

# Responsible AI for road safety

## 1 Introduction to work package 4

WP4 is a unique work package consisting of 3 DCs working in synergy and covering three distinct yet complementary aspects of responsible AI in road safety: ethical theories (DC1), AI methods (DC2) and AI applications (DC3). It aims to develop a new theoretical framework for justice in AI for road safety, embedding for the first time a combination of values like fairness, transparency, privacy, explainability and responsibility. It will further operationalise the explainability of AI for optimal policy support by benchmarking AI models from the econometrics and ML domains in concrete policy questions, and the equity in AI by developing new AI-based knowledge discovery and modelling tools dedicated to less resourced countries (LMICs) – in line with the first research goal (RG) of the project: To develop responsible, fair and impactful AI for road safety with the below objectives:

- To create a new knowledge framework for justice in the design, data and models of AI applications in road safety, with a focus on disadvantaged groups, e.g., Low-to-Middle-Income Countries (LMICs), Vulnerable Road Users (VRUs), and women;
- To operationalise the concept of AI explainability as a key enabler of justice in decision-making, and benchmark the value of both machine learning (ML) and econometric AI models in this respect;
- To design and develop new AI-based knowledge discovery and modelling tools for less resourced countries (LMICs);

and the key exploitable results are:

- Theoretical framework for justice in AI for road safety;
- Model agnostic explainable AI tool for road safety.

## 2 Advances beyond the state-of-the-art

So far, WP4 has gone beyond state-of-the-art in drawing attention to the ethical implications of the applications of AI for road safety that have been widely overlooked under the shadow of the ethical considerations that connected and autonomous vehicles demand. It has presented a new categorization for AI applications in road safety based on the tasks their respective AI models and techniques fulfil, including: future forecasting, correlation analysis, anomaly detection, group formation, optimization, and analytics support. It has identified possible ways that AI applications in road safety give rise to ethical concerns. It has shown that integration of AI in decision-making creates a large and complex socio-technical system that complicates the distribution of tasks and obligations. Moreover, AI systems have the potential to disturb conditions for blameworthiness of individuals (such as moral agency and transgression of norms).

AI outputs often lack epistemic explanation, and thus, their integration with human decision-making is often influenced by automation bias. They may reduce options of action, influence desires and norms, and degrade personal skills and self-trust. AI can reduce the control of individuals over their accessibility to others through the collection and storage of data. In addition, it can enable profound inferences that loosen the control of the individuals over the information about themselves. Lastly, limitations in training datasets and the design choices embedded in AI models can result in impartial outputs. Additionally, access and usability of AI systems are oftentimes not similar for different individuals.

As a theoretical output, WP4 has identified aspects pertaining to the road safety domain in particular that can make the use of AI in this context more ethically challenging, including: the serious consequences of wrongdoings, the need for well-grounded and traceable decision-making and action, the engagement with norms and desires, the embeddedness in everyday life, the extensive presence in public space, and the inequal current state.

The work package has not stopped at the theoretical level, though. It has established a tree-based methodology to review the existing AI analytic methods in road safety based on the scale of analysis (microscopic, mesoscopic, and macroscopic road safety assessment) and purpose of analysis (prediction or causal / associational inference). In addition, it has developed a conceptual methodology for benchmarking the AI methods and data in the form of a radar diagram with the following dimensions: prediction ability, explainability / interpretability, unobserved heterogeneity and endogeneity, feature selection, model assumptions, and computational efficiency.

In addition to the theory and methodology, WP4 has gone beyond the state-of-the-art in showcasing this methodology via a complete framework for automated road safety assessment in Low- and Middle-Income Countries (LMICs), addressing a fundamental challenge: building models that are both accurate and fair when critical infrastructure for vulnerable road users appears rarely in the data. Specifically, it has developed a deep learning pipeline for automated visual road assessment, validated across datasets around 560,000 road segments. It has presented an interpretable machine learning workflow combining gradient boosting models with SHAP analysis to transparently identify key risk drivers from structured road data. It has developed a reusable data engineering pipeline for processing raw road survey video, including a road-based splitting protocol that ensures valid generalization testing.

A summary of the main achievements, innovations, contributions and outcomes of each DC is presented in a table in the next section.

### 3 Scientific outputs and publications

During the first reporting period, the DCs have actively contributed to the dissemination of their research through academic manuscripts, conference presentations and journal submissions. Key outputs include papers (working, under review, or published / presented) demonstrating strong engagement with both academic and industrial communities, as shown on Table 1.

**Table 1: WP4 Scientific Outputs**

Doctoral Candidate	Scientific output
DC 1 – Bahareh Khajepour	<b>Conference paper (under review):</b> Artificial intelligence for road safety: Ethical implications, Road Safety and Simulation conference (RSS), Naples, 2026.
DC2 – Manjinder Singh	<b>Working papers (in progress):</b> A tree-based methodology for a systematic AI model selection based on the scale and purpose of analysis A radar-diagram conceptual model of benchmarking AI methods with road safety data
DC3 – Amirhossein Hassani	<b>Conference paper (accepted and presented):</b> Efficient Net-Swin Transformer for Automated iRAP Road Safety Attribute Extraction in Low- and Middle-Income Countries. The 12th International Congress on Transportation Research, 2025.  Journal article (published): Shahid, M., Gregurić, M., Hassani, A., & Ševrović, M. (2025). Optimizing Car Collision Detection Using Large Dashcam-Based Datasets: A Comparative Study of Pre-Trained Models and Hyperparameter Configurations. Applied Sciences, 15(13), 7001. <a href="https://doi.org/10.3390/app15137001">https://doi.org/10.3390/app15137001</a>

## 4 Next Steps and Ongoing Research

The DCs in this work package have created a living document on how the theories, methods and applications fit into each other and create a broader narrative for responsible AI in road safety. The document starts with one example among different ethical dimensions in AI for road safety: responsibility. It crafts the definition and theoretical need for responsibility and establishes the existing explainable AI methodologies (SHAP) for implementing responsibility in data analysis. It then showcases how these methodologies can be applied in a low-to-middle income country. The document is a living document and the DCs will be working on it during the next phases of the project. The plan is to add more ethical dimensions to the theoretical section of this document (by DC1), as well as additional methodological ways to embed the added ethical dimensions in the road safety analysis (by DC2), and finally to implement the added theories and proposed methodologies in more case studies in LMIC.