

Agentic Safety Verifier

Antje Loyal, Rujiao Yan, Mirosław Robakowski,
Christian Wirth (AUMOVIO SE)



SAFE AI
ENGINEERING

Motivation

Data-driven safety assurance for AI-enabled and automated systems requires the systematic **collection of quantitative evidence** across predefined operational scenarios, requirements, and safety objectives.

Standards such as *ISO 26262* [1] emphasize evidence-based safety arguments throughout the development lifecycle, while *ISO/PAS 8800* [2] extends this perspective to AI-specific safety considerations, highlighting the need for traceable and representative validation data.

However, **identifying relevant evidence within large-scale video repositories** remains a significant challenge.

Idea

The **Agentic Safety Verifier (ASV)** is an agentic framework that unifies different tasks along the complete lifecycle of an AI System. The ASV can be used to:

- Create a **scene description** and add (predefined) tags like ODD specific wording (Operational Design Domain)
- Perform statistical analysis of **ODD coverage** and identification of data gaps
- Carry out a **semantic video search** for specific scenarios
- Automatically **check for violation of safety constraints** or compliance with requirements
- Identify **ODD definition gaps**

Using semantic retrieval capabilities, the system enables engineers to efficiently **discover, organize, and trace scenario-specific evidence**, reducing manual review effort while increasing coverage and reproducibility.

By connecting safety requirements with relevant video observations, the approach supports **scalable validation workflows** and strengthens confidence in data-driven safety assessments for modern intelligent systems.

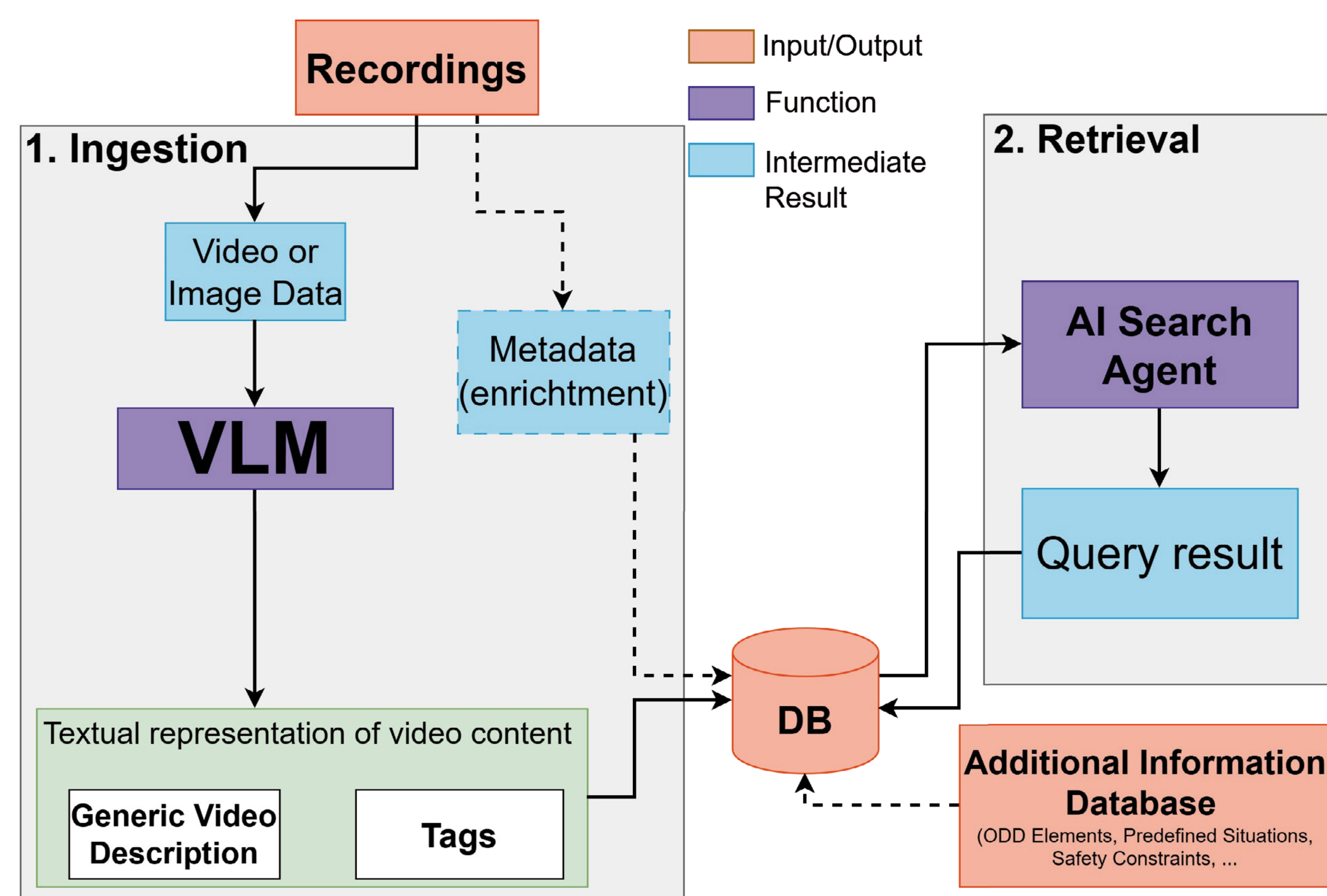


Figure 1: Building the database of textual representations by ingestion of VLM descriptions. Following this first step, various use cases can then be realized, here a semantic search is shown.(© AUMOVIO SE)

Architecture

- (Base) Build a metadata database, where each recording owns ODD-related and general tags, operational situations and safety constraints
- A **Vision Language Model (VLM)** is used to generate general scene descriptions and tags regarding operating environment for each recording
- Video retrieval with an **AI search agent** based on a free text query
- Recordings are **matched with predefined tags** (e.g. ODD definitions) and safety constraints (currently predefined according to *SOTIF* [3]).
- (Addition) With an **additional VLA** also planned actions or outputs from different models can be evaluated, even when the safety driver intervened

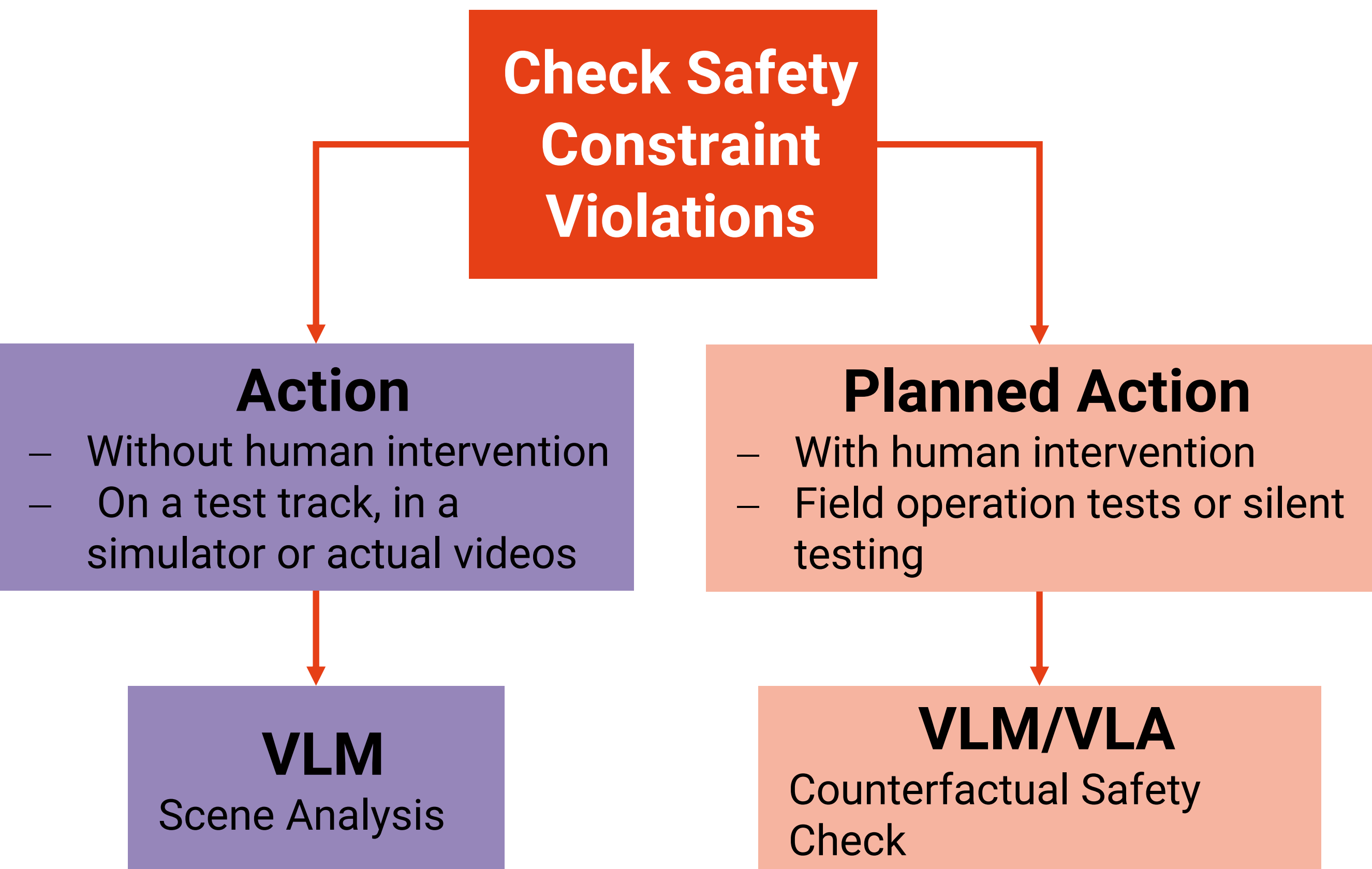


Figure 2: Two options of checking safety constraint violations. Real decisions and data or analysis of planned action without actual carrying them out. VLA: Vision Language Action (© AUMOVIO SE)

Outlook:

In the next steps we want to:

- Add 3D grounding for better 3D understanding
- Add counterfactual reasoning [4] for different planned actions and including model comparison
- Integrate VLA World Model [5] to also use videos from test drivers or other models to evaluate planned actions

References:

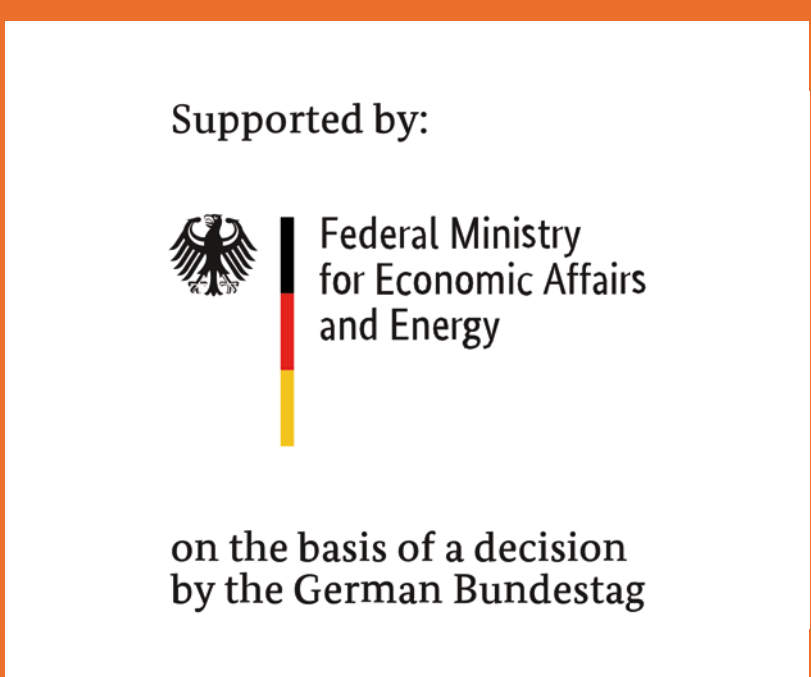
- [1] ISO 26262:2018 Road vehicles functional safety <https://www.iso.org/publication/PUB200262.html>,
- [2] ISO/PAS 8800:2024 Road vehicles – Safety and artificial intelligence <https://www.iso.org/standard/83303.html>
- [3] ISO 21448:2022 Road vehicles – Safety of the intended functionality <https://www.iso.org/standard/77490.html>
- [4] Z. Peng et al., "Counterfactual VLA: Self-Reflective Vision-Language-Action Model with Adaptive Reasoning", 2025
- [5] F. Jia et al., "DriveWorld-VLA: Unified Latent-Space World Modeling with Vision–Language–Action for Autonomous Driving", 2026

For more information regarding the poster, contact:

Antje.Loyal@aumovio.com,
Rujiao.Yan@aumovio.com



Safe AI Engineering is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Energy.



[linkedin.com/company/ki-familie](https://www.linkedin.com/company/ki-familie)
www.safe-ai-engineering.de