of the interviewees), which currently lacks any support to monitor more than 700 cameras simultaneously. The alert system enables real-time monitoring, unlike now when such control can only be done retrospectively. The Mobility Service could also monitor in real-time the effectiveness of the choices made in terms of urban mobility. The survey conducted, the interviews carried out, and the checks performed by the personnel of the Municipality of Trento corroborate this conclusion.	
	Safer mobility in critical areas	Surveys for Local Police
	Evaluation and Comments The evaluations are carried out on the basis of the responses obtained via the surveys (Section 17.2.1) and interviews (section 17.2.6). The evaluations are organised in terms of the impact of each use case on the societal objectives The survey confirmed the MARVEL framework has the potential to reduce time for planning urban security plans and mobility plans <u>MT1: Monitoring if Crowded Areas</u> It is possible to monitor the presence of vehicles in a pedestrian area. This aspect contributes to improving the perceived level of security. <u>MT2: Detecting Criminal and Anti-Social Behaviours</u>	

Pilot	Societal Objectives	Measurements
	It's possible to monitor the presence of vehicles in a pedestrian area. This aspect contributes to improving the perceived level of security. <u>MT3: Monitoring of Parking Places</u> The analysis of short-term data provides the possibility to extract behaviours of vehicles, buses, bicycles, pedestrians, etc., elaborating mobility and security plans that are more objective and verifiable over time (80% of the voters). All these aspects contribute to improving the quality of life for citizens. (100% of interviewees) <u>MT4: Analysis of a Specific Area</u> The analysis of historical data series allows for the identification of behaviours of entities moving within the city territory, forming hypotheses, developing mobility plans, and verifying their effectiveness (80% of the voters), Thus improving the quality of life. (for 92% of the people).	
	Better life quality and increased security for citizens	Surveys for Local Police
	Evaluation and Comments The evaluations are carried out on the basis of the responses obtained via the surveys (Section 17.2.1) and interviews (section 17.2.6). The evaluations are organised in terms of the impact of each use case on the societal objectives, <u>MT1: Monitoring if Crowded Areas</u> The possibility of true real-time monitoring, combined with the sending of appropriate alerts, allows for faster intervention (for the 54% of voters by more than 30%.) and enhances the perceived sense of security for citizens (for 90% of the people). The ability to analyze historical data enables a better understanding of crowd behavior in specific situations. Such information can be used to better prepare security plans for events organized in public places (escape routes, arrangement of infrastructure, etc.). <u>MT2: Detecting Criminal and Anti-Social Behaviours</u> The possibility of real-time monitoring, coupled with the sending of appropriate alerts, allows for quicker interventions (for the 63% of voters by more than 30%.) and enhances the perceived sense of security among citizens (for 90% of the people). Additionally, it provides the opportunity to increase trust in the local police force and foster active collaboration among the various involved stakeholders. <u>MT3: Monitoring of Parking Places</u> The system allows real-time monitoring, reducing response times (for the 63% of voters by more than 30%) and increasing the perceived sense of security for citizens (for 100% of the people). Analysis of historical data enables the examination of anomalous situations, their occurrences, and parking usage. <u>MT4: Analysis of a Specific Area</u> The real-time monitoring features enable interventions in reduced timeframes (for the 66% of voters by more than 30%.). The analysis of short-term data provides the possibility to extract behaviors of vehicles, buses, bicycles, pedestrians, etc., elaborating mobility and security plans that are more objective and verifiable over time. All these aspects contribute to improving the quality of life for citizens (for 92% of the people).	

14.3 Business goals evaluation and Impact

The business goals and corresponding evaluation methods have been suggested and reported in D1.2 during the earlier months of the project. The business KPIs are related to a number of functional and non-functional KPIs, namely low latency and detection time, increased efficiency in data-driven planning/decision-making, end-user experience, citizens (or target groups) satisfaction, and scalability. The business goals will be applied to the GRN and MT use cases, as shown in Table 47. Since the system is not in production, (at the end of the project) goals such as revenue and profit increase, cost reduction etc. cannot be quantitatively determined. As such, the goals are evaluated on the basis of qualitative surveys (Section 17.1.1) and interviews with experts (Section 17.1.7).

Table 47: Mapping of the business-related use case requirements into DataBench business metrics. The table includes actual pilots' requirements as well as high-level expected goals (Table 7.1 in D1.2)

Metric	GRN		MT	
	Requirement/Impact	**Evaluation method**	**Requirement/Impact**	**Evaluation method**
Revenues increase	GRN may improve the value of the current product making it more attractive for other potential clients.	Market Study	N/A	N/A
	Evaluation and Comments GRN explored new markets in law enforcement and support for active modes. GRN engaged in new project proposals related to the use cases implemented.		N/A	
Profit increase	For the related public administration, more efficient use of taxpayer funds towards safe and functional roads.	Survey	For the related public administration, more efficient use of taxpayer funds towards safe and functional public spaces.	System usability System efficacy Comparison w.r.t. the current situation
	Evaluation and Comments According to transport experts, the collection of data to inform decision-making is of paramount importance. However, the first step is to allocate public funds for the collection of data. This data should be made available to interested parties, such as researchers, urban planners and shared among all public authorities. This should lead to more efficient transport systems and therefore more effective and efficient use of public funds.		**Evaluation and Comments** According to the statements of the Commander of the Local Police and the Head of Urban Security Unit, the system would allow a rationalisation in the use of available manpower and thus a better use of public funds to ensure public safety. Having the possibility of reducing the number of operators dedicated to viewing the video surveillance cameras thanks to the alert function in the case of potentially dangerous situations would allow for more personnel on the territory directed to the areas where critical situations are found by the system, also reducing intervention times.	
Cost reduction	Either cost reduction or an increase in productivity using data-driven decisions. Instantly available data can reduce the required human efforts (or allow better use of human efforts)	Survey	Cost reduction due to increase in productivity and in more effective data-driven decisions. The use of the multiple sensors deployed in the city leads to reductions in the cost of human monitoring as well as in the reduction of social costs due to misbehaviour. Similarly, better use of the recorded data could reduce the planning cost for mobility improvement.	System usability System efficacy
	Evaluation and Comments According to transport experts, the collection of quality and relevant data will improve the decision-		**Evaluation and Comments** According to the statements of the Commander of the Local Police and the Head of Urban	

	making processes. The reduction in costs will take place over time, following the initial investments in such systems. In law enforcement, the MARVEL system will lead to better use of human resources, who will be involved in the reasoning and validation tasks instead of manually searching for events in video feeds, as is with the current processes. According to experts, social costs will decrease only if deterrent is proportional to the improper behaviour.		Security Unit, the system would allow a rationalisation in the use of available manpower and thus a better use of public funds to ensure public safety. Having the possibility of reducing the number of operators dedicated to viewing the video surveillance cameras thanks to the alert function in the case of potentially dangerous situations would allow for more personnel on the territory directed to the areas where critical situations are found by the system, also reducing intervention times. On the other hand, according to the evaluations of the Commander of the Local Police, the Head of Urban Security and the Mobility Office, the possibility of having long-term datasets available practically in real-time would make it possible to draw up urban security and mobility plans in a shorter time than the current situation. Finally, the monitoring of such plans would be easier and more immediate	
Time efficiency	Reduced time in processing data for data-driven decisions. Quicker intervention for real-time monitoring scenarios.	Survey	Reduced time in processing data for data-driven decisions. Quicker intervention for real-time monitoring scenarios.	End-user experience
	Evaluation and Comments According to transport experts, the MARVEL framework will definitely reduce the time for processing the raw data and can also result in more data and subsequently better decision-making. For real-time situations disruptive events or associated symptoms are on average detected earlier and this should result in a quicker response.		**Evaluation and Comments** According to the statements of the Commander of the Local Police and the Head of Urban Security Unit, the system would allow a rationalisation in the use of available manpower and thus a better use of public funds to ensure public safety. Having the possibility of reducing the number of operators dedicated to viewing the video surveillance cameras thanks to the alert function in the case of potentially dangerous situations would allow for more personnel on the territory directed to the areas where critical situations are found by the system, also reducing intervention times. This aspect is also validated by the results obtained from the various Surveys submitted to the Local Police and the Mobility Office according to which in all use cases the intervention time would be reduced on average by at least 30% compared to the current response times.	
Product Service quality	Satisfaction of cyclists and other vulnerable road users. Improvement of data-driven design. Improvement of the internal end-user experience and administrator opinion on the	Survey	Improvement of the internal end-user experience and administrator opinion on the use of the real-time sensor feeds.	Surveys to the Local Police Citizens' report and independent analyses Traffic-related measures

	use of the real-time sensor feeds		Improvement of the efficacy and efficiency of the data-driven decision-making process. Improvement of public opinion and reduction in citizens' reports.	Annual reports such as the quality of life report of the newspaper "Il Sole 24 Ore"
	Evaluation and Comments From surveys, perceived increase in safety through real-time sensors and more law enforcement is to the satisfaction of Cyclists and other VRUs. The possibility of collecting and processing more data results in a perceived improvement in data-driven Design. SmartViz provides a comprehensive data collection tool for the urban planner and a usable Decision Making Toolkit for the manager working with real-time information.		**Evaluation and Comments** The opinions of the Commander of the Local Police and the Mobility Office, also surrogated by the feedback collected through the surveys, indicate how the graphical interface of the MARVEL framework helps to understand the collected data in a simple and intuitive way. The functionalities developed for the UI allow potential users to fully understand the potential of the system, resulting in an improved decision-making process and, consequently, in better quality of the local territory, the city surveillance, the quality of life of the citizens, their perceived security, and the experience of tourists.	
Business model innovation	More efficient transport networks that encourage green and active modes of transport, contributing to more responsible public spending and a better quality of life.	Survey	N/A	N/A
	Evaluation and Comments Mobility is one of the main issues to tackle to reach carbon neutrality and improve quality of life. Encouraging green, shared and active modes of transport promises to reach this objective. According to transport experts, this is achievable through the collection of both historical and real-time data.		N/A	

15 Feedback from External Advisory Boards and External Stakeholders

MARVEL relayed on an innovative methodology to develop and deliver a robust scientific prototype for increased intelligence in a smart city environment aiming to extend to the degree possible the typical cycle (requirements, development, and evaluation) to a more agile approach that relies on four phases: (i) the *Baseline phase* (M1-M8), (ii) the *Innovation phase* (M9-M18), (iii) the *Experimentation phase* (M19-M30), and (iv) the *Consolidation phase* (M31-M36).

During the Baseline phase, which was the driving force of the project, the MARVEL architecture and the experimental protocol were defined, a SOTA update was conducted, and a market analysis was initiated. The Innovation phase started at M9, focusing on exploring various approaches for project components, delivering the Minimum Viable Product (MVP) as a proof-of-concept demonstrator, and releasing the 1st Integrated prototype of the MARVEL framework at M18.

The Experimentation phase, started at M19, converged technological innovations, executed field trials for the use cases that were developed during RP1, and developed remaining cases for the 2nd release of the MARVEL framework at M30. During the Consolidation phase which was started at M31, the MARVEL Data Corpus and the final version of the MARVEL framework were released. Moreover, the developed use cases were evaluated not only by the involved pilots but also by external evaluators.

Throughout these phases, external partner feedback played a pivotal role in refining the MARVEL framework. External parties were categorised into MARVEL External Advisory Boards (EAB) and external stakeholders. The MARVEL Advisory Board (AB) and Ethics Advisory Board (EB) consisted of independent experts providing valuable feedback on project achievements. External stakeholders, encompassing public authorities, Traffic Managers, Law Enforcement Agencies, data scientists, engineers, architects, project managers, and citizens, contributed insights on smart city technologies.

Feedback was collected after every MARVEL framework release at partner-organised events. Notable events include meetings with AB and EB (one meeting per year), the MARVEL Info Days that took place after each release of the framework, and several other events organised by the MARVEL partners. The subsequent sections detail the feedback collected after the submission of D6.2 at M22 and MARVEL's actions in response.

15.1 MARVEL External Advisory and Ethics Boards

The MARVEL External Advisory Board (EAB) consists of the MARVEL Advisory Board (AB) and the MARVEL Ethics Board (EB). The main task of the MARVEL Advisory Board (AB) was to offer external, independent analysis and recommendations concerning the project's achievements, bringing valuable expertise to ensure the full realization of MARVEL's objectives. Similarly, the primary responsibility of the EB members was to supervise, offer guidance, assess, and, where necessary, bring concerns to the attention of the PC and consortium partners regarding potential ethical considerations within the project, with a special focus on the processing of personal data (PD).

The MARVEL Advisory Board consists of four (4) independent members external to the MARVEL consortium:

Table 48: External Advisory Board members

Member	Position	Company
Yannis Theodoridis	Professor	University of Piraeus, Greece
Mr. Darko Pekar	CEO	Alfanum doo Novi Sad, Serbia.
Mr. Antoine van Ruymbeke	President	MILSET Europe
Mrs. Nuria de Lama (joined at M17)	Consulting Director	IDC4EU

The MARVEL EB is composed by three (3) independent members external to the MARVEL consortium:

Table 49: External Ethics Board members

Member	Position	Company
Dusko Martic	Head of Content Protection and DPO	United Media Group
Anastasiya Kiseleva	Data Protection Expert & Legal Researcher	Vrije Universiteit Brussels
Julia Pshenichnaya (left at M17)	Cognitive process automation practice lead	IBM France
Dr. Ghasan Bhatti (joined at M18)	Director & IoT-NGIN Project Coordinator	Capgemini

The MARVEL EB is assisted by three consortium members:

- **Mr Djordje Djokic** (PN), MARVEL's Data Protection Manager (DPM);
- **Prof. Dragana Bajovic** (UNS), MARVEL's Scientific & Technical Project Manager (STPM);
- **Ms Despina Kopanaki** (FORTH), member of the MARVEL's coordination team and WP9 Leader.

As reported in D6.2, during the first year, MARVEL organised two special events with its advisory boards. On April 28, 2021 the first MARVEL EB meeting was successfully held, bringing together the MARVEL consortium and the three EB members. Subsequently, on December 16, 2021, the MARVEL Advisory Board (AB) kick-off meeting took place in an online format. Valuable feedback was received during these meetings, as elaborated in D6.2, accompanied by a comprehensive report of the consortium's subsequent actions in response to the feedback.

On September 1st and 2nd, 2022, Ms. Nuria de Lama, a member of the MARVEL AB since M17, was invited by the PC to participate in the review preparation meeting held in Heraklion, Crete. The gathering occurred two weeks before the official Interim Review meeting of the project. During the meeting, the consortium, alongside Ms. de Lama, actively collaborated on all review materials, including presentations for various sessions and demonstrations of pilot use cases and individual components. Ms. de Lama, leveraging her extensive experience in RIA projects, provided insightful perspectives and valuable feedback over the two days, significantly aiding partners in refining the materials for the impending review. Valuable feedback received on the exploitation and long-term sustainability of the MARVEL framework for the second reporting period as summarised in Table 50 below.

Table 50: MARVEL AB Feedback – M21

<table>
<tr><th>Topic</th><th>Question/Comment/Suggestion</th><th>Reply summaries during the AB meeting from the consortium</th></tr>
<tr><td>Market Analysis</td><td>A very well documented report that aligns with review specifications. However, it is generic and provides limited utility for the exploitation of MARVEL results. No real competitor Analysis presented since the is no actual comparison with MARVEL offerings and Benchmarking. SWOT and PEST analysis are both generic and require updating to offer more relevant insights.</td><td>During Y3, under ITML's leadership, MARVEL strategically refined its approach for commercialization and societal impact. A thorough market readiness assessment for 16 components recognised by the Innovation Radar was performed to evaluate market preparedness and gather cost and effort estimations. This effort was complemented by the collection and development of Business Model Canvases for these components.</td></tr>
<tr><td>Proposed Business Models</td><td>Certain models presented are impractical to follow in practice, as they are described in a generic manner. The lack of clarity on choosing one model over another can impact decision-making in the development process. No clear path provided.</td><td rowspan="2">Collaboration with Horizon Results Booster services further strengthened exploitation roadmaps and business strategies, resulting in detailed market analyses and tailored financial models for the MARVEL Framework.
In D7.6, the exploitation results of the MARVEL framework along with the 15 components are presented. The deliverable also includes competitor analysis and SWOT and PEST updated analyses.</td></tr>
<tr><td>MARVEL Components</td><td>From an exploitation point of view, the way that the components are presented is generic and offers limited practical value for commercial use.</td></tr>
<tr><td>Standards</td><td>What is meant in KPI 16.1 by "identified standards"? Are there plans to implement them, comply with them, or consider contributing to them? Additionally, is there a strategy to promote MARVEL as the ideal solution aligned with some of these standards, potentially creating a marketing advantage for the final solution based on standards?</td><td>KPI16.1 has been achieved. Numerous components have benchmarked their outcomes from R1 and R2 against specific benchmarks and the SOTA in their respective fields, such as inference models. In this context, more than four benchmarks or shared datasets, including LEAF, UrbanSound8k, MOT, CelebA, etc., have been utilized for performance assessment. This has been achieved through participation in challenges, such as DCASE, or leveraging existing datasets like DISCO. Furthermore, a comprehensive white paper has been released titled “An overview of standards involved in smart cities in the context of MARVEL”[26].</td></tr>
<tr><td>iKPI related to SMEs</td><td>How is the participation of SMEs in providing data ensured?</td><td>Both KPI-O5-E3-1 and iKPI14.1 have been achieved. Through several info days and dissemination/ communication events, a community of around 27 organisations has been established that seems willing to utilise the work done. This includes academic partners from universities, colleges, and research institutes, as well as industries and 8 SMEs. Moreover, the city of Gozo agreed to contribute to the Corpus initiative and provide data for research purposes. More information can be found in D2.5.</td></tr>
<tr><td>Synergies</td><td>While there are numerous collaborations with Big Data projects, engagements with Smart City projects should also be explored.</td><td>MARVEL has initiated collaborations with the EVEREST[27] and EMERALDS[28] EU-funded projects, resulting in a series of collaborative activities. Additional details can be found in D7.7.</td></tr>
</table>

[26] https://doi.org/10.5281/zenodo.10276358

[27] https://emeralds-horizon.eu

[28] https://everest-h2020.eu

Exploitation	Consider all regulations, not just GDPR. Numerous acts related to Data Governance and Data Platform Operation exist. Currently, Data Spaces is a prominent topic in the European Commission, and all the mentioned activities will have an impact on exploitation and decision-making.	MARVEL actively participated to all the activities organised by ICT-51 group of projects under the umbrella of EUH4D[29]. During the organised workshops, MARVEL had the opportunity to present how it contributes to the Data Spaces.

To mark the project's conclusion, a joint final meeting of the AB and the EB was held online on October 20, 2023. This three-hour session aimed to showcase the project's achievements and gather valuable feedback from all key stakeholders. While unforeseen circumstances prevented the full member attendance originally anticipated, Prof. Yannis Theodoridis, Ms Nuria de Lama, and Ms Anastasiya Kiseleva provided their insights. However, after the meeting, all the necessary material, i.e., presentations, deliverables, meeting minutes, etc. were shared with all the members. Table 51 below summarises the feedback received along and the specific actions taken by the consortium to address each point.

Table 51: MARVEL EAB Feedback – M34

Question/Comment/Suggestion	Reply summaries during the AB meeting from the consortium
Send the KPIs to the EAB (Prof. Theodoridis) to check and provide feedback	KPIs were sent and feedback was provided.
Extreme-scale analytics. The amount of data is not clear. Is MARVEL able to handle massive amounts of datasets?	D5.7 which was submitted at M36 showcased the capabilities of the MARVEL framework for large-scale deployments, as well as extending it with third-party components. The deliverable reported on the scalability and extensibility requirements for a large deployment in a Smart City and the approach considered for scalability and extensibility. At the time of the EAB meeting, the deliverable was in progress. The feedback from the EAB was taken into consideration.
Custom-fashion per use case. Make sure that you present it as an advantage of your solution.	One of the core advantages of the MARVEL framework is its flexibility and adaptability to any smart city that has IoT devices such as cameras and microphones. With a rich portfolio of 35 components, MARVEL is able to tackle a wide variety of challenges that smart cities are facing. The presentations were tailored to present this advantage.
Present in a slide which components were used in which use cases? To prove that there is a balance in the use cases and the number of components used per use case	This slide was incorporated in the Scientific & Technical presentation of the MARVEL solution. The usage of all the MARVEL components across several use cases is showcased.
Make sure the data from the DMP are aligned with those of the Corpus. Also what are the datasets you leave behind	This exercise was performed during the preparation of the RP2 Technical Report and the DMP was updated accordingly to reflect all the datasets that were used within the project.
Make sure it will be clear how competitive these technologies/components are. Are they really needed in the use cases? If the use cases were more complex, would they be able to handle the complexity?	The EU Innovation Radar recognised 16 of MARVEL's innovations as excellent, marking a significant achievement for the project. Comprehensive exploitation plans were developed (D7.6 delivered at M36) for each innovation, including analyses of alternative solutions, unique value propositions, selling points, target markets, adopters, competitors, use models, timing, etc. MARVEL demonstrated its ability to address diverse challenges by constructing several components from scratch to address pilot-defined problems within the use cases. The robust research and technical team

[29] https://euhubs4data.eu

	positions MARVEL to further develop components as needed in the future.
Pilots: Need to convince that you have reached the business KPIs. Make part of the presentation	The evaluation of the MARVEL use cases took place very close to the end of the project due to the severe circumstances that the project faced. All the KPIs (functional, non-function, etc.) are reported in this deliverable. Moreover, during the review preparation, the pilots included all the necessary KPIs after each use case to prove the impact of the use cases and the MARVEL framework as a whole.

15.2 Feedback from external stakeholders

This section provides summaries of the feedback received during project info days, workshops and conferences from external stakeholders.

15.2.1 ICSC2022 (26-29.09.2022) MARV: Multimodal and AI-Responsible data processing and deliVery in smart cities

FORTH, UNS, and AU, as MARVEL partners, jointly organized a full-day Workshop on "Multimodal and AI-Responsible data processing and delivery in Smart Cities (MARV)" at the IEEE International Smart Cities Conference (ISC2) 2022 in Paphos, Cyprus, held from September 26-29, 2022. MARV Workshop aspired to the convergence of a set of technologies in the areas of AI, analytics, multimodal perception, software engineering, HPC as part of an Edge-Cloud Computing Continuum paradigm that goes beyond traditional Big Data, conventional architectures heavily capitalizing on distributed resources and heterogeneous data sources in smart city environments, while implementing privacy preservation techniques at all data modalities and at all levels of its architecture. The ultimate objective was to highlight data-driven real-time application workflows to enable fast and accurate insights and optimised decision-making in modern cities, showcasing the potential to address societal challenges very effectively, from increasing public safety and security, optimising energy consumption from a multimodal/multidomain perspective to analysing traffic flows and urban mobility for effective city planning.

The Workshop featured two keynote talks by Prof. Petar Popovski from Aalborg University, a world-leading expert in Communication Theory, Wireless Communications, Internet of Things, and Smart Grid Communications, and MARVEL partner Farhan Sahito (PN), a world-leading expert in international privacy compliance. Additionally, two panels comprised experts like Prof. Petar Popovski, Dr. Farhan Sahito, Prof. Dejan Vukobratovic, and MARVEL partners Dr. Alessio Brutti (FBK), Prof. Dragana Bajovic (UNS), Prof. Alexandros Iosifidis (AU), Dr. Claudio Cicconetti (CNR), and Mrs. Despina Kopanaki (FORTH). The Workshop also included presentations of six contributing papers.

The workshop was attended by 20-30 participants which corresponds to a large percentage of the overall (physical) conference participation. Engaging discussions took place throughout the day, involving several researchers. Notably, discussions with two municipalities were a highlight, with expressed interest in the MARVEL framework. It was emphasised that a municipality, when introducing new services for citizens, lacks the privilege of failure, necessitating thorough an extensive period of testing before official release. After the event, one meeting was organised, where MARVEL technologies and the framework were presented to the representatives of the city of Lugano.

15.2.2 Vision zero – Belgrade

MARVEL project partners (ZELUS, ITML, INTRA, UNS and GRN) participated in the Vision Zero for the Balkans Road Safety Conference 2023 on May 15th in Belgrade, Serbia. UNS presented MARVEL's achievements during the IRF Master Class 1, titled "Making AI Work for Road Safety". In parallel, MARVEL held a booth at the conference with a monitor screen demonstrating the project's technologies and use cases. More than 200 people attended the conference, most of which were originated from the public sector and transportation domain. During the Master Class, attendees gained insight into how the project utilises its developed technologies for real-time traffic management, public safety, environmental monitoring, and resource optimisation in urban environments. By the end of the event, our representatives were able to engage with more than 30 individuals who showed interest in the project. Although this is considered a dissemination activity that is thoroughly reported in D7.7, our engagement with attendees made clear their positive feedback on their first look at MARVEL, and that is also demonstrated by the fact that during the event, there was a significant increase in MARVEL website and social media engagement.

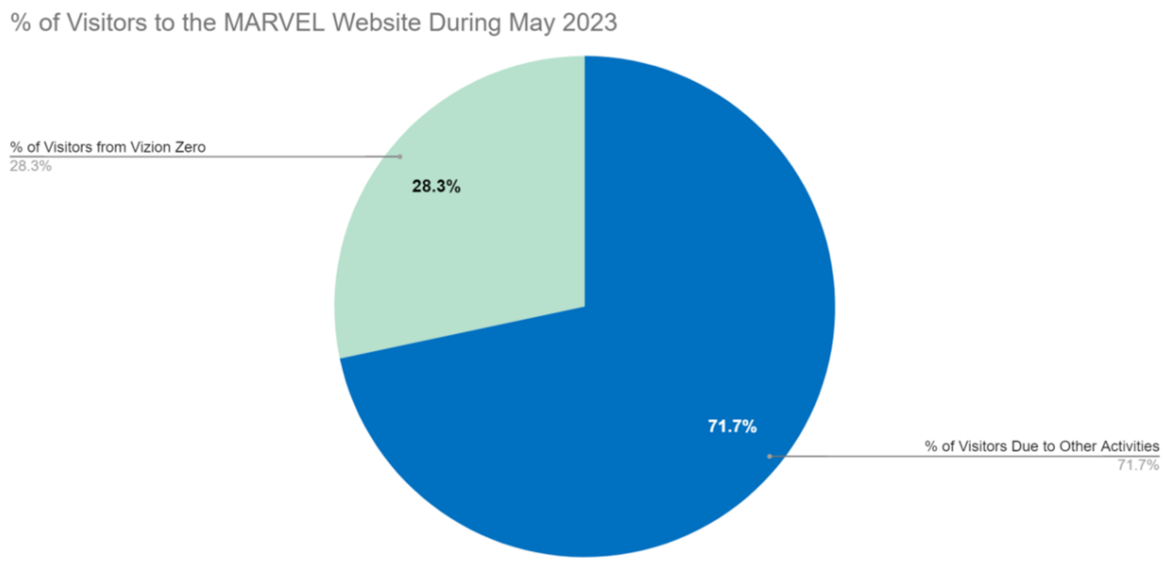


Figure 25: Visitors to the MARVEL Website during May 2023

The Vision Zero conference provided a significant opportunity for the MARVEL project to reach out to the broader community involved in road safety, particularly in the Balkans region. The feedback and connections made during the conference will no doubt provide further momentum to MARVEL's mission of transforming urban environments with extreme-scale, AI-driven analytics. Looking ahead, MARVEL continues to lead in developing and applying extreme-scale analytics and AI for smart cities, driven by the mission to make urban environments safer, more efficient, and more sustainable.

15.2.3 EBDVF2022 (21-23.11.2022)

MARVEL sponsored the European Big Data Value Forum 2022 (EBDVF 2022) that was held on November 21-23, 2022 in Prague, Czech Republic. The event was organised by the Big Data Value Association (BDVA), IT4Innovations (VSB – Technical University of Ostrava), Plan4all and Charles University in Prague in collaboration with the European Commission (DG CNECT). The aim is to bring together industry professionals, business developers, researchers and policymakers from Europe and other regions of the world to advance policy actions, and industrial and research activities in the area of data and AI. The special theme of the 2022 event was "At the Heart of the Ecosystem for Data and AI".

MARVEL was present with a physical booth. The objective was via the booth to communicate and share more information about the MARVEL project, its technologies and the real-life pilot use cases and find more stakeholders and smart cities. The booth attracted more than 100 people and partners managed to engage a great portion of them to the next activities. It is worth noting that those who visited our booth were already familiar with the MARVEL project, demonstrating a level of awareness, and they came with specific questions about our technologies and upcoming plans.

In summary, MARVEL's participation in EBDVF 2022 has the potential to positively impact its visibility, stakeholder engagement, and collaborations within the data and AI community, contributing to the project's overall success.

15.2.4 Second Info Day in Malta (28.11.2022)

The scope and objectives of the MARVEL 2nd Info Day, held on November 28, 2022, in Valletta, Malta were to invite individuals interested in the project to showcase its progress. The event aimed to inspire potential early adopters and investors, identify societal and expert concerns about the project, inform the public, and encourage academic research. Additionally, the event was seeking to facilitate connections among attendees engaged in similar work and ensure effective dissemination of information.

During the info day, an electronic questionnaire was organized on site to receive feedback from the attendees and the results revealed diverse interests among the participants, with 66% expressing interest in research and academia and notably, 37% are already implementing big data analytic solutions.

The most voted valuable innovation in the MARVEL framework was Audio-visual analytics-AI, with citizens emphasising safety, privacy, and user-friendly interfaces. City municipalities prioritized data analysis, management, traffic control, and security.

In terms of citizen preferences, functionalities like mobility justice, road traffic monitoring, and faster response from authorities were highlighted. City municipalities expressed interest in automation, real-time decision-making, and surveillance support. Challenges included privacy concerns and technological considerations.

Concerning long-term data, citizens and city municipalities saw its value in improving urban planning, traffic analysis, and identifying behavioural patterns. The MARVEL framework was perceived to benefit most the police force and public administration.

Regarding automation in traffic control rooms, functionalities like automated event reporting, live updates, and efficient alert systems were underscored. The ability to revisit events and manage information in a user-friendly manner is crucial for quick responses.

For drone applications, slow mobility traffic monitoring, crowd guidance, and trash management were seen as impactful, enhancing safety and security in various scenarios.

Privacy and security were voted as the most relevant challenges for applying MARVEL, and the majority agreed that the framework respects privacy. However, concerns about scaling, robustness, ethics, and computational resources were raised.

Some suggested improvements included incorporating social media inputs and addressing citizen monitoring concerns.

In summary, the Malta Info Day highlighted diverse interests, positive feedback on MARVEL's potential, and considerations for privacy and technology. Long-term data, control room functionalities, and drone applications emerged as integral aspects of the project's success. In addition, the Ministry of Gozo and the Gozo Regional Development Authority were highly interested in the project and eventually decided to adopt the framework on a specific case.

15.2.5 EBDVF2023 (October 2023)

For the third year running, MARVEL sponsored the European Big Data Value Forum (EBDVF), a cornerstone event for the data-driven community. On October 26, 2023, the project organised a dedicated workshop, titled "Empowering Smart Cities with Multimodal Extreme-Scale Data Analytics," with the active collaboration of UNS, ZELUS, ITML, GRN, INTRA and MT.

The goal of the workshop was to provide a comprehensive overview of MARVEL's goals and achievements. Professor Sotiris Ioannidis presented the project's vision, followed by in-depth talks from Dr. Vivian Kiousi and Dr. Nikola Simic showcasing its innovative solutions. The session then delved into practical applications with presentations and demonstrations by Professor Adrian Muscat and Mr. Thomas Festi on use cases developed for Malta and Trento. Finally, Mr. Tassos Kanellos and Mrs. Stella Markopoulou delivered a live platform demonstration, followed by a lively Q&A session moderated by Mrs. Despina Kopanaki.

With over 30 active participants, the discussions at the end of the session covered diverse topics, such as data utilisation, bias mitigation, model integration, scalability, and market readiness. The audience's keen interest and insightful questions reflected the potential of MARVEL to transform urban environments.

15.2.6 Smart City Expo (Nov 2023)

As part of the project's final event, organised by FORTH, MARVEL participated at the Smart City Expo World Congress 2023, held from November 7-9. Recognised as a premier platform for urban innovation, this event provided an excellent opportunity for MARVEL to demonstrate its commitment to enhancing smart cities through advanced multimodal data analytics.

MARVEL's impactful presence was showcased through its booth, where attendees were welcomed to explore innovative solutions designed for smart cities. MARVEL's booth was a hub for showcasing cutting-edge data analytics solutions and engaging with experts. Live demonstrations gave attendees first-hand experience of the technologies poised to transform urban environments. Additionally, MARVEL's online e-booth provided comprehensive information and guidance, extending accessibility beyond the physical event.

The booth attracted over 150 participants, representing diverse backgrounds such as smart cities, municipalities, policymakers, researchers, transport managers, consultants, and students. Notably, attendees expressed particular interest in the MARVEL AI subsystem and the use cases developed within the project. Demonstrations played a crucial role in facilitating a deep understanding of MARVEL's capabilities. The positive impact continued after the Expo, with numerous invitations extended to contacts made during the event, encouraging them to attend MARVEL's subsequent 3rd Info Day.

15.2.7 Third Info Day – online (4.12.2023)

The MARVEL 3rd Info Day, held on December 4, 2023, was a pivotal gathering for Smart Cities professionals, enthusiasts and stakeholders. During this engaging online event, distinguished experts in the field showcased MARVEL's main achievements, and participants delved into the benefits of the MARVEL project, experiencing its functionalities through real-world applications. The event was attended by 47 participants in total and it facilitated knowledge sharing and brought together representatives from four forward-thinking cities and municipalities around Europe, specifically from Belgium, Romania and Greece. Furthermore, participants seized the opportunity to network, fostering collaborative discussions and idea exchange.

The Info Day participants were also requested to provide their valuable feedback and insights by taking a brief survey that was presented in the context of the Info Day. The survey was based on a structured questionnaire, containing 9 questions that were carefully selected to obtain information about the following aspects:

- User acceptance of the MARVEL solution (Q#1, Q#2)
- Validation of the MARVEL value proposition (Q#3, Q#4, Q#5)
- Usability evaluation of the MARVEL framework (Q#6)
- Implementation barrier/risk assessment (Q#7)
- Market assessment (Q#8, Q#9)

In total, 19 responses were received from the 3rd Info Day participants.

The individual questions are presented in detail along with the received respective responses and their analysis in D7.6. The main conclusions that can be drawn from this analysis are listed below:

- There is a very high acceptance rate of the MARVEL solution among the respondents (89%).
- A substantial majority of respondents (74%) find that MARVEL sufficiently addresses data privacy issues. However, such issues are also recognised as the most important concern in the implementation of MARVEL and should therefore remain as a main focus area for potential improvements in the future.
- The main audience that can be addressed by MARVEL includes (a) transport and traffic management authorities and (b) law enforcement agencies.
- MARVEL should primarily focus on use cases related to densely crowded areas and events and to detecting road accidents, while the detection of criminal/anti-social activity in remote and sparsely populated areas does not appear very prominent as a use case by comparison.
- The technical aspects of the MARVEL solution that can bring the highest added value include data management, AI, Data Corpus-as-a-Service and its potential for scalability, while the ones that are seen as less relevant to the needs of Smart Cities are the solution's potential for customisation and the use of multi-channel microphones.
- There is a very high approval of the usability of the Decision Making Toolkit (SmartViz) of the MARVEL framework among respondents.
- Future endeavours should be made towards ensuring that the costs of the computational infrastructure required to implement MARVEL are realistic and that relevant concerns should be addressed by marketing strategies.
- Efforts should be made via marketing campaigns to effectively communicate the added value of MARVEL so as to overcome any lack of expertise in organisations that may potentially adopt MARVEL, while continuing to optimise the usability of the solution at the same time.

- While not seen as prominent potential obstacles, the effectiveness of AI components, scalability issues, the cost of the solution and data collection methods should remain as points of attention.
- MARVEL is targeting a new market with little competition that is seen as an opportunity for marketing the solution.
- There is a substantial critical mass of potential adopters of the MARVEL solution that justifies the efforts and investments to release it to the market.

16 Summary and Conclusions

This document reported the final assessment and impact analysis from a use case point of view, including whether the components satisfy the use case technical requirements, whether end-users perceive the tools developed as useful in practice, i.e., improve business processes and to what extent the use cases implemented can potentially address societal challenges.

The framework is useful if the technical components meet the use case requirements. The MARVEL framework consists of several components that interact to deliver a dynamic E2F2C infrastructure. The feedback from the External Advisory and Ethics Board has helped in shaping some of the component requirements. The performance of most AI components was measured during both RP1 and RP2, first against industry standards benchmarks and then on real-world data obtained from the use cases and under standard operation on the field. These components which have achieved the intended KPIs process the AV data to obtain information that is useful to the use case. The end-to-end framework latency was measured during RP2. The latency achieved satisfied the requirements of the use cases.

Evaluating the framework from a user point of view required the engagement of professionals who are the target users of the various use cases, i.e., transport planners/engineers, policymakers and law enforcement and security personnel. For this reason, the opinion of the end-users was gathered using formal surveys and interviews. The surveys included opinions on how much useful the tools developed fulfil their intended application. Experts welcomed the framework and could see further examples of where the system can be useful in their professional practice. The opinions gathered can also serve as useful feedback to any future implementation and fine-tuning.

Understanding to what degree societal challenges are addressed is a more difficult exercise since in most cases measurements can only be carried out once a use case or a tool is widely deployed over a significant time period that can span a number of years. Such a task is therefore impossible to achieve due to the limited timeframe of the project and instead evaluations were limited to sampling the perceived impact of the use cases. The first Info Day organised in Trento collected the first feedback from smart city decision-makers, which served to fine-tune the use cases prior to the RP1 integration phase. Following the RP1 integration phase, surveys limited to a couple use cases have surveyed both personnel working in the field of interest, on efficiency and user experience, as well as the public, for their opinion on meeting societal challenges. The opinions gathered in these surveys and events served as feedback to the RP2 integration period. In addition, the 2nd Info Day that took place halfway through the RP2 period consolidated the feedback from smart city and big data external stakeholders. The final impact analysis was carried out towards the end of the project using surveys and interviews. This exercise studied the perceived impact on society in general as well as the perceived impact on business processes. Results strongly indicate that the framework or similar systems are needed to address societal challenges in urban environments.

17 Appendix : Survey Methodologies and Documents

Non-functional KPIs and subjective parameters were evaluated via surveys and interviews. This section summarises the methodology adopted and includes all printable survey documents for each pilot.

17.1 Malta pilot - Introduction to methodologies and documents

To evaluate the use cases in the Malta pilot and assess the impact on society and business processes two methods were used: (a) online surveys, and (b) live interviews. The former is covered in sections 17.1.1-17.1.6, whilst the latter is given in section 17.1.7. The results from both methods are used to complete the evaluation tables related to the Malta Pilot.

17.1.1 Malta pilot – Online Surveys

Five online surveys were prepared. One survey (S5), which considered the Malta pilot as a whole was addressed to the general public, whereas the other four surveys (S1-S4) considered each use case separately and were addressed to professionals working in the transport sector and specific to the respective use case. The list included academic researchers, traffic managers, law enforcement officers, transport and urban planners, and policymakers. Each survey included textual descriptions of for example the use case, a recorded video demo and a number of questions. The surveys were distributed via email. Sections 17.1.2-17.1.6 elaborate on each survey and give a summary of the results.

17.1.2 S1 – survey for GRN1

This survey was intended for the evaluation of the GRN 1 Use Case – Safer Roads by experts within Cyclist NGOs, researchers, urban planners and transport engineers. The survey was mailed to 11 potential respondents and a total of 6 experts replied to the survey.

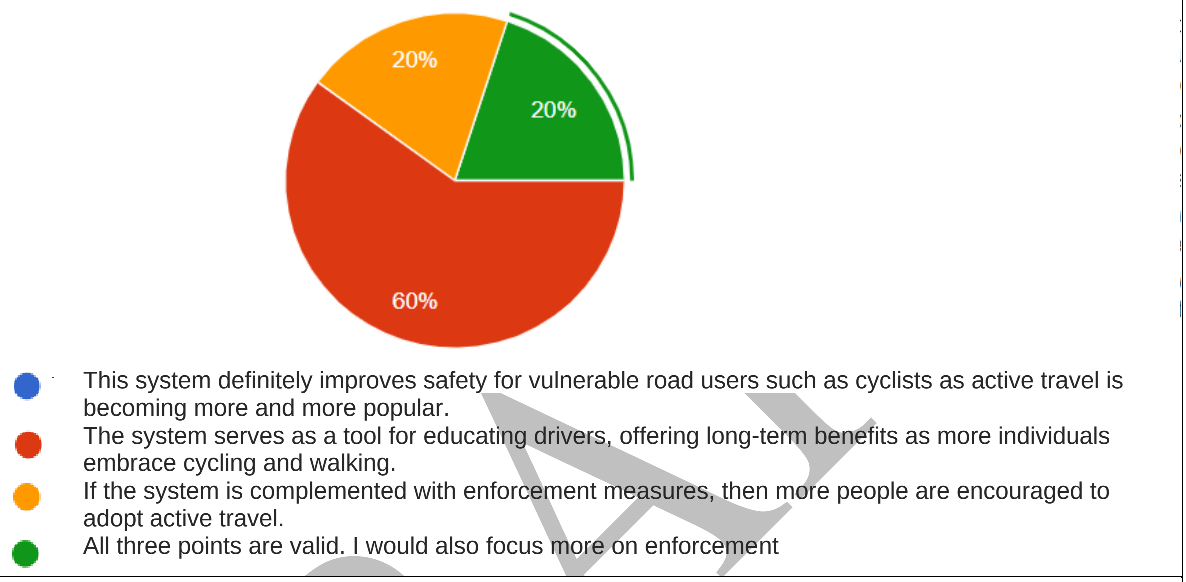
The survey starts with a high-level explanation of the GRN1 use case, followed by questions related to the impact of the pilot. This is followed by a short video of the SmartViz UI and some questions to assess the usability of the UI.

Safer Roads- The Malta Pilot
Impact Evaluation

The video showcased a smart sign detecting cyclists and/or pedestrians, and alerting drivers of their presence, implicitly asking them to take greater care, such as slowing down and keeping eyes on the road.

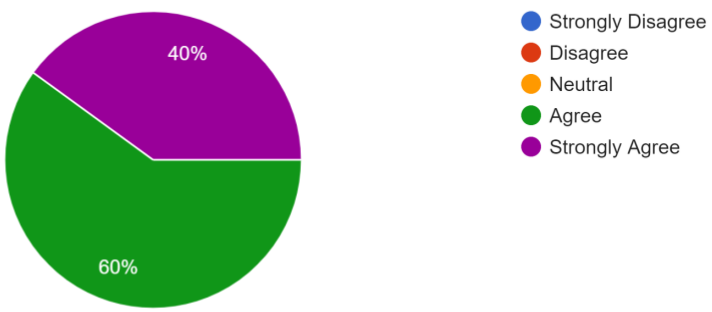
What are your thoughts on this system?

5 responses



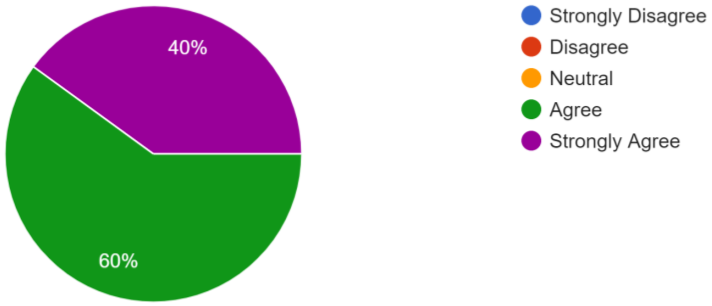
Do you agree that storing bicycle and pedestrian detections, along with timestamps, over an extended period is valuable for planning cycling and walking infrastructure?

5 responses



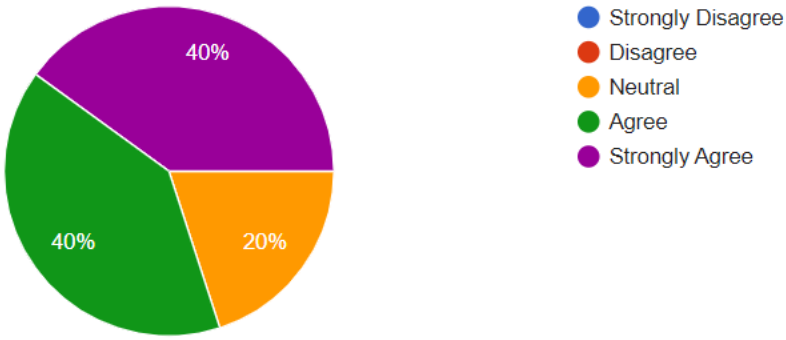
Would you agree that collecting weather data in parallel with the detection of bicycles and pedestrians is beneficial in the study of active travel modes? (for example to discover correlations)

5 responses



Do you believe it is time to explore variable speed limits, adjusting the sign posted speed limits in real-time, based on weather, road conditions, or the presence of vulnerable road users? The aim of this approach is to reduce average trip times while promoting active travel.

5 responses



Would you like to leave any additional comments that will help us in the evaluation?

2 responses

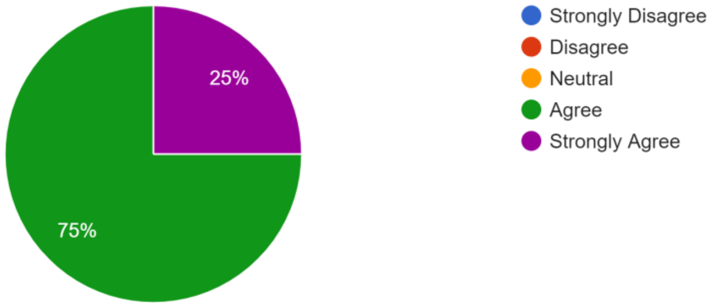
We need more of this technology

With regard to variable speed limits, I would like to explore the concept a bit more. Hence my choice - neutral

User Interface Evaluation

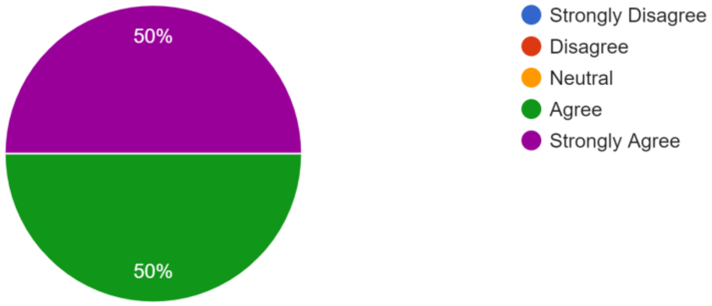
To what extent do you believe that the GRN1-Safer Roads objectives are effectively supported by the SmartViz toolkit, assisting you in achieving the end goal of this use case?

4 responses



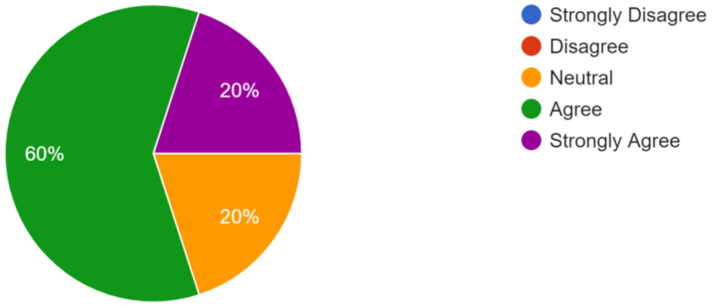
I agree that the summaries widget (shown in the picture below) assists me in monitoring the counts of the respective entities and detections.

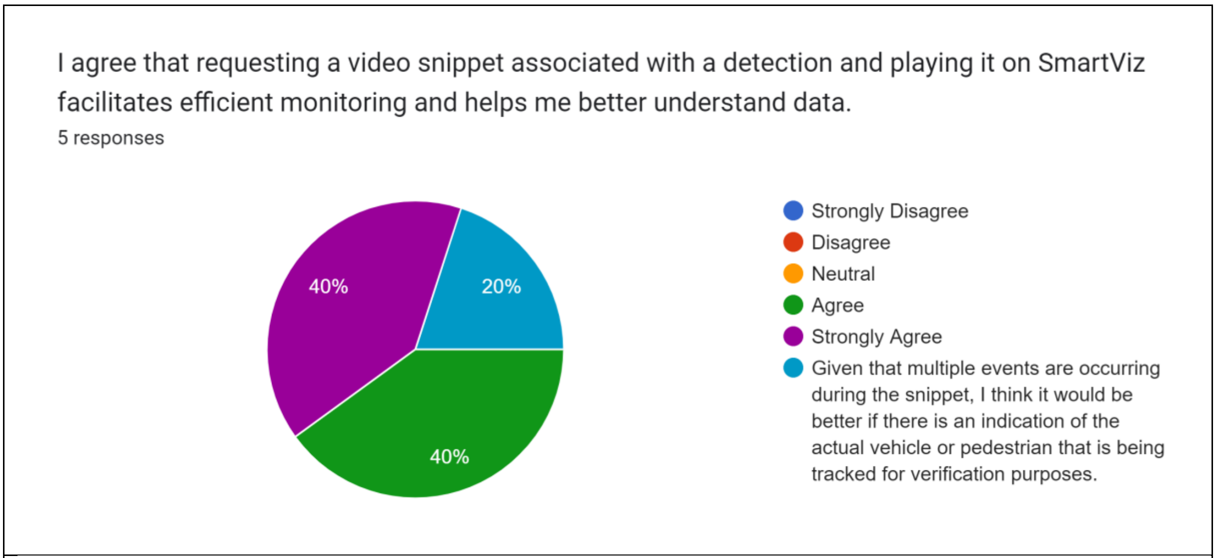
4 responses

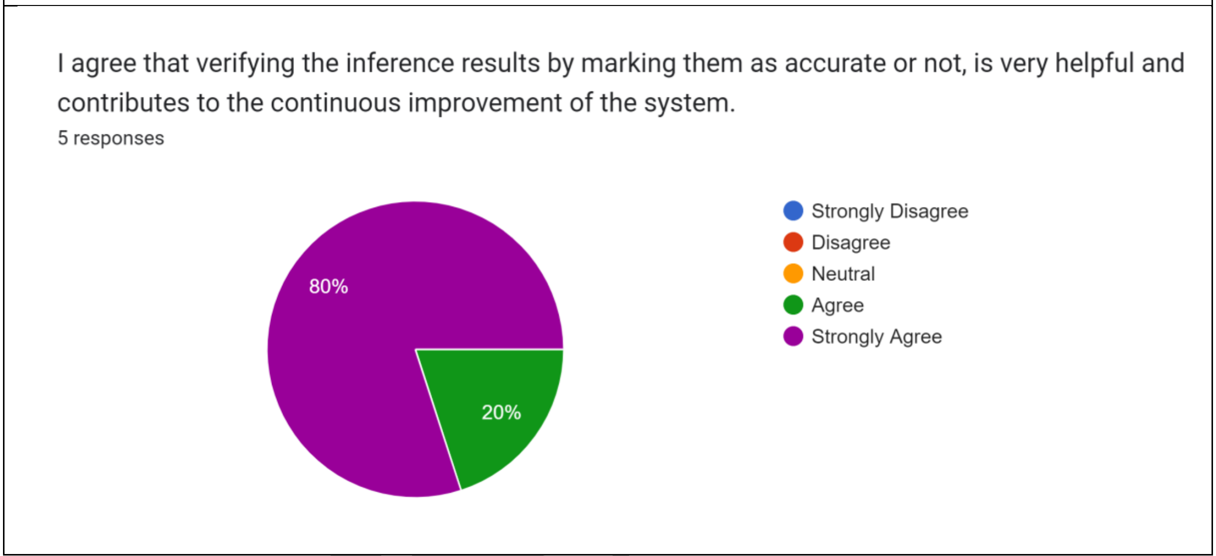


I believe that the alerting functionality (shown in the picture below) notifies me clearly and adequately for a detection in the monitored area along with the relevant information.

5 responses







17.1.3 S2 - survey for GRN2

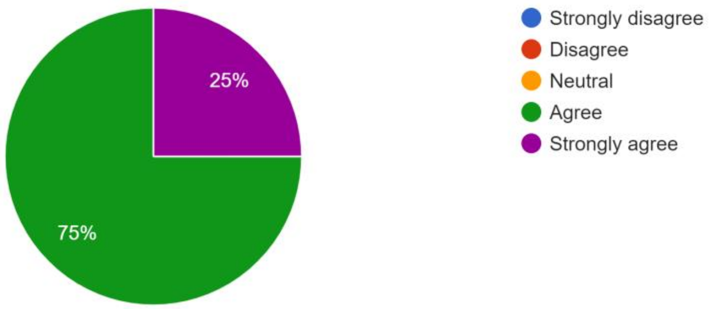
This survey was intended for the evaluation of the GRN 2 Use Case Analysing Road User Behaviour by experts in Transport and Law Enforcement and Road Safety. The survey was sent to 13 potential respondents and a total of 5 experts replied to the survey.

The survey starts with a high-level explanation of the GRN2 use case, followed by a number of questions related to the impact of the pilot. This is followed by a short video of the SmartViz UI and some questions to assess the usability of the UI.

Analysing Road User Behaviour - The Malta Pilot
Impact Evaluation

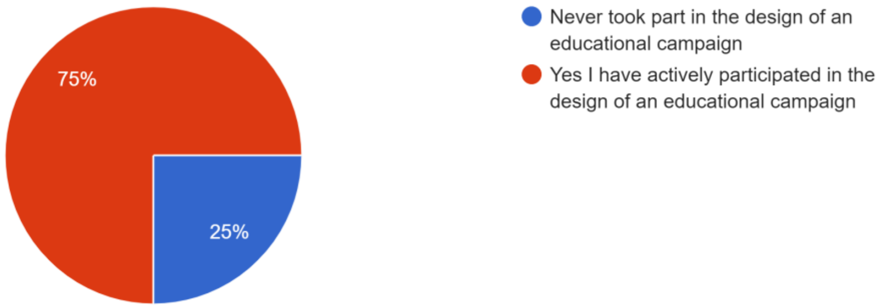
Do you agree that educational campaigns and law enforcement go hand in hand?

4 responses



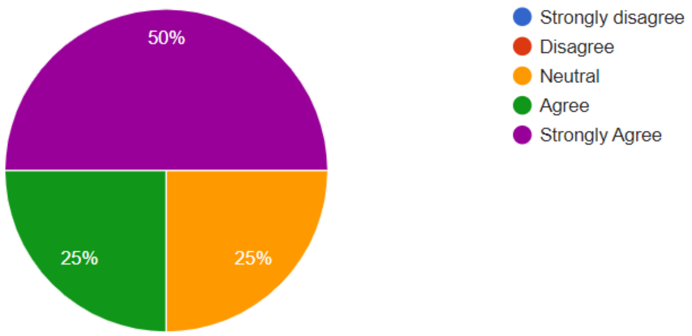
Have you ever been involved in the development of an educational campaign that targets road etiquette?

4 responses



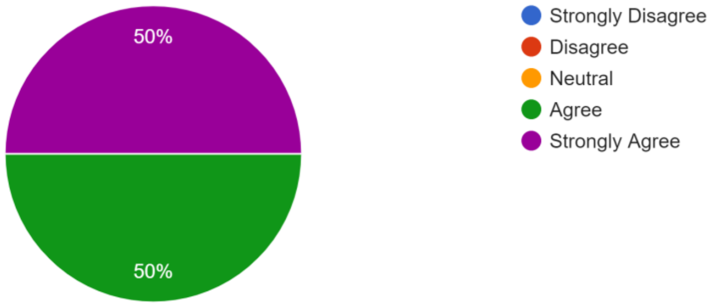
The MARVEL framework can be used to identify (through the AI models) which objectionable behaviours or contraventions are the most common on the road and therefore help in deciding which behaviour to target in an educational campaign, thus maximising the impact. Would you agree that such a feature within the tool is useful?

4 responses



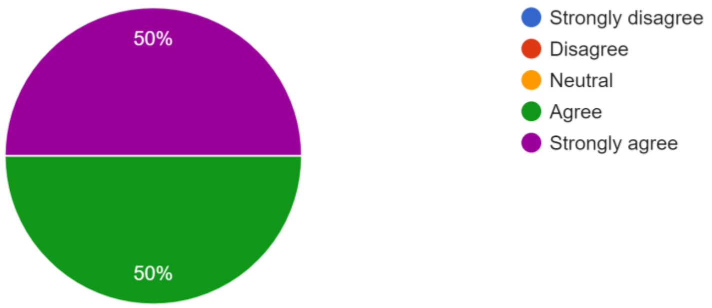
The framework can also be used to evaluate the effectiveness of an educational campaign as demonstrated in the video. Would you agree that such a function is useful in practice?

4 responses



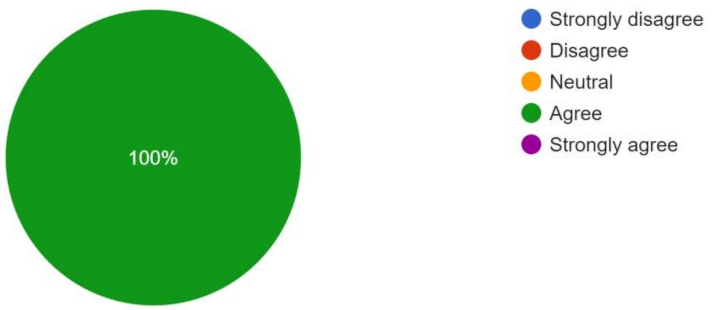
The framework can also be used to evaluate the effectiveness of law enforcement campaigns. Do you agree that such a function is useful in practice?

4 responses



From what you have seen, do you agree that this system is easy to use?

4 responses



Would you like to leave any additional comments or feedback that will help us in the evaluation?

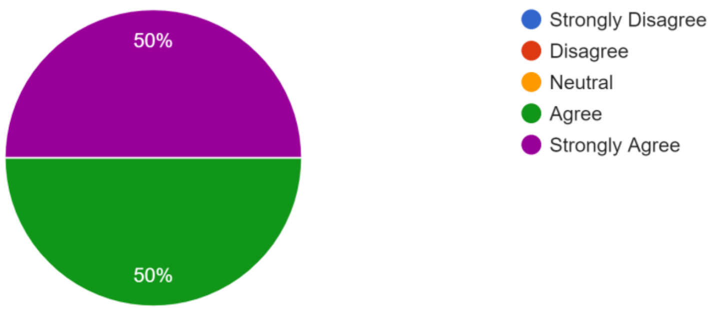
1 response

n/a looks good well done

User Interface Evaluation

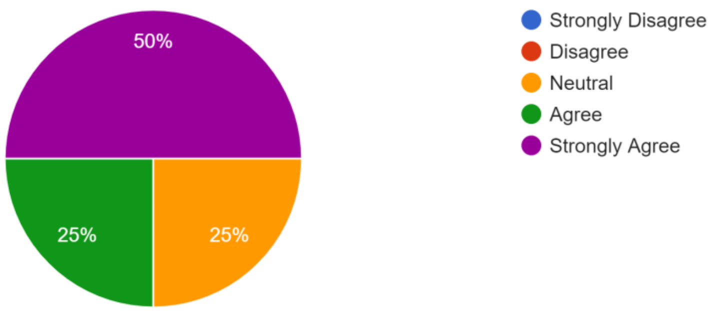
To what extent do you believe that the comparison feature in SmartViz allows and helps you to compare, analyze, and find patterns in the detected events within each time period?

4 responses



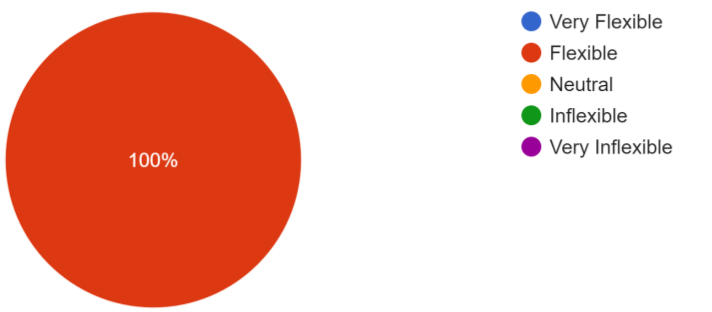
To what extent do you find convenient the option to download the dashboard as a PDF version or in a JSON file for storing data and using them according to your preference for future or offline reference?

4 responses



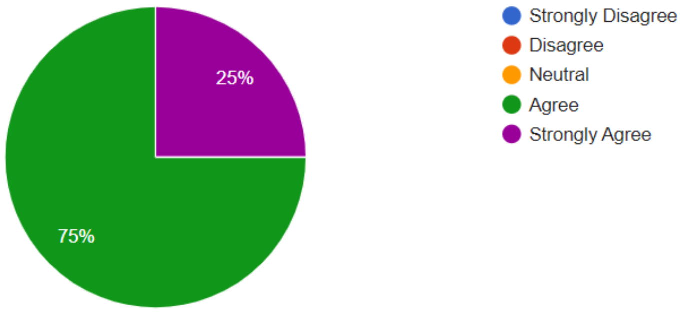
How would you rate the flexibility of SmartViz in allowing you to customize the dashboard according to your needs? Please consider the options to resiz...eorder, and hide widgets for optimal configuration

4 responses



To what extent do you agree that the timeline feature, allowing scrolling and moving across time, with synchronized updates to all other widgets in the dashboard, enhances the ease of data exploration?

4 responses



17.1.4 S3 – survey for GRN3

This survey was intended for the evaluation of the GRN 3 Use Case – Traffic Conditions and Anomalous Events by traffic managers/engineers and control room personnel. The survey was sent to 12 potential respondents and a total of 3 experts replied to the survey.

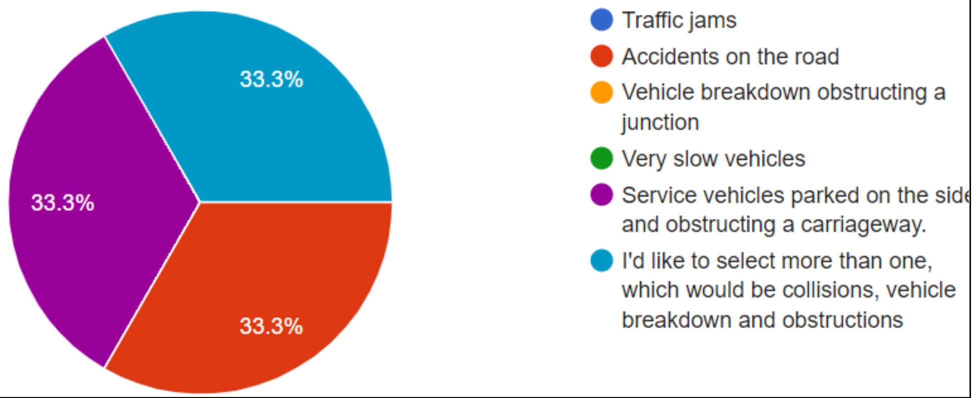
The survey starts with a high-level explanation of the GRN3 use case, followed by a number of questions related to the impact of the pilot. This is followed by a short video of the SmartViz UI and some questions to assess the usability of the UI.

Traffic Conditions and Anomalous Events - The Malta Pilot

Impact Evaluation

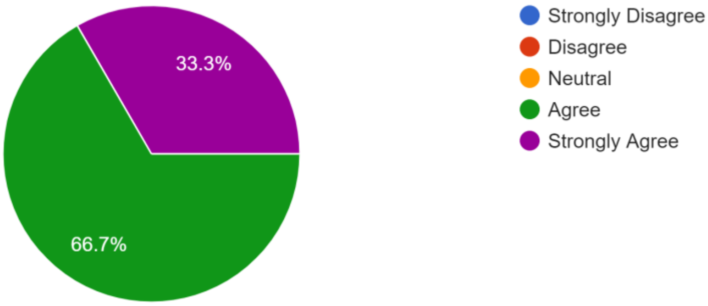
You have just watched a video of a tool that uses AI to detect multiple anomalous events and traffic conditions and therefore can be used as an aid in the traffic control room. Which of the following events would you consider most useful to detect during your everyday work?

3 responses



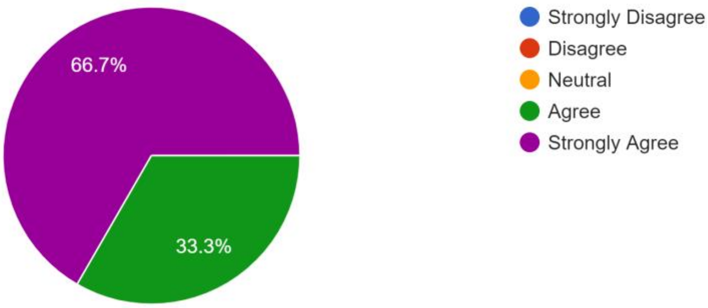
Do you agree that this tool will help you in your day-to-day job, for example, it helps in increasing efficiency?

3 responses



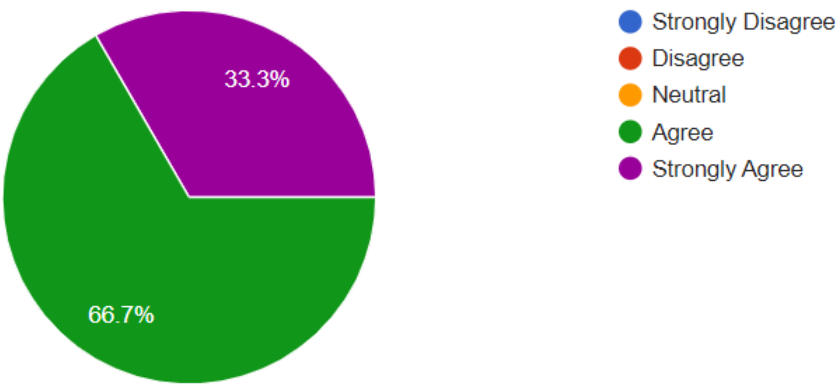
Do you agree that this tool will help you to simultaneously monitor a larger number of locations?

3 responses



In case of a disruptive event, for example, an accident, would you agree that the use this tool would help in detecting the event earlier and thus result in a quicker respons when dispatching police/civil/medical teams to locations?

3 responses



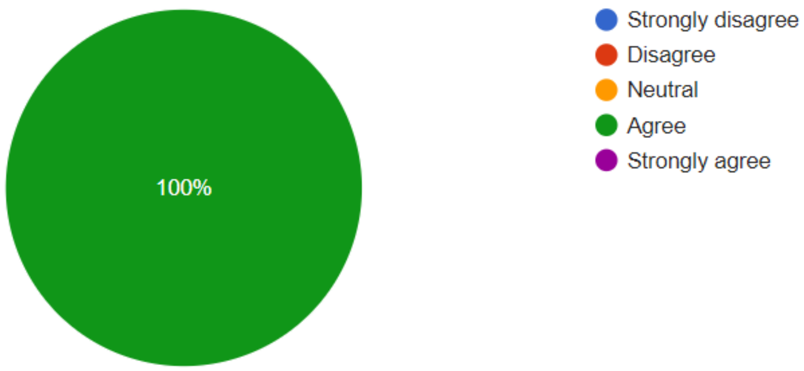
From what you have seen, do you agree that this system is easy to use?

3 responses

User Interface Evaluation

I believe that the added value, usability, and overall effectiveness of the **Real-time Map** widget in showcasing detected anomalies and events within the MARVEL project is high and allows users to interact and monitor from a high level the locations.

3 responses

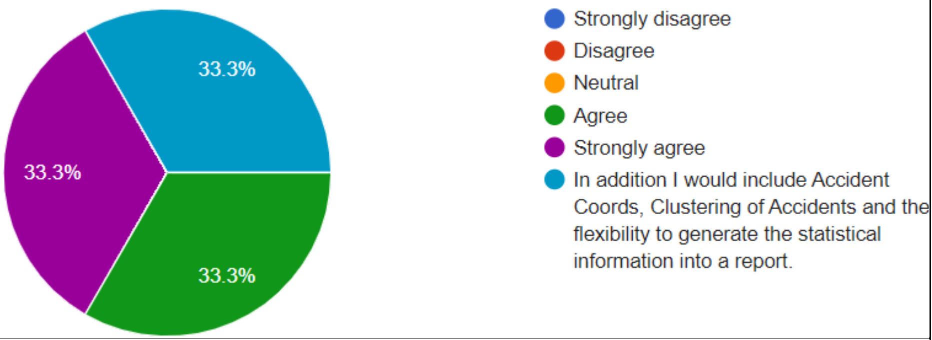


I believe in the usability and functionality of the Video Stream Player widget, considering its ability to allow users to see in real-time the anonymised camera feed, and control the playback for the cameras within the MARVEL project?

3 responses

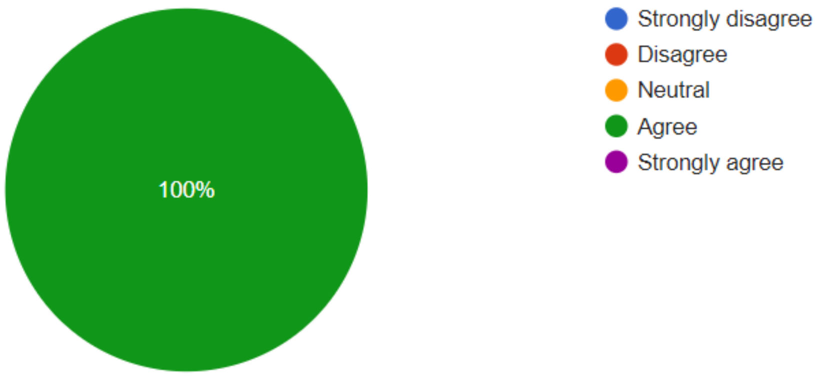
To what extent do you believe that the **statistical** representation widget, coupled with the speed line charts, assists in enhancing your understanding of the data and provides a visual aid for making informed decisions?

3 responses



I agree that the details widget presents the incoming events, anomalies in a concise and clear way for me to drill into more detailed information regarding the detections and proceed with my investigation.

3 responses



17.1.5 S4 - survey for GRN4

This survey was intended to Evaluate the GRN 4 Use Case – Junction Traffic and Trajectory collection by experts in policy, infrastructure engineering, planning and research. The survey was sent to 9 potential respondents and a total of 4 experts replied to the survey.

The survey starts with a high-level explanation of the GRN4 use case, followed by a number of questions related to the impact of the pilot. This is followed by a short video of the SmartViz UI and some questions to assess the usability of the UI.

Junction Traffic and Trajectory Collection - The Malta Pilot

Impact Evaluation

You have just seen a video of a tool which uses AI to collect long-term data, more specifically, counts of vehicles, cyclists and other road users an...the infrastructure, and develop policy in mobility?

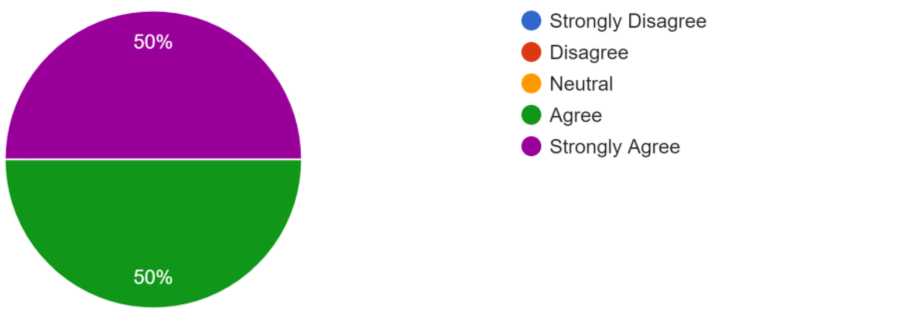
4 responses

For which function do you believe this AI tool would be most useful in practice? Please select one from the list below:

4 responses

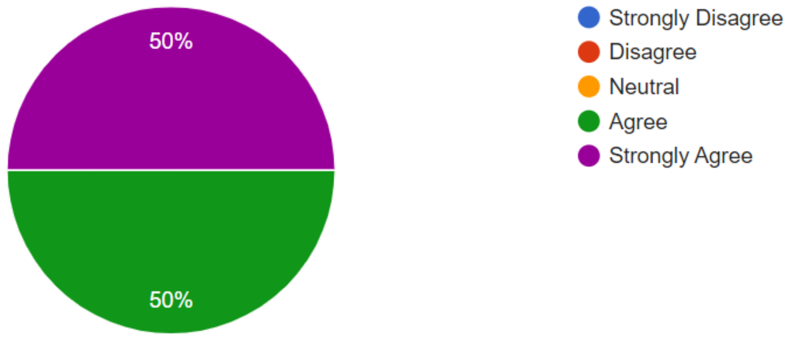
Do you agree that the collection of long term data (when compared to short-term data) covering several months and years would result in better mobility policy decision-making and evaluation?

4 responses



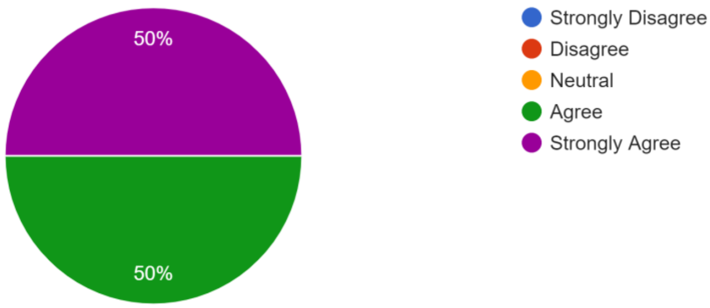
Do you agree that collecting long-term data covering several months or years, in comparison to short-term data, would significantly enhance the decision-making and evaluation process for mobility policies?

4 responses



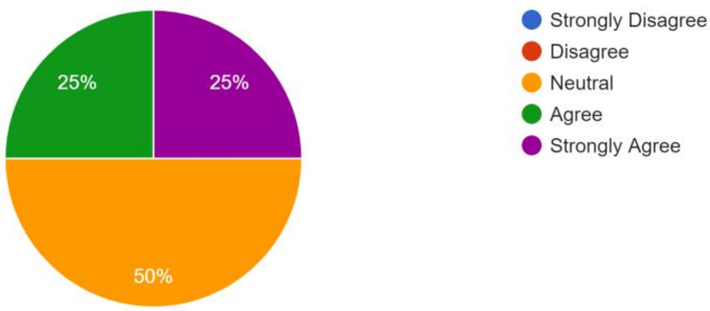
Do you agree that long-term data, when compared to short-term data, is more useful in the design of road infrastructure?

4 responses



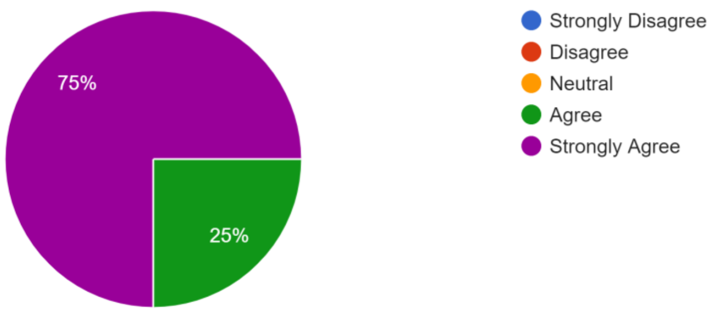
From what you have seen, do you agree that the tool is intuitive and easy to use?

4 responses



Do you agree that collecting data on all types of road users, i.e vehicles, and horse drawn pedestrians and cyclists would help decision makers...d for the needs of a wider variety of road users?

4 responses



Would you like to leave any additional comments or feedback that will help us in the evaluation?

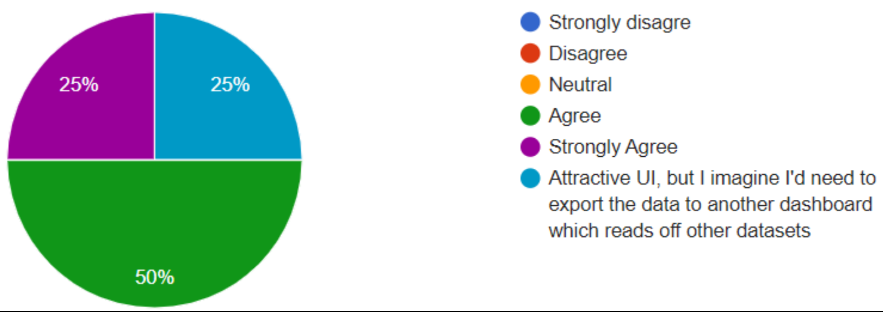
2 responses

Would be a very valuable dataset to complement transport modelling!

User Interface Evaluation

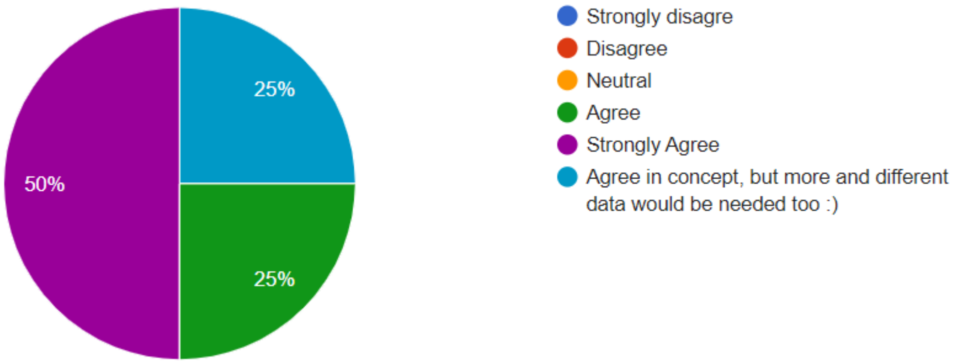
To what extent do you believe the timeline feature, allowing scrolling and moving across time with **synchronized** updates to all other widgets in the dashboard, enhances the ease of data exploration and investigation, especially **in conjunction** with the detailed information provided by the details widget.

4 responses



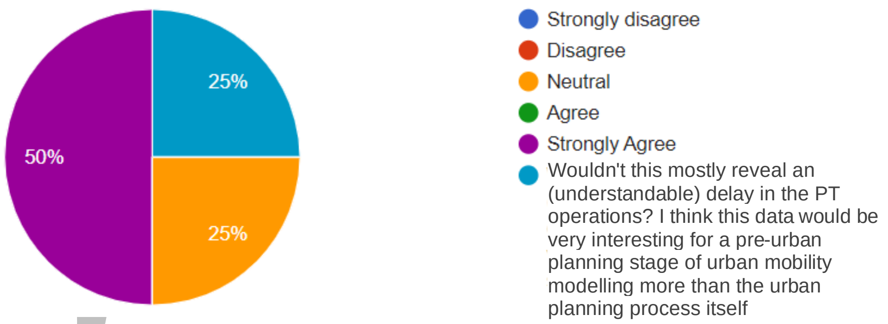
I agree that the trajectories widget, with the ability to **showcase the paths of individual entities** on the road and the option to animate and filter them, contribute to the understanding of road usage for urban planning purposes.

4 responses

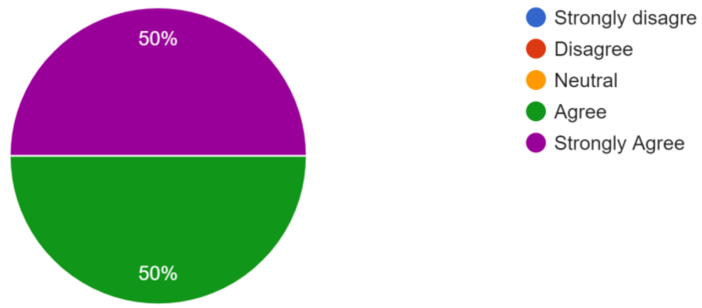


With the addition of the bus schedule to the timeline, we noticed a potential **hidden correlation** between the scheduled bus times and the passing buses with a 5-minute difference to the original schedule. The discovery of a hidden correlation between the inference data and the public buses schedule, contribute to your insights as an urban planner, and it might influence decisions related to road infrastructure planning.

4 responses



I believe that the filtering options in this dashboard, particularly when selecting a specific camera and filtering by vehicle type, facilitates efficient d...xploration and decision-making for an urban planner.
4 responses



17.1.6 S5 – AI Systems for Sustainable Mobility - The Malta Pilot survey

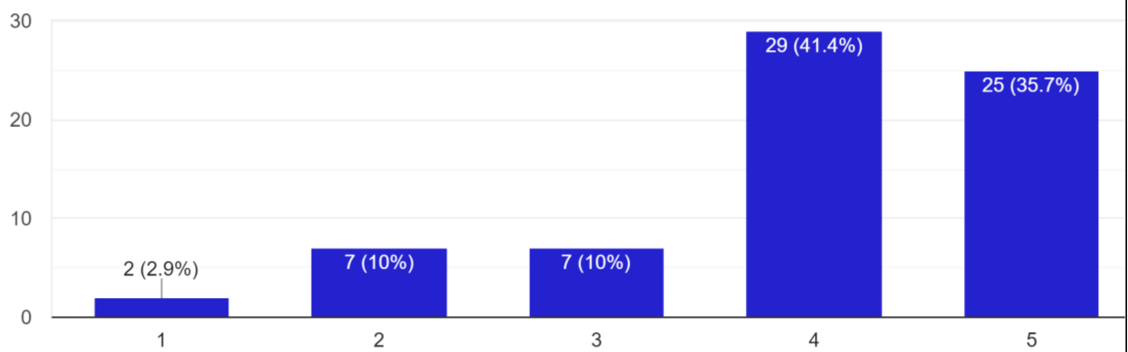
This survey was intended to evaluate the public response to the Malta pilot. The survey was shared with the general public through the MARVEL social media and was then also reshared by the partners on their own pages. The surveys were also shared via email to students and staff via the University of Malta newsletter. In total, 70 responses were submitted.

The survey starts with a general high-level explanation of all the MARVEL pilots. Then for each use case, a brief explanation is given of the use case and specific questions are asked about the use case.

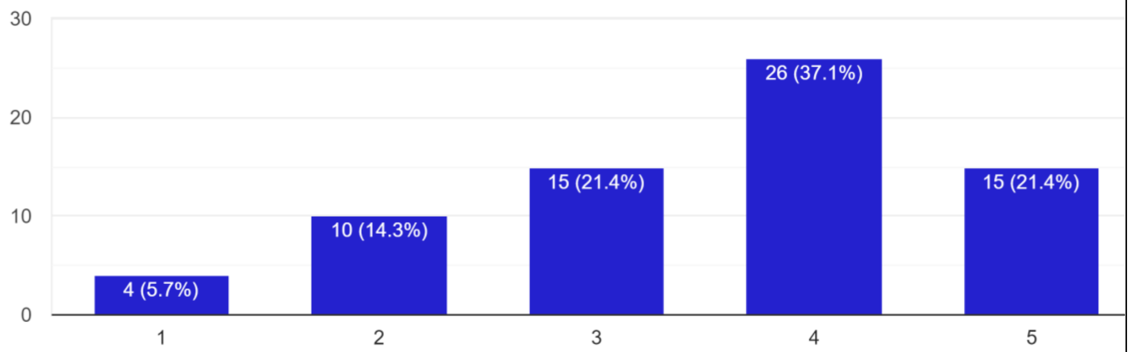
Note: Unless indicated otherwise, for all questions which are answered on a linear scale from one to five; one represents 'Strongly Disagree' and five represents 'Strongly agree'.

Safer Roads for Cyclists and Pedestrians
Description: In this use case the AI detects cyclists and pedestrians on the open shared roads. An electronic sign alerts car drivers of the presence of the cyclists or pedestrians. Therefore intuitively the car drivers should take greater care, for example they slow down.

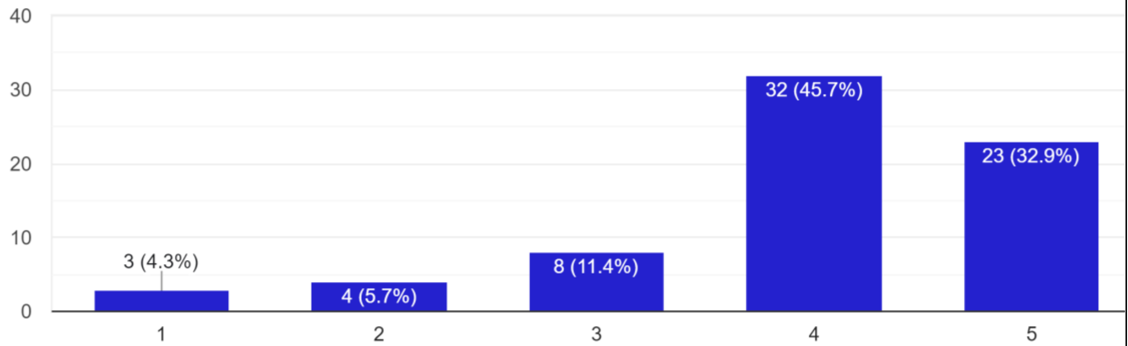
Do you agree that if such a system is successful (i.e car drivers take greater care) cyclists or pedestrians will feel more safe on the road?
70 responses



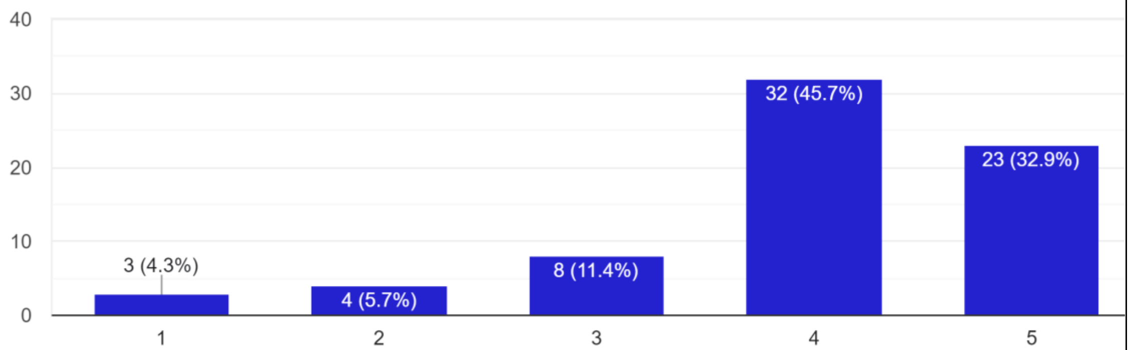
Do you agree that such a system encourages more people to cycle on the road?
70 responses



What about you? Would you be more inclined to use an active mode (cycling or walking) of transport, knowing that roads are safer?
70 responses



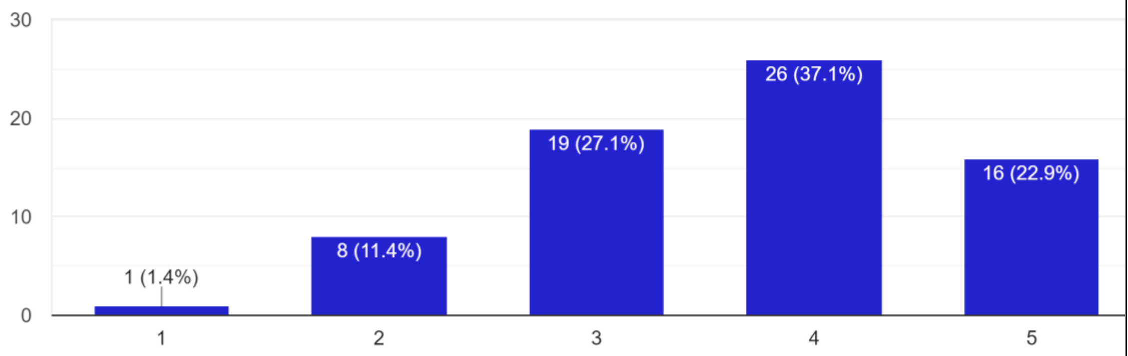
What about you? Would you be more inclined to use an active mode (cycling or walking) of transport, knowing that roads are safer?
70 responses



AI for Road User Behaviour Analysis
Description:
AI can be used to detect improper behaviour on the roads, for example cars not stopping at stop signs, cutting corners and not observing right-of-way etiquette. These bad habits are typically difficult to pin down using law enforcement and targeted educational campaigns may be the way to go.

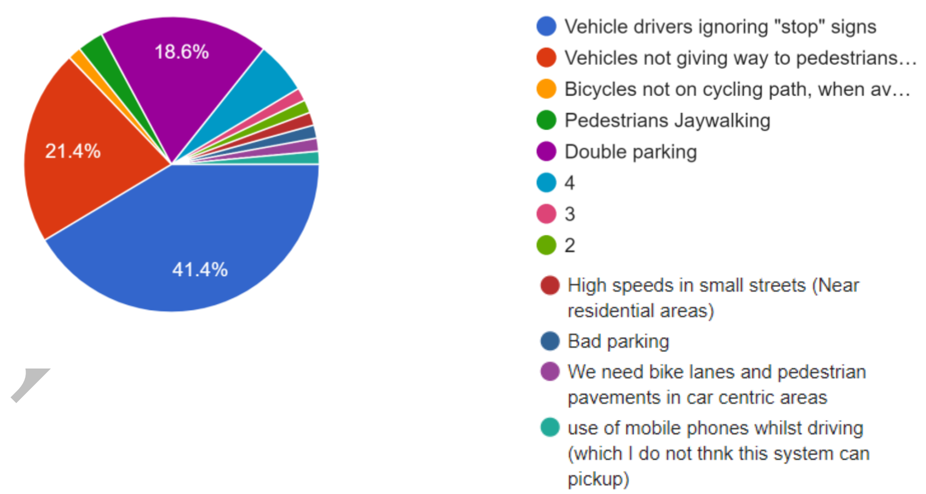
Do you agree that targeted educational campaigns can be effective at reducing certain types of bad behaviour?

70 responses



From the following list of objectionable behaviours, which one would benefit society most if it had to be detected by the AI tools?

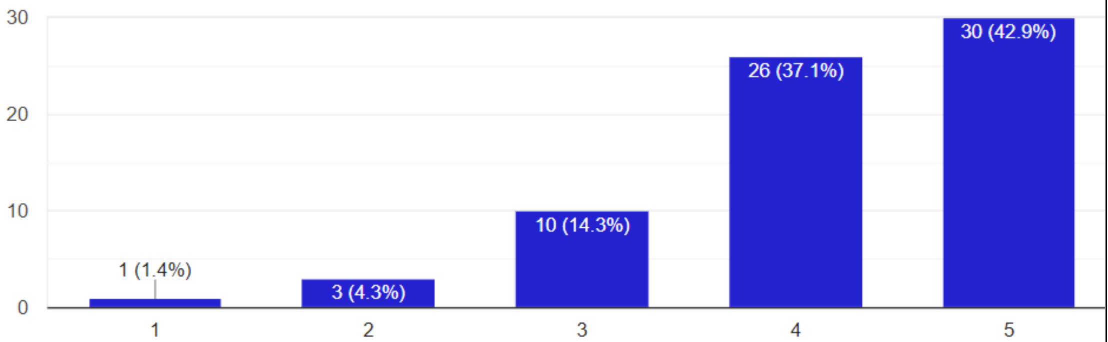
70 responses



Large trucks passing through towns lower the quality of life for residents. The AI system can also be used to detect vehicle violations such as large trucks driven through residential roads.

Would you agree with law enforcement agencies to use these AI models in an effective way, for example, but not only, detecting trucks passing through residential roads?

70 responses



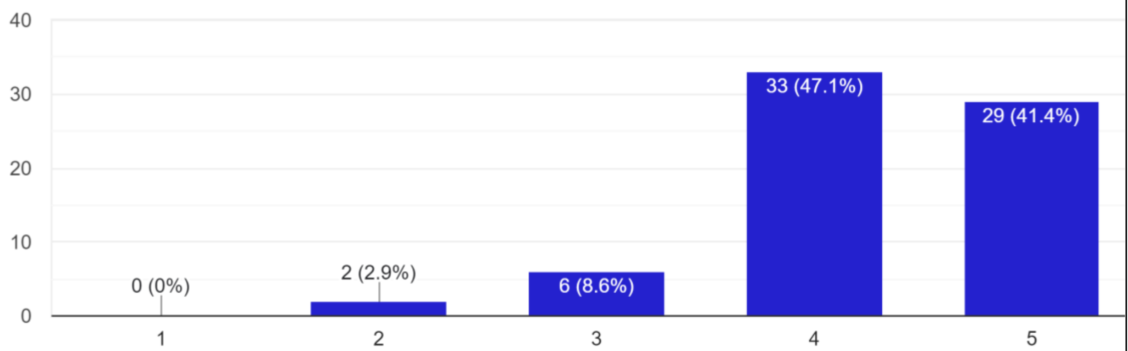
AI for Anomalous Events and Traffic Condition Detection

<u>Description:</u>

Traffic managers and personnel in control rooms observe live video feeds of various junctions to estimate traffic flow, and react when they detect an anomaly, for example an accident at the junction or anomalous traffic flows or queues. The AI system in MARVEL automatically observes the video feeds in real time and alerts the traffic managers of any anomalies.

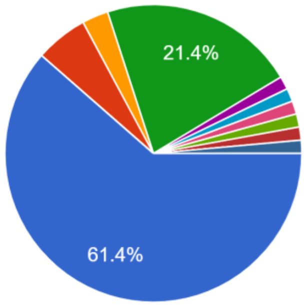
Do you agree that the timely detection of unusual traffic and accidents on the road can reduce journey time/ result in less congestion?

70 responses



In addition the AI is continuously measuring the traffic flow and this data can be used to predict journey time and, in case of congestion, re-route t...e following statements do you agree with the most?

70 responses



- Predicting journey times on main roads in advance allows drivers to consider other routes
- Predicting journey times on main roads does not make any difference in the daily commute
- I am against re-routing since such a system makes use of residential roads
- I am against re-routing since such a system improves journey times, but also car use and on the contrary society should look into reducing car dependency.
- Good data feed into modeling:)
- While providing alternate road usage can lead to decreased journey time for drivers, it can also cause more dangerous roads in residential areas. You cannot fix the traffic issue by just "using another road", more viable transport systems are needed, as well as a new way of thinking about how roads are designed (eg. using smaller, slower roads to minimize both traffic, and accidents)
- I am not against re-routing, but I strongly feel society looking into reducing car dependency is majorly more important. That does not mean that re-routing traffic is a bad idea.
- All of the above
- I am in favour of using AI models to estimate journey times and alerting drivers to make use of alternative routes, however, I would prefer if society would reduce its dependency on cars but instead switch to utilising alternative modes of transport
- Re-routing yes, but only using other non-residential roads

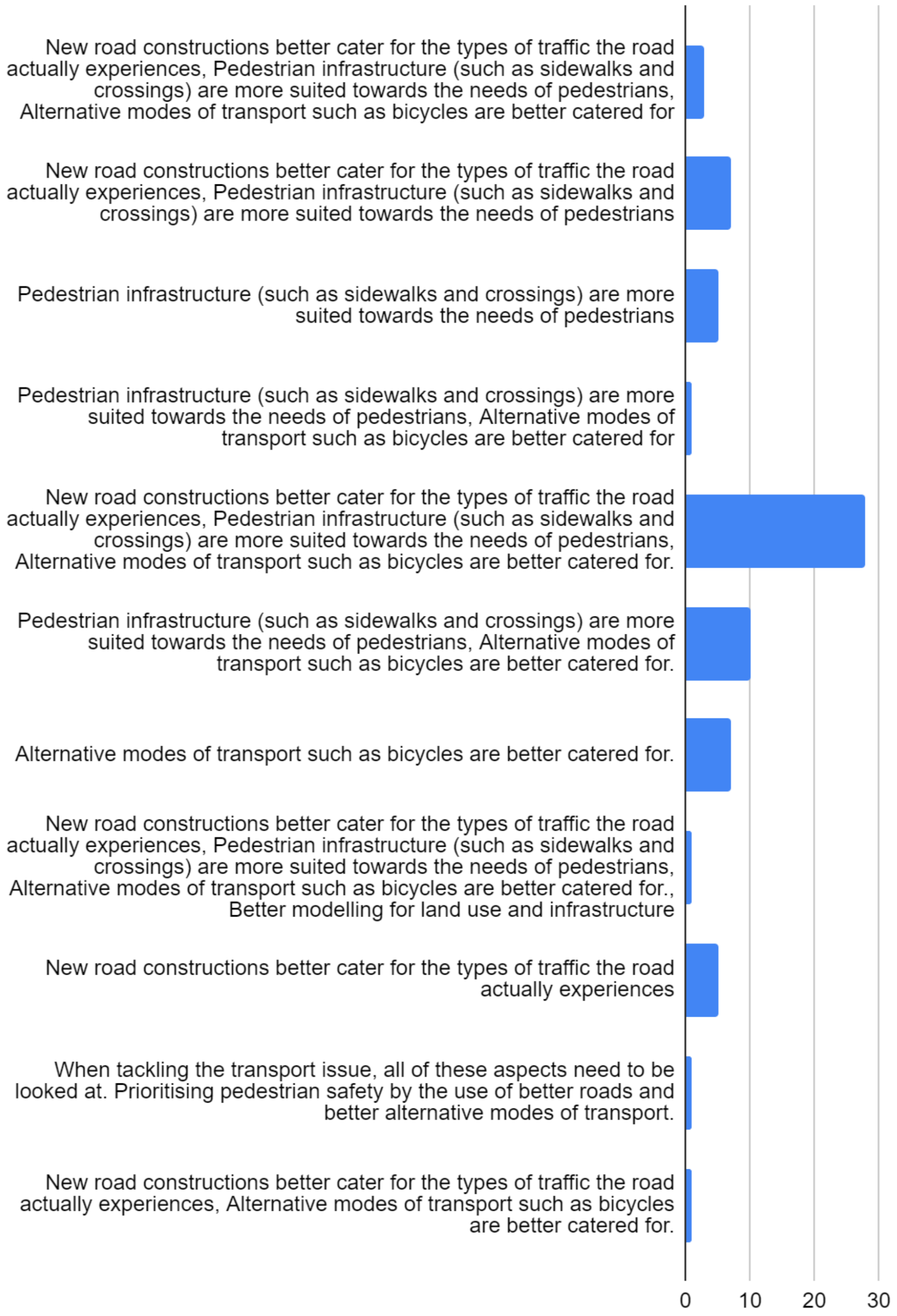
AI for Junction Traffic Trajectory Collection

Description:

The AI collects data, that can be used in for example studying movements of pedestrians, paths taken by cyclists, car drivers and heavy vehicles. This data can be used for long-term analytics, for example to inform road maintenance and safety, and for studying the impact of past and new transport policy.

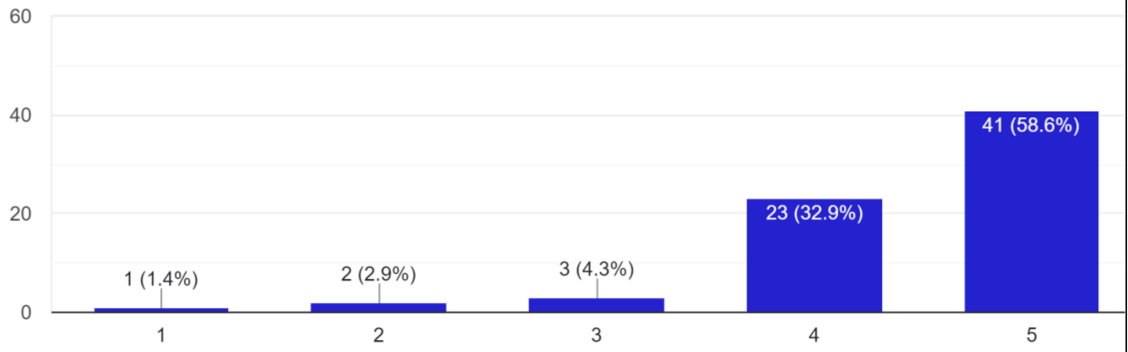
Which of the below matters can improve if authorities have access to long term data on roads? Choose all the options that apply

69 responses


New road constructions better cater for the types of traffic the road actually experiences, Pedestrian infrastructure (such as sidewalks and crossings) are more suited towards the needs of pedestrians, Alternative modes of transport such as bicycles are better catered for
New road constructions better cater for the types of traffic the road actually experiences, Pedestrian infrastructure (such as sidewalks and crossings) are more suited towards the needs of pedestrians
Pedestrian infrastructure (such as sidewalks and crossings) are more suited towards the needs of pedestrians
Pedestrian infrastructure (such as sidewalks and crossings) are more suited towards the needs of pedestrians, Alternative modes of transport such as bicycles are better catered for
New road constructions better cater for the types of traffic the road actually experiences, Pedestrian infrastructure (such as sidewalks and crossings) are more suited towards the needs of pedestrians, Alternative modes of transport such as bicycles are better catered for.
Pedestrian infrastructure (such as sidewalks and crossings) are more suited towards the needs of pedestrians, Alternative modes of transport such as bicycles are better catered for.
Alternative modes of transport such as bicycles are better catered for.
New road constructions better cater for the types of traffic the road actually experiences, Pedestrian infrastructure (such as sidewalks and crossings) are more suited towards the needs of pedestrians, Alternative modes of transport such as bicycles are better catered for., Better modelling for land use and infrastructure
New road constructions better cater for the types of traffic the road actually experiences
When tackling the transport issue, all of these aspects need to be looked at. Prioritising pedestrian safety by the use of better roads and better alternative modes of transport.
New road constructions better cater for the types of traffic the road actually experiences, Alternative modes of transport such as bicycles are better catered for.
0
10
20
30

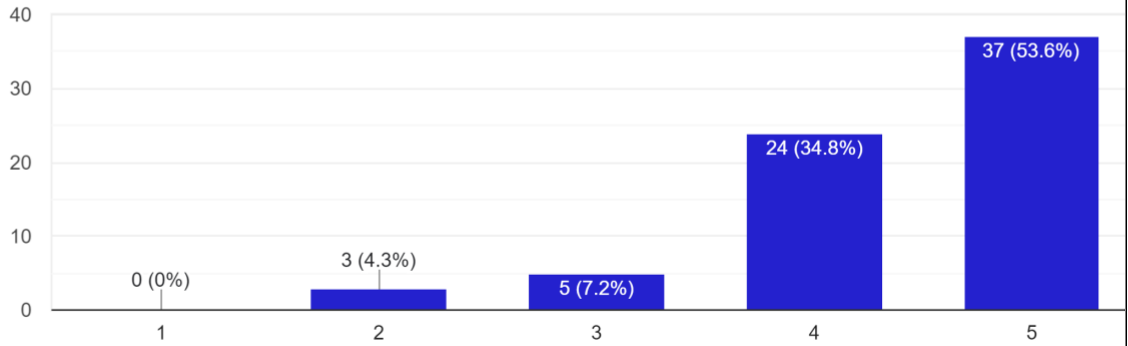
During the planning of roads, do you agree that long term data should be collected on all types of road users including pedestrians and bicycles and not just cars?

70 responses



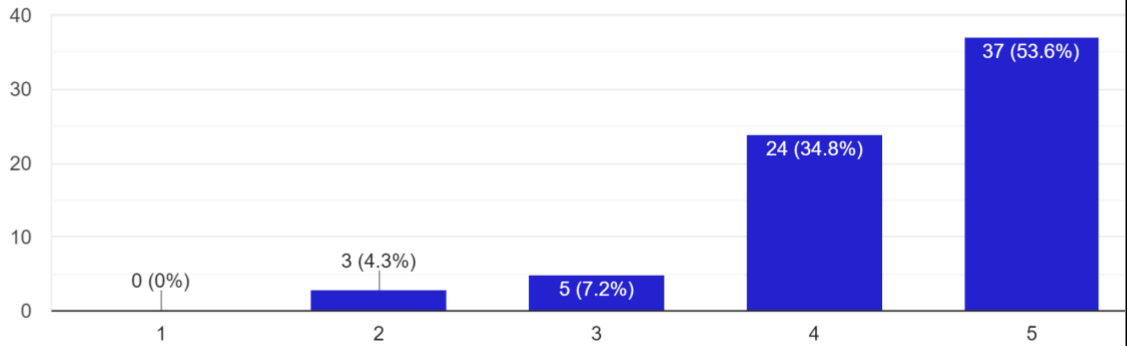
Do you agree that long term data in the analysis of road use would result in more efficient allocation of physical resources for road transport?

69 responses



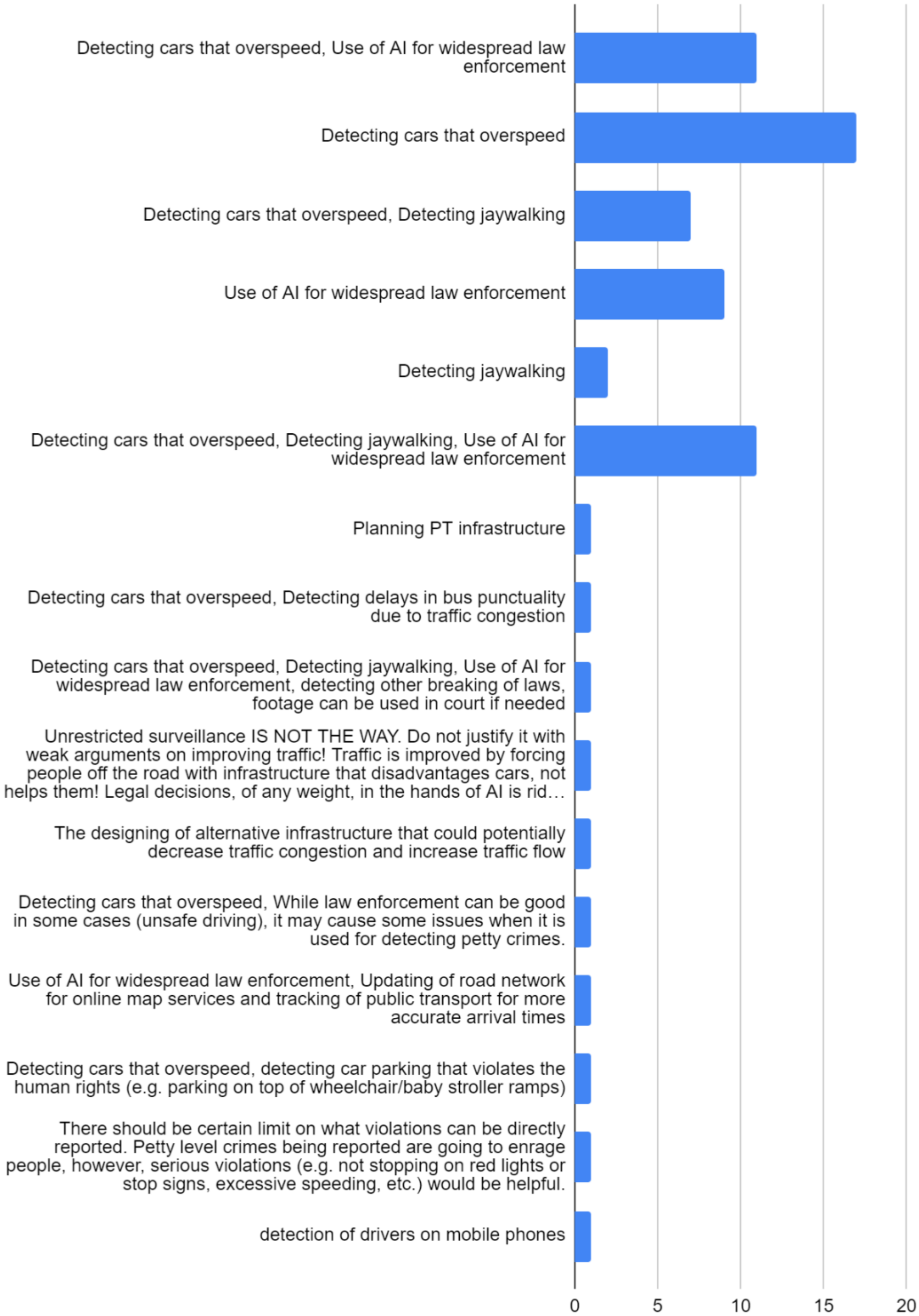
Do you agree that long term data in the analysis of road use would result in more efficient allocation of physical resources for road transport?

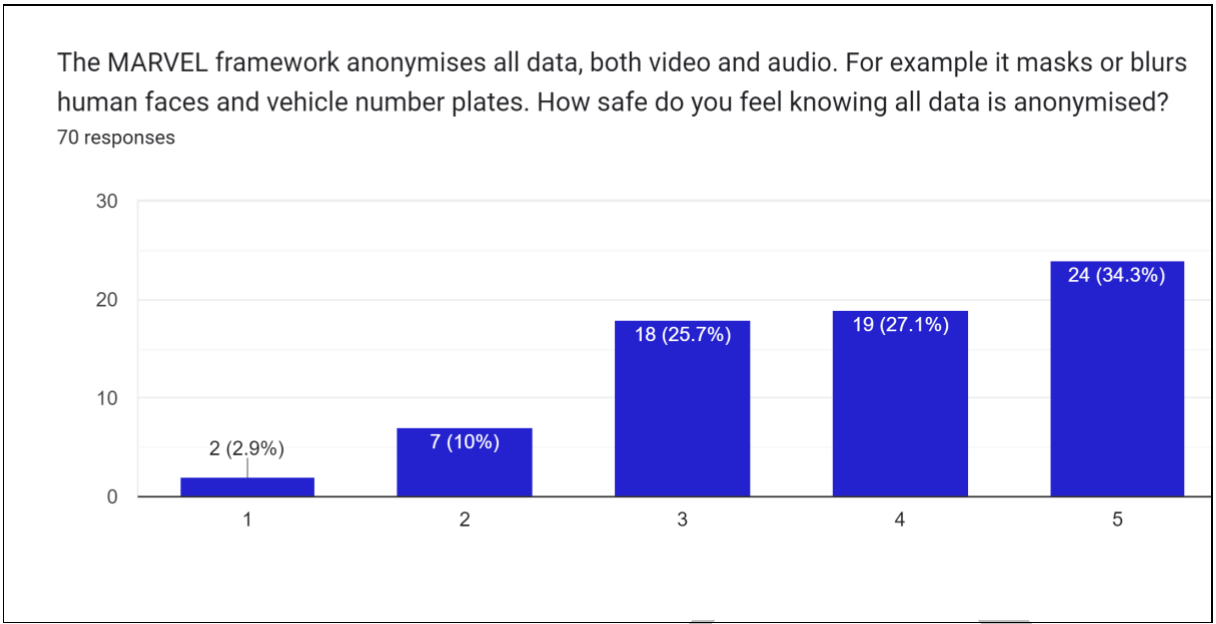
69 responses



Further questions on the use of AI in mobility.

AI in the MARVEL framework can be trained to do other tasks. Which of the following tasks would society benefit from: 67 responses





17.1.7 Malta pilot and Gozo MVP– Interviews

Five of approximately one hour in length interviews were organised with selected professionals working in the transport and related sectors; Urban planner, Transport and Traffic consultant, Law enforcement manager, and municipality managers and directors. The interviews consisted of an introductory presentation, including demos, followed by a discussion on an appropriate selection of topics from the following list:

Functional Objectives:

- Collection of short and long-term data for (a) Trend analysis; (b) Seasonal analysis; (c) Long-term Planning; (d) Maintenance scheduling.
- Traffic control room semi-automation.
- Tool that helps in law enforcement.
- Tool that helps in educational campaigns.
- Increasing safety on the road for cyclists and pedestrians.

Business Objectives:

- From a public administration point of view, the better use of data leads to:
    - More efficient use of taxpayer funds towards safe and functional roads.
    - Cost reduction due to increase in productivity.
    - More effective data-driven decisions.
- The use of the multiple sensors deployed in the city leads to:
    - Reductions in the cost of human monitoring.
    - Reduction of social costs due to misbehaviour.
- Reduced time in processing data for data-driven decisions leads to:
    - Quicker intervention for real-time monitoring scenarios.

  - Satisfaction of cyclists and other vulnerable road users.
- Data is presented in an intuitive and easy to use manner.
- Potential expansion of the MARVEL framework.
  - Any other locations in Gozo that can potentially benefit from MARVEL.

Societal Objectives

More responsible public spending and a better quality of life.

- Increased uptake of active modes of travel, such as walking and cycling.
- Improved road etiquette among road users.
- More efficient use of the physical transport resources.
- Timely maintenance of the physical transport resources.

17.2 Trento pilot - Introduction to methodologies and documents

To evaluate the use cases in the Trento pilot and assess the impact on society and business processes two methods were used; (a) online surveys, and (b) live interviews. The former is covered in Sections 17.2.1-17.2.5, whilst the latter in Section 17.2.6. The results from both methods are used to complete the evaluation tables related to the Trento Pilot.

17.2.1 Trento pilot – Online Surveys

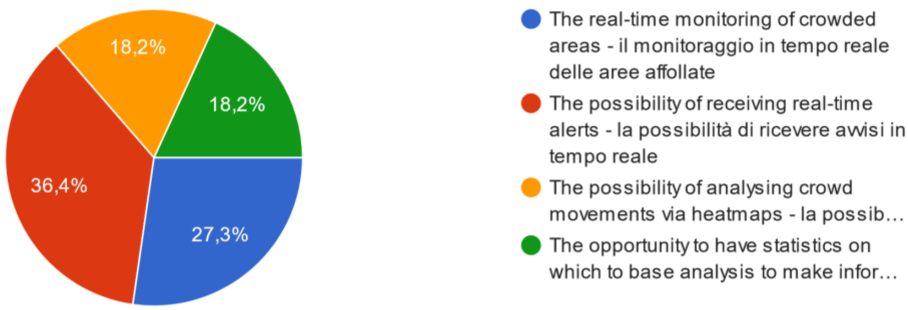
Four online surveys were prepared, considered each use case and were addressed to specific professionals working in the transport and urban planners, such as law enforcement officers and policy-makers. Each survey included textual descriptions of for example the use case, a recorded video demo and a number of questions. The surveys were distributed by email and sections below details the results per question.

17.2.2 Survey for MT1-Crowd Monitoring - Trento Pilot

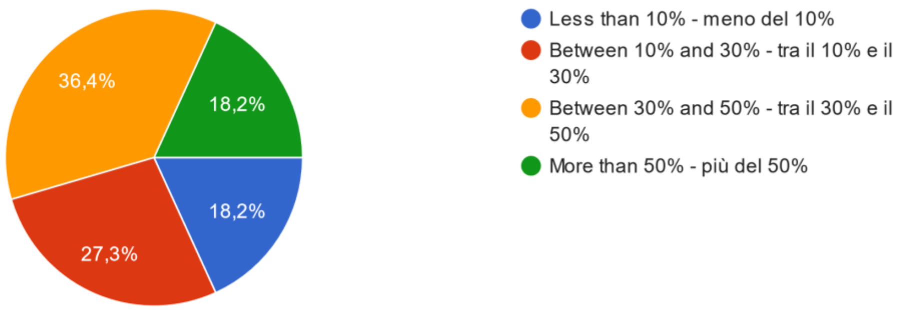
This survey was intended to evaluate the use case MT1 – Crowd Monitoring by Local Police of the Municipality of Trento. A total of 11 experts replied to the survey: the Commander of Local Police, the Head of Urban Security Unit and some Central Operations Officers.

The survey starts with a high-level explanation of the MT1 and the presentation of the SmartViz UI, then some questions about the impact of the pilot and other questions related to the usability of the UI.

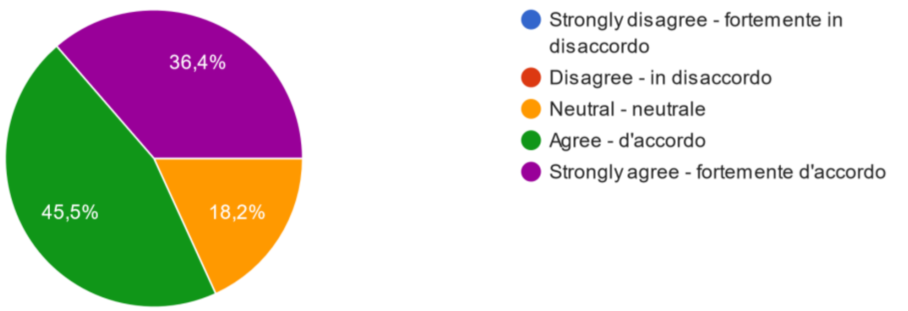
After watching the video showing a system to monitor and verify events such as exceptional crowding, suspicious or abnormal behaviour of ind...ema ha trovato più interessante o impressionante?
11 risposte



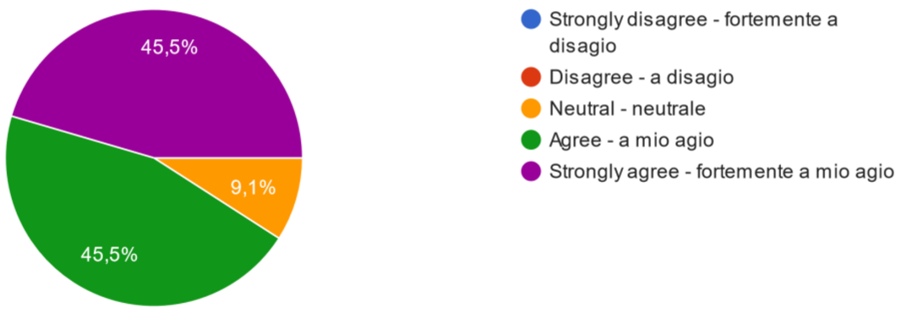
By having the possibility of receiving real-time alerts, how much do you estimate the response time in case of need for intervention can be reduced com... rispetto allo scenario attuale senza tale sistema?
11 risposte



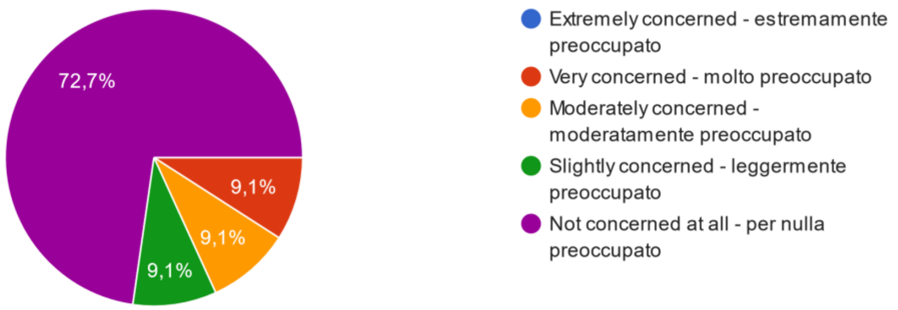
The system can also store crowd movement records, complete with date and time, over a long period of time. Do you agree that this data is useful...za, in particolare per eventi pubblici medio-grandi?
11 risposte



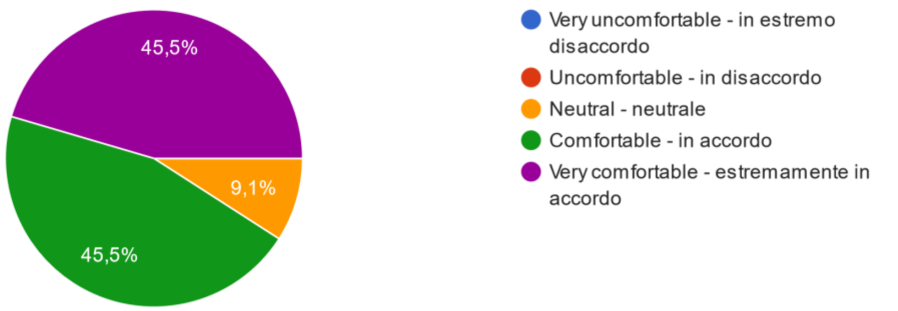
Would you be comfortable with the presence of such a monitoring system in public spaces you frequent (e.g., malls, transportation hubs)?   Ti se... (ad esempio, centri commerciali, hub di trasporto)?
11 risposte



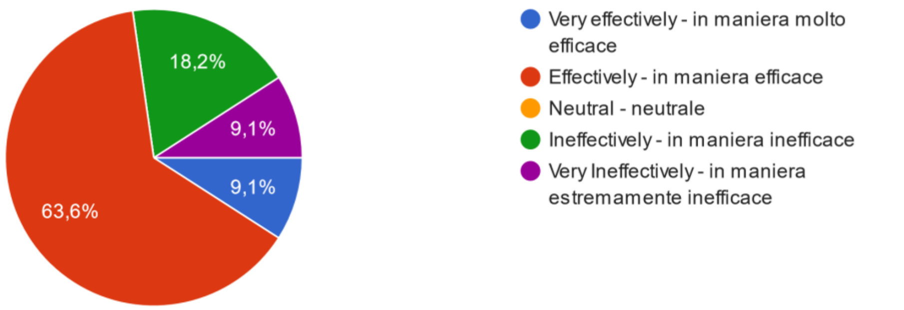
To what extent are you concerned about privacy issues related to the implementation of such a system in public spaces?   In che misura è preoccupa...l'attuazione di tale sistema negli spazi pubblici?
11 risposte



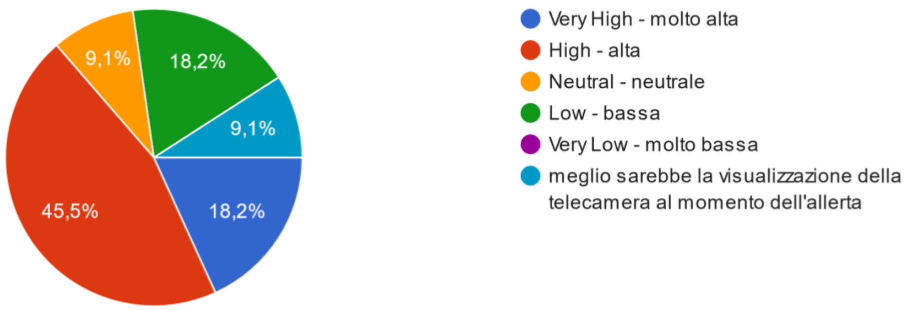
In general, do you think that a system like MARVLE framework can improve the local area, city surveillance, citizens' quality of life, their perceive...ta e migliorare l'esperienza complessiva per i turisti?
11 risposte



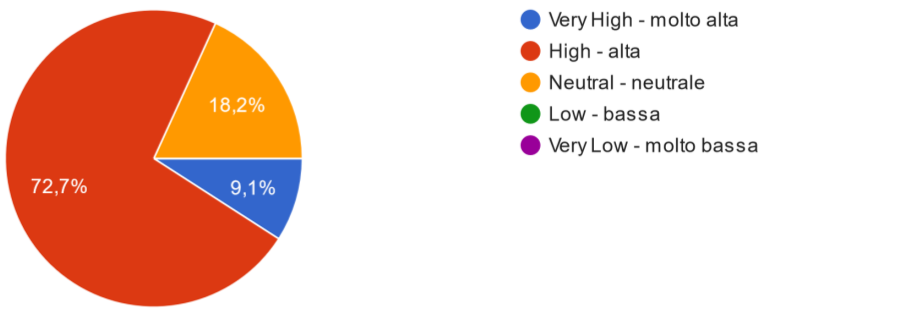
How well does the SmartViz UI enable Local Police officers to focus on relevant views, analyze detected anomalies, and determine appropriate cours...e durante gli eventi pubblici nelle aree affollate?
11 risposte



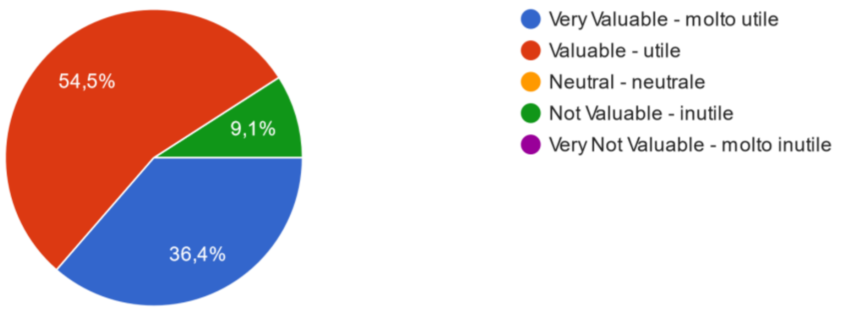
To what extent does the alerting functionality in SmartViz, particularly through the Alerts widget, enable you to promptly respond to anomalous events...tendo la sicurezza pubblica nelle aree affollate?
11 risposte



To what extent does SmartViz enhance your ability to compare two different time periods, providing a comprehensive understanding of temporal pattern...temporali e dei cambiamenti nelle aree affollate?
11 risposte



How valuable do you find the availability of weather information in SmartViz for correlating weather conditions to detected anomalies and alerts in public...malie e agli avvisi rilevati nelle piazze pubbliche?
11 risposte

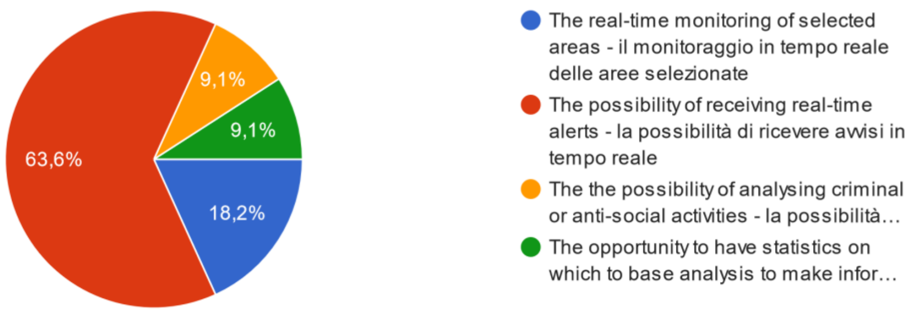


17.2.3 Survey for MT2-Detecting criminal and anti-social behaviours

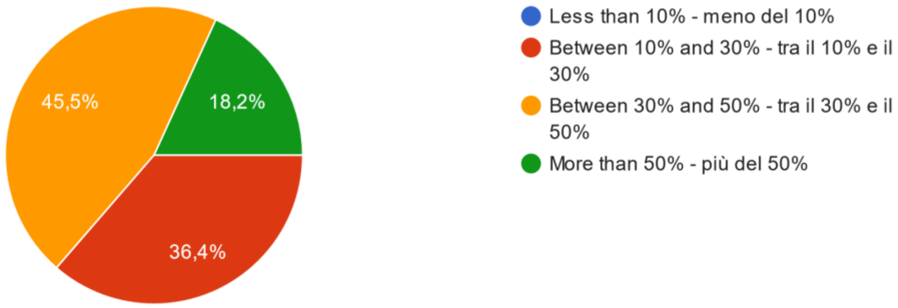
This survey was intended to evaluate the use case MT2 – Detecting criminal and anti-social behaviours by Local Police of the Municipality of Trento. A total of 11 experts replied to the survey: the Commander of Local Police, the Head of the Urban Security Unit and some Central Operations Officers.

The survey starts with a high-level explanation of the MT2 and the presentation of the SmartViz UI, then some questions about the impact of the pilot and other questions related to the usability of the UI.

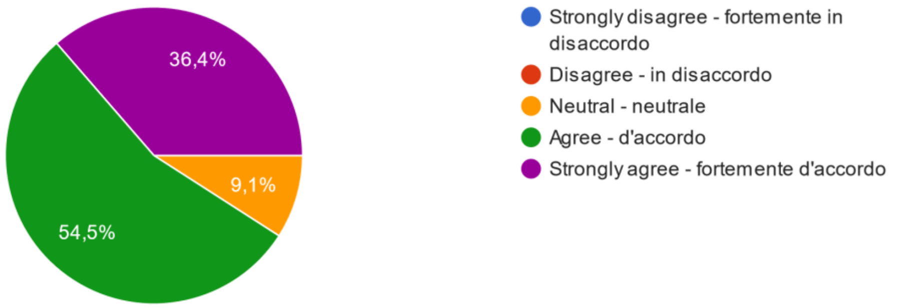
After watching the video showing a system to monitor and identifying criminal or anti-social activities, which specific aspects of the system di...tema ha trovato più interessante o impressionante?
11 risposte



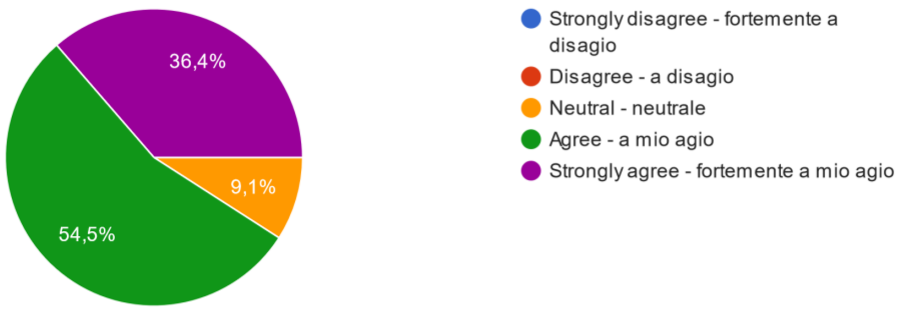
By having the possibility of receiving real-time alerts, how much do you estimate the response time in case of need for intervention can be reduced com... rispetto allo scenario attuale senza tale sistema?
11 risposte



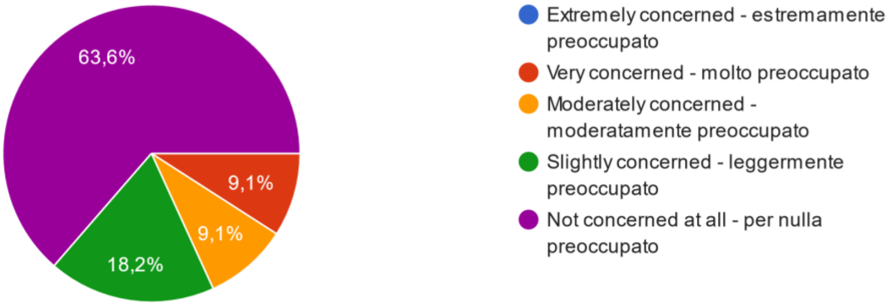
The system additionally has the capability to store records of criminal or anti-social activities, including date and time, over an extended period. D...zione strategica delle misure di sicurezza urbana?
11 risposte



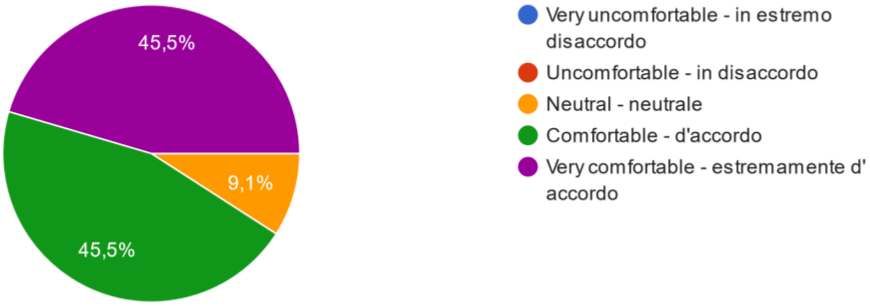
Would you be comfortable with the presence of such a monitoring system in public spaces you frequent (e.g., malls, transportation hubs)? Ti se... (ad esempio, centri commerciali, hub di trasporto)?
11 risposte



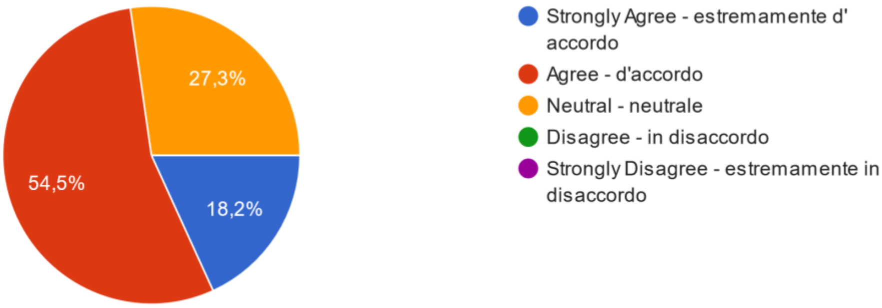
To what extent are you concerned about privacy issues related to the implementation of such a system in public spaces? In che misura è preoccupa...ll'attuazione di tale sistema negli spazi pubblici?
11 risposte



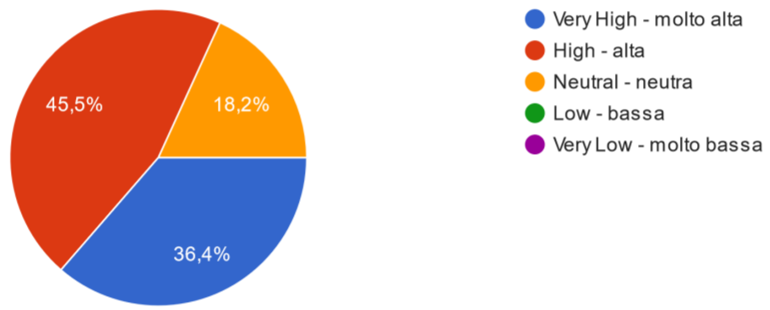
In general, do you think that a system like MARVLE framework can improve the local area, city surveillance, citizens' quality of life, their perceive...ta e migliorare l'esperienza complessiva per i turisti?
11 risposte



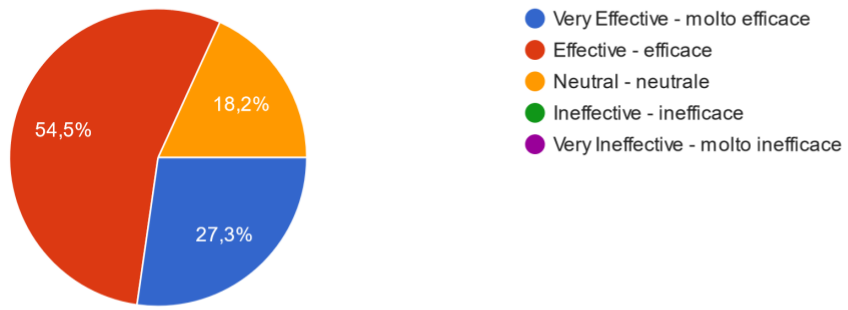
In your opinion, does the Word Cloud widget effectively fulfill its purpose of showcasing the most prevalent keywords and descriptions, as generated b...a migliore comprensione degli eventi identificati?
11 risposte



To what extent does the police intervention functionality contribute to the user's ability to flag specific events within the visualisation interface an... per un'ulteriore attenzione da parte delle autorità?
11 risposte



How effective do you find the statistics widget, particularly the bar chart on the left, in visually representing the numbers of categories associated wit...iate alle attività criminali o antisociali rilevate?
11 risposte

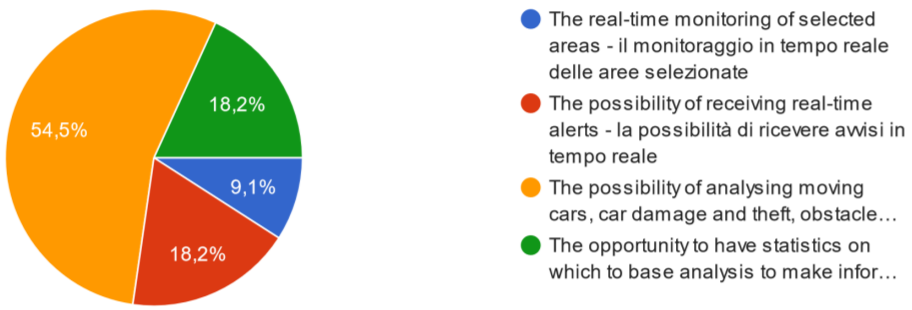


17.2.4 Survey fot MT3-Monitoring of parking places

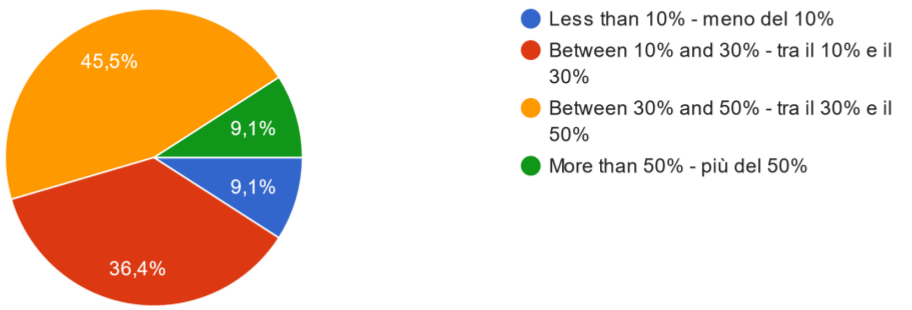
This survey was intended to evaluate the use case MT3 – Monitoring of parking places by Local Police of the Municipality of Trento and Mobility Department. A total of 11 experts replied to the survey: the Commander of Local Police, the Head of the Urban Security Unit, some Central Operations Officers and two people working in the mobility and urban planning sectors of the Municipality.

The survey starts with a high-level explanation of the MT3 and the presentation of the SmartViz UI, then some questions about the impact of the pilot and other questions related to the usability of the UI.

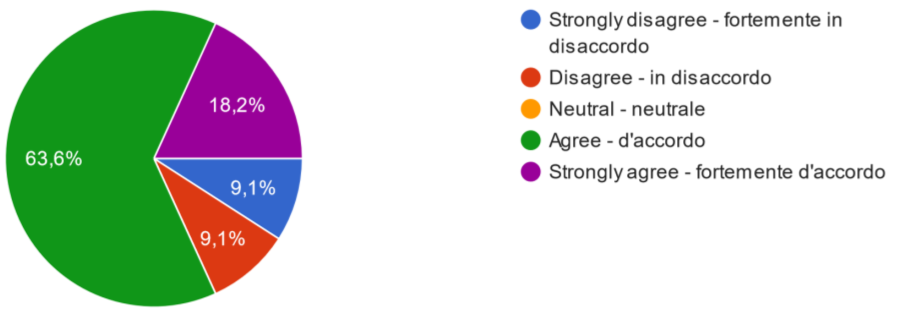
After watching the video showing a system to monitor and identify moving cars, car damage and theft, obstacles, etc. in a car park, which specific...stema hai trovato più interessante o impressionante?
11 risposte



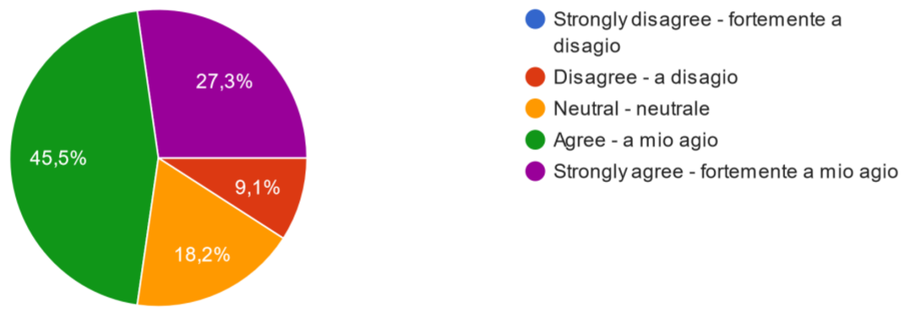
By having the possibility of receiving real-time alerts, how much do you estimate the response time in case of need for intervention can be reduced com... rispetto allo scenario attuale senza tale sistema?
11 risposte



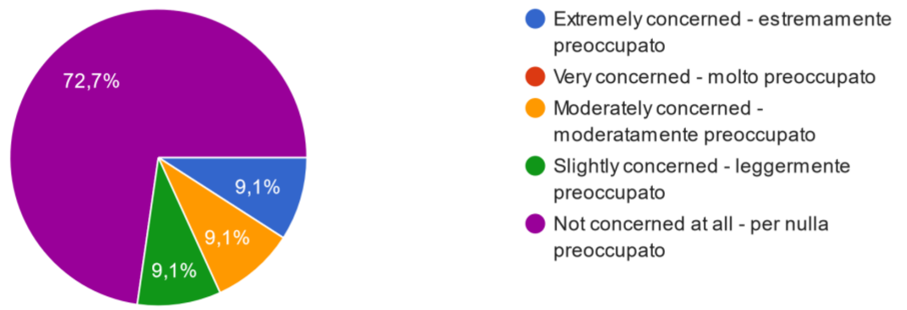
The system is able to store records of moving cars, instances of car damage and theft, as well as the identification of obstacles in a car park, includ...e misure di sicurezza urbana e dei piani di mobilità?
11 risposte



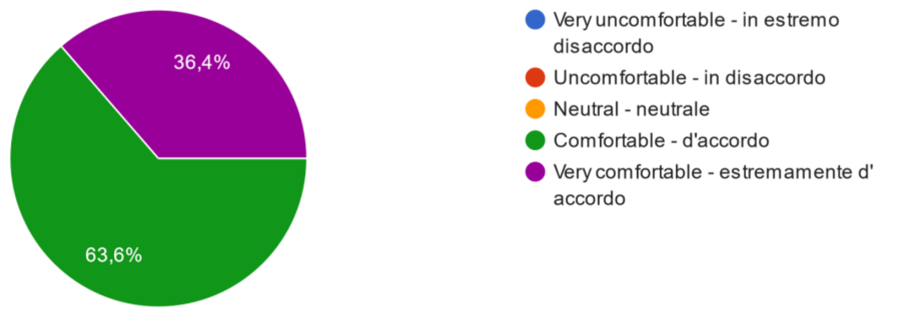
Would you be comfortable with the presence of such a monitoring system in public spaces you frequent (e.g., malls, transportation hubs)? Ti se... (ad esempio, centri commerciali, hub di trasporto)?
11 risposte



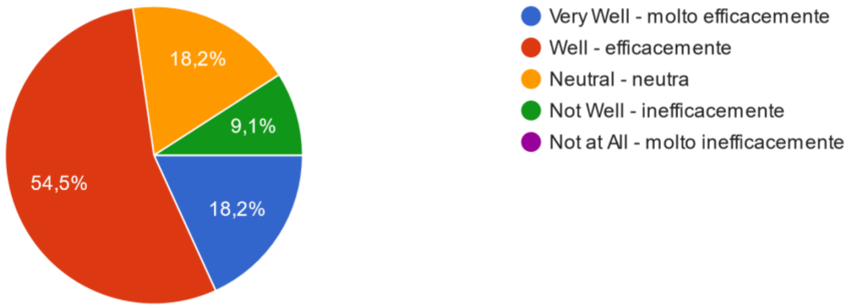
To what extent are you concerned about privacy issues related to the implementation of such a system in public spaces? In che misura è preoccupa...ll'attuazione di tale sistema negli spazi pubblici?
11 risposte



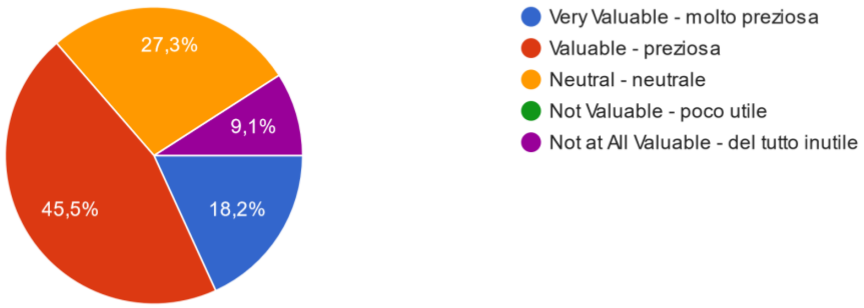
In general, do you think that a system like MARVLE framework can improve the local area, city surveillance, citizens' quality of life, their perceive...ta e migliorare l'esperienza complessiva per i turisti?
11 risposte



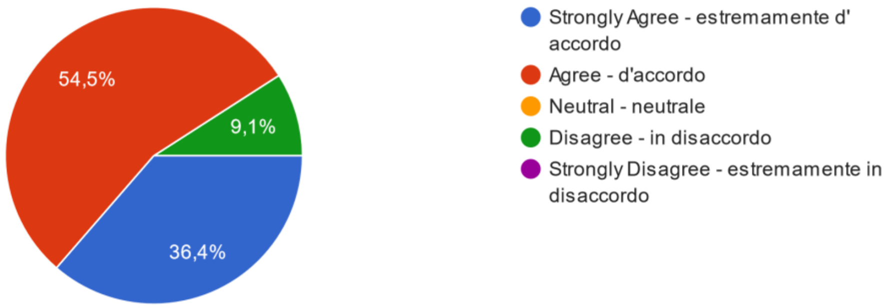
How well does the SmartViz UI provide features like the timeline distribution of vehicles, total number of vehicles, clustering of vehicles/events, ...e di monitorare correttamente l'area di parcheggio?
11 risposte



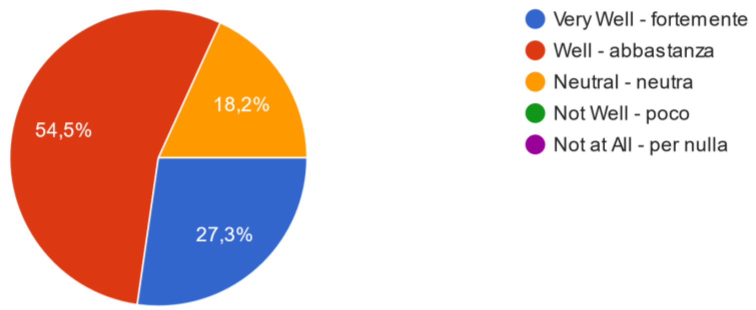
In your experience, how valuable is the ability to download visualized data in JSON format for further analysis, and does it enhance the overall ut...raggio dei parcheggi e del rilevamento di anomalie?
11 risposte



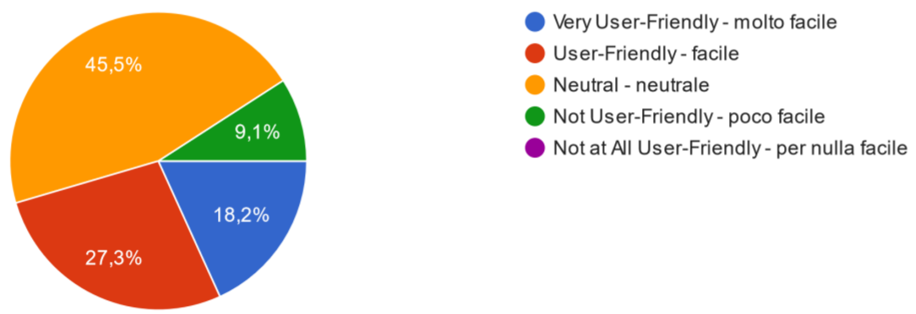
In your opinion, does the summaries widget fulfill its purpose of providing a quick and easy overview of anomalies in the parking lot, supporting Local Po...ale nelle loro azioni di monitoraggio e correzione?
11 risposte



To what extent does the flexibility of the SmartViz UI, including the ability to rearrange and resize widgets, enhance your user experience and interacti...esperienza utente e l'interazione con il cruscotto?
11 risposte



How user-friendly and efficient is the audio player functionality in SmartViz, allowing users to select an event detected in a microphone stream and play ...fono e riprodurre il corrispondente snippet audio?
11 risposte



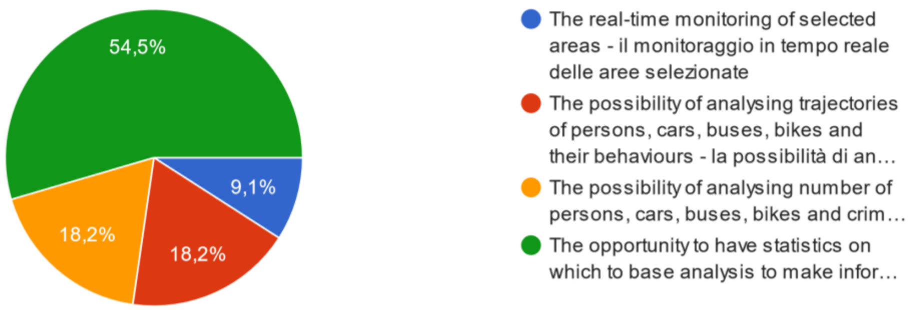
17.2.5 Survey for MT4-Analysis of a specific area

This survey was intended to evaluate the use case MT4 – Analysis of a specific area by Local Police of the Municipality of Trento and Mobility Department. A total of 12 experts replied to the survey: the Commander of Local Police, the Head of the Urban Security Unit, some Central Operations Officers, two people working in the mobility and urban planning sectors of the Municipality and a policymaker.

The survey starts with a high-level explanation of the MT4 and the presentation of the SmartViz UI, then some questions about the impact of the pilot and other questions related to the usability of the UI.

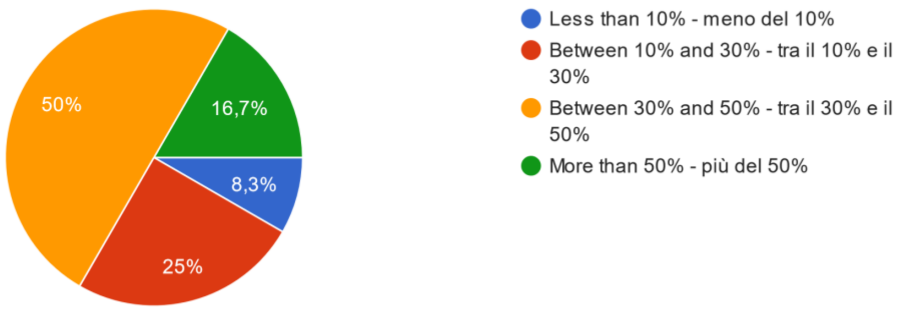
After watching the video showing a system to monitor and identify number of persons, cars, buses, bikes and their behaviours, criminal or anti-social...stema ha trovato più interessante o impressionante?

11 risposte



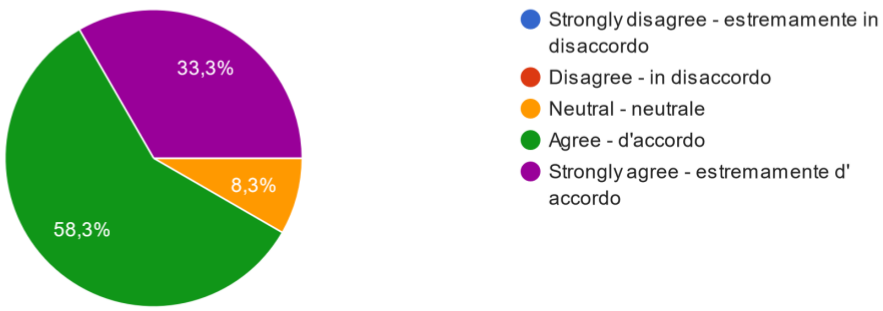
By having the possibility of real-time monitoring, how much do you estimate the response time in case of need for intervention can be reduced compa...essere ridotto rispetto ad ora senza tale sistema?

12 risposte

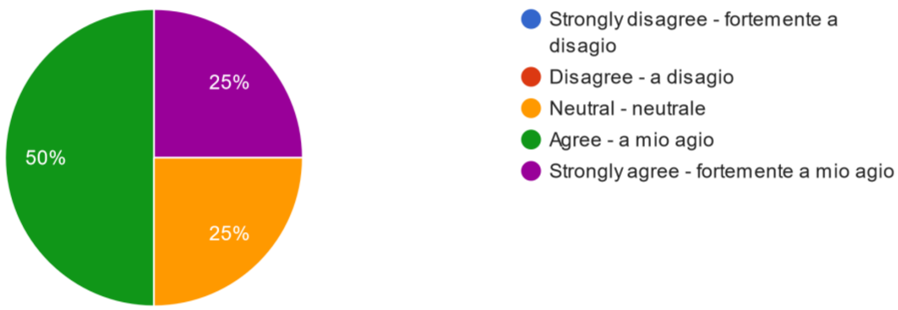


The system can also store number of persons, cars, buses, bikes and their behaviours, criminal or anti-social activities records, complete with date an...icare piani di sicurezza urbana e piani di mobilità?

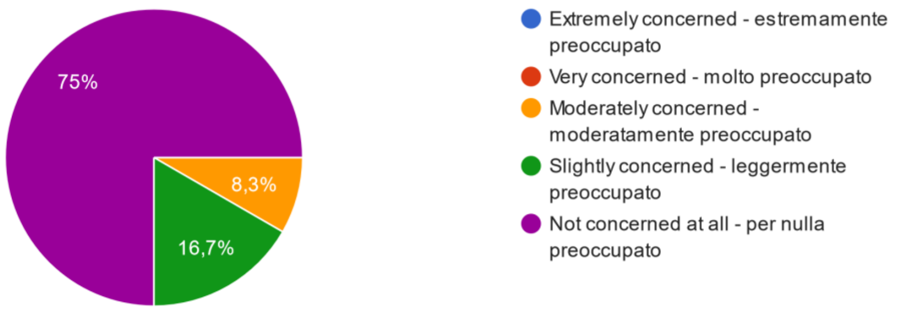
12 risposte



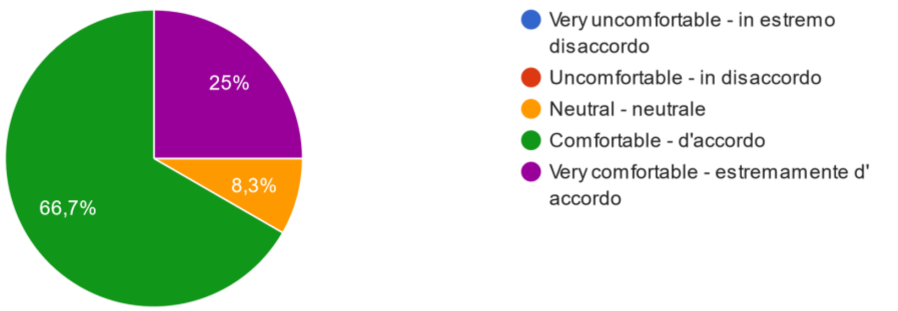
Would you be comfortable with the presence of such a monitoring system in public spaces you frequent (e.g., malls, transportation hubs)? Ti se... (ad esempio, centri commerciali, hub di trasporto)?
12 risposte



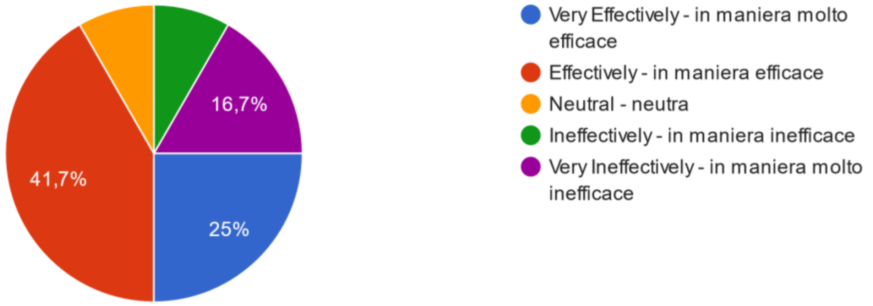
To what extent are you concerned about privacy issues related to the implementation of such a system in public spaces? In che misura è preoccupa...ll'attuazione di tale sistema negli spazi pubblici?
12 risposte



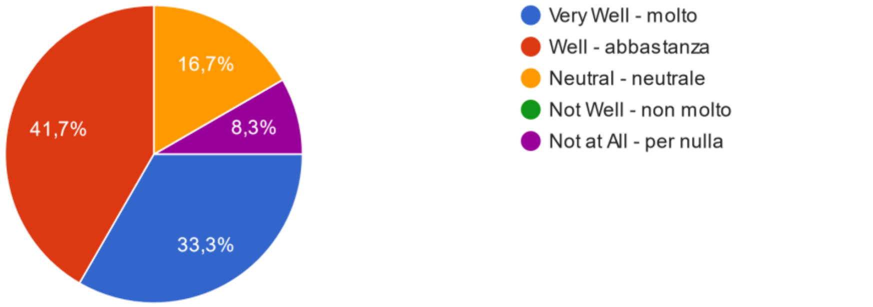
In general, do you believe that a framework like MARVEL has the potential to enhance local area and city surveillance, improve citizens' quality of life, c...a e migliorare l'esperienza complessiva per i turisti?
12 risposte



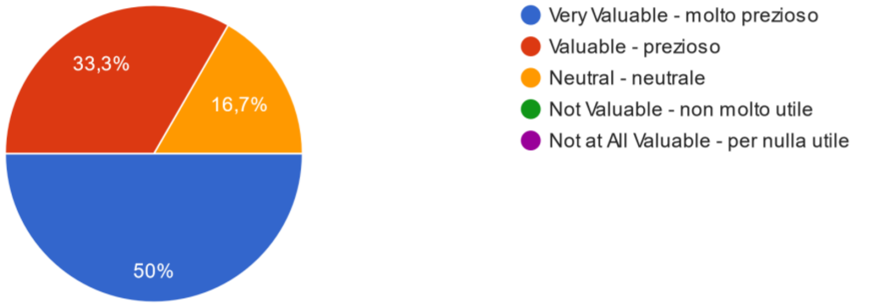
How effectively does the user interface in SmartViz help achieve the goals and objectives of monitoring antisocial behaviors, as well as supportin...ficiente del traffico e della pianificazione urbana?
12 risposte

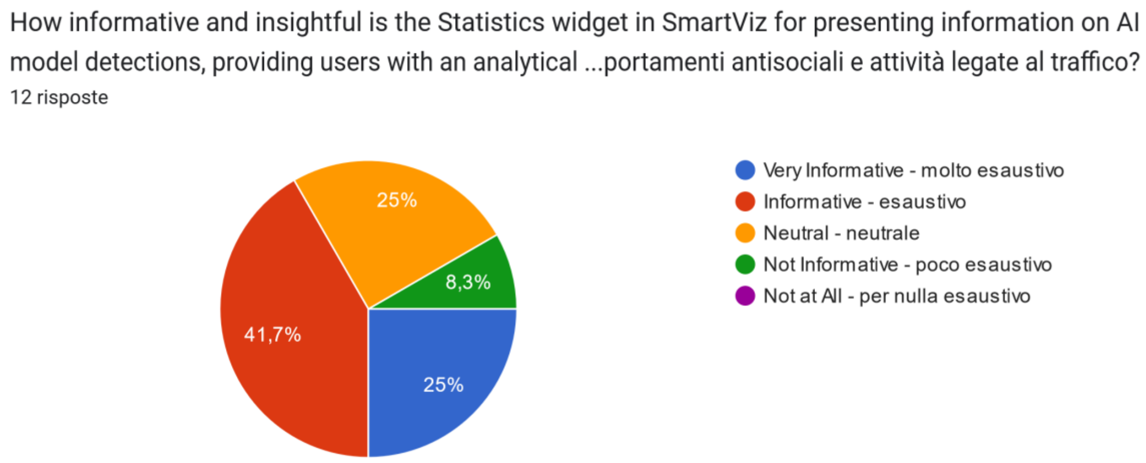


To what extent does the audio player functionality in SmartViz enhance the user experience, allowing users to select an event and play the corres...snippet audio o video a fini di convalida e analisi?
12 risposte



How valuable is the Vehicle Trajectories widget in SmartViz for users in terms of illustrating vehicle paths, adjusting time periods, and filtering by vehicle...menti antisociali e delle attività legate al traffico?
12 risposte





17.2.6 Trento pilot – Interviews

Three one-hour interviews were organised with the Local Police of the Municipality of Trento and the Mobility Department. The interviews consisted of an explanatory presentation, including demos, followed by a discussion on the following topics:

- effectiveness of the system (time, costs, benefits, etc.);
- ease of use of the system;
- impacts (quality of life of citizens, privacy, etc.).

17.3 UNS experiment - Introduction to methodologies and documents

To evaluate the use cases in the UNS experiment and access the impact on society and business processes two methods were used; (a) online surveys, and (b) live interviews. The former is covered in section , 17.3.1 whilst the latter in section 17.3.2. The results from both methods are used to complete the evaluation tables related to the UNS experiment.

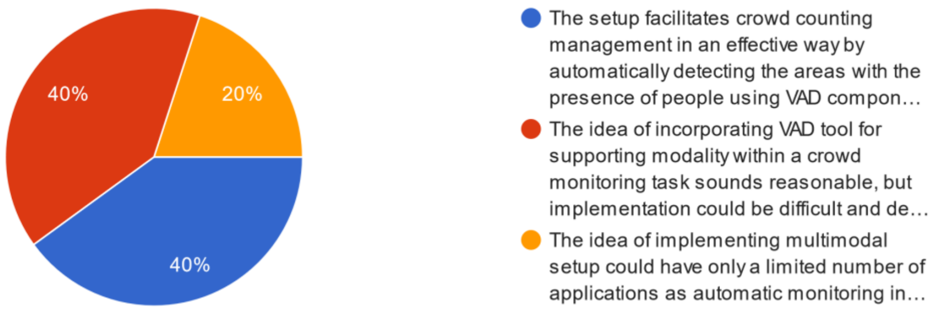
17.3.1 UNS1 survey

I hereby confirm that I am 18 years of age or older. I am aware that completing and submitting this anonymous questionnaire implies that I am particip... informed consent on the conditions listed above.

5 responses

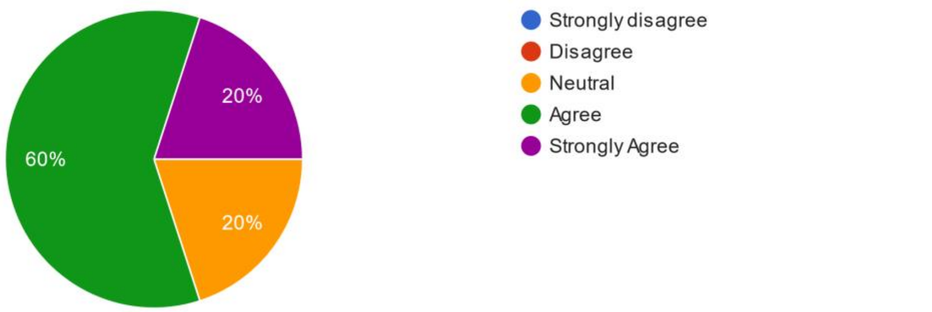
How effective do you find the multimodal setup which incorporates VAD and VCC tools?

5 responses



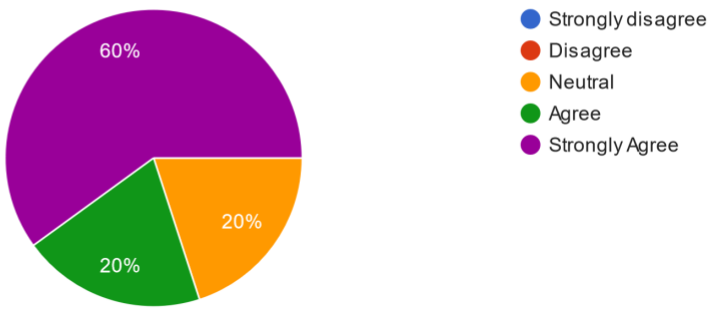
Do you agree that the presented crowd counting technology is accurate?

5 responses



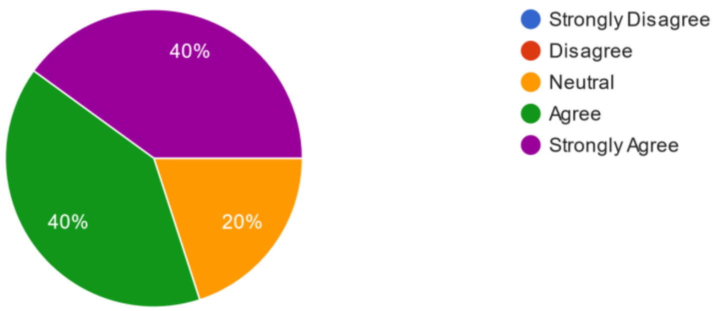
Do you find that the incorporated automatic alarm can improve prevention of overcrowding at covered areas of large public events?

5 responses



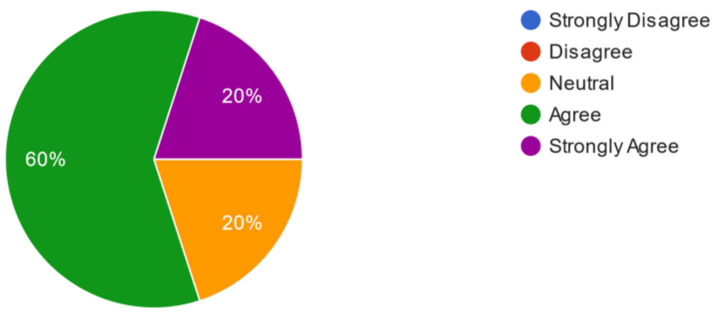
From what you have seen, do you agree that the UNS1 setup can seamlessly be adapted for the usage of new equipment for audio-visual monitoring?

5 responses



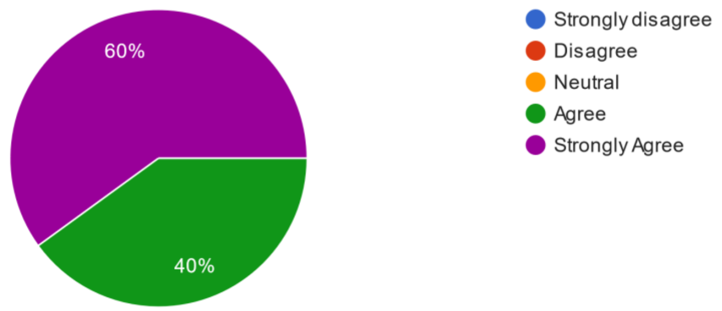
Taking into account that both audio and video streams are being anonymised at the Edge layer before sending data to the Fog and Cloud servers ... UNS1 architecture supports secure transmission?

5 responses



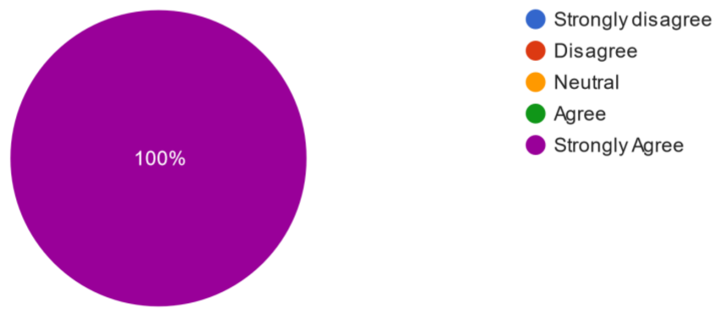
I agree that the Heatmap Representation widget provides a valuable and effective means of visualizing and understanding the movements of the crowd.

5 responses



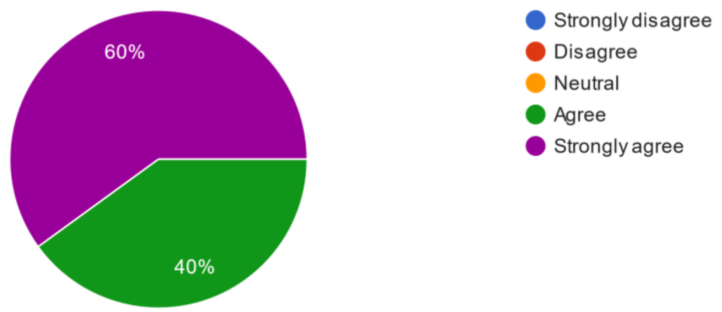
The ability to replay the consecutive created heatmaps from a selected time period is a valuable way to see the crowd density and distribution acr...nderstanding the context and the crowd movement.

5 responses



I agree that the Video Stream Player widget is effective in providing a visual representation of anonymized camera feeds and contributes to the overall usability of the MARVEL project.

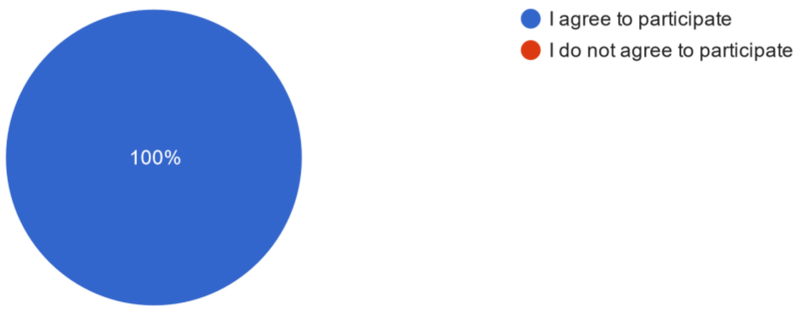
5 responses



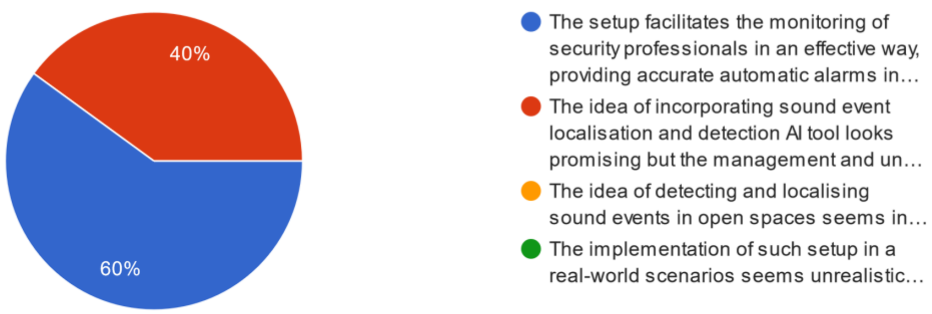
I agree that SmartViz, accelerates decision-making in detecting and addressing anomalous events during the drone experiment.
5 responses

17.3.2 UNS2 survey

I hereby confirm that I am 18 years of age or older. I am aware that completing and submitting this anonymous questionnaire implies that I am particip... informed consent on the conditions listed above.
5 responses

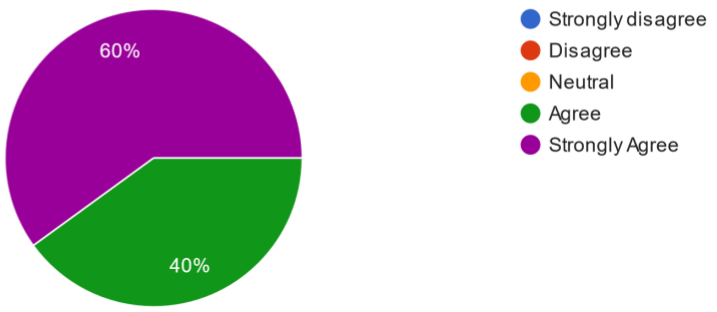


How effective do you find the implementation of such a setup in urban city areas for monitoring open spaces?
5 responses



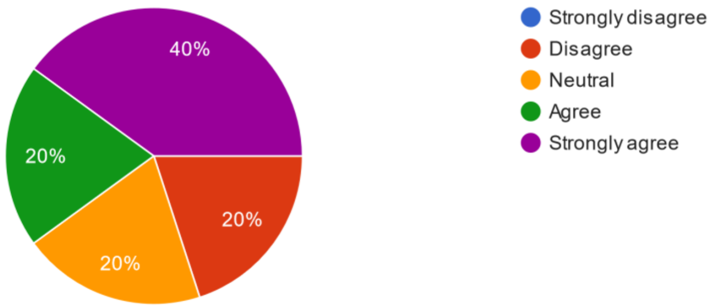
From what you have seen, do you agree that the UNS2 setup can seamlessly be adapted for the usage of several microphone array boards placed i... number of audio channels to simplify monitoring?

5 responses



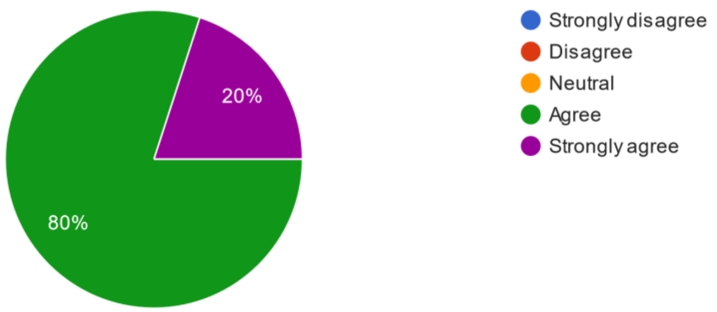
As the UNS2 setup incorporates audio anonymisation component alongside voice activity detection that precedes it, do you agree that the integration ...es in maintaining citizen privacy and data security?

5 responses



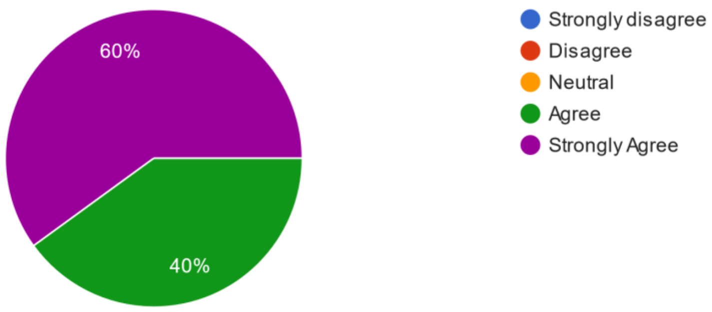
From what you have seen, do you agree that the implementation of new equipment due to continuous long-term upgrades of the UNS2 setup would not be a major problem?

5 responses



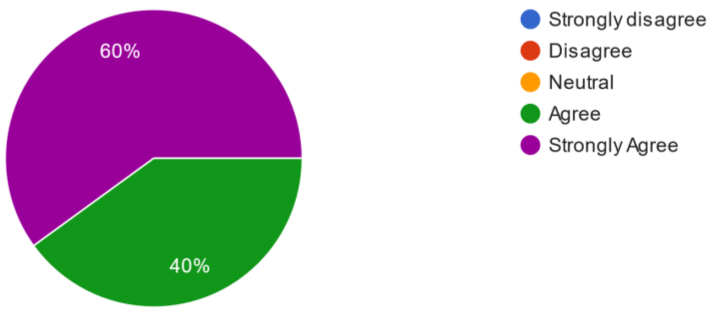
I think that the Sound localisation map widget enables one to explore the spatial characteristics of sound and make data-driven observations or decisions based on the displayed information.

5 responses



I think that the audio playback functionality enhances the user's ability to explore and understand the content within the broader context of their anal...nsights and facilitating informed decision-making.

5 responses



I agree that the details widget (Sound Event Localisation), presents the incoming detections, in a concise and clear way for me to combine them along...e detailed information regarding the detections.

5 responses