

Monitoring

Antje Loyal (AUMOVIO SE)
Dr. Xinxing Wang (Bertrandt Group)



SAFE AI
ENGINEERING

Motivation

Artificial Intelligence (AI) is increasingly embedded in safety-critical systems, where it must operate under strict risk constraints while remaining performant and adaptive. Monitoring perception systems by systematically analyzing outputs and operating context to **detect, control or mitigate unsafe behavior**, is therefore a crucial safety measure.

Monitoring provides an **additional safety layer** that helps identify potential failures during operation.

Application of Monitors

There is no single type of “AI Monitor”. Monitors exist for different scenarios, lifecycle steps or inputs and outputs in general. In our project we distinguish between:

- **Monitors** that can influence driving decisions directly and
- **Field monitors** that collect information over time e.g., to trigger a retraining.

What most of the monitors have in common, is that they **estimate the uncertainty** of the data or the model output.

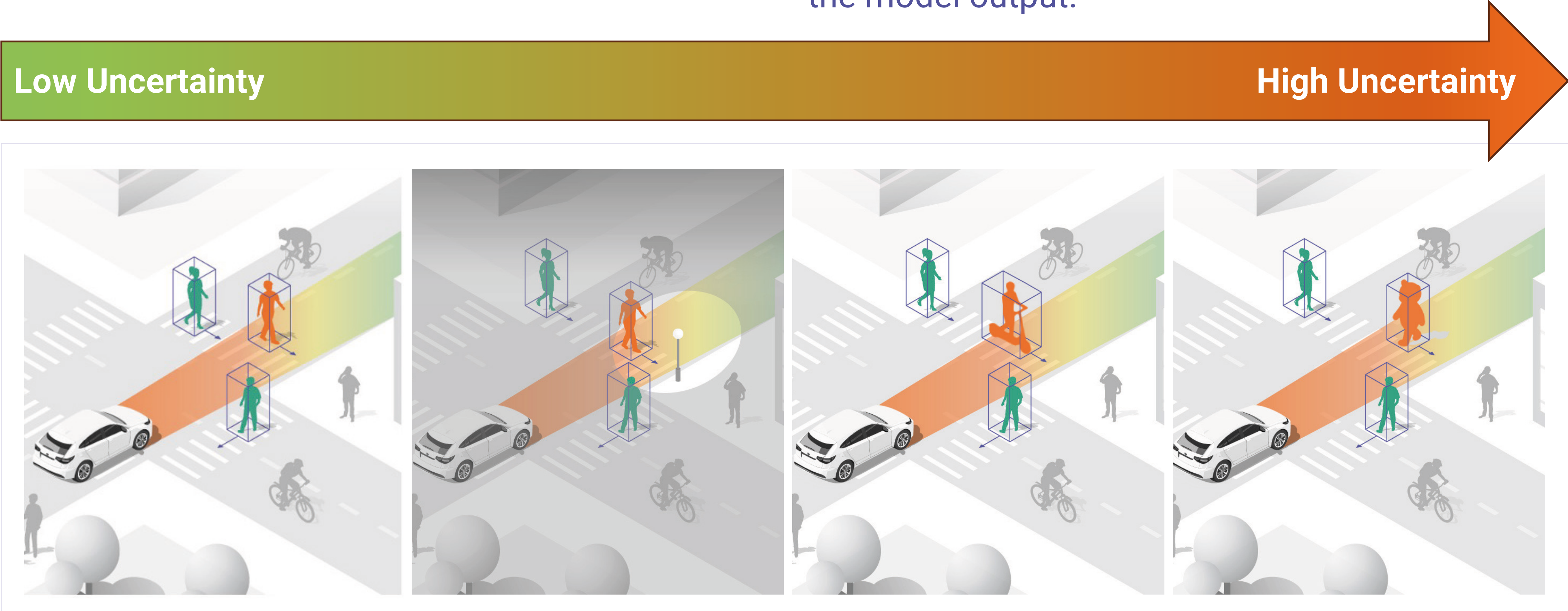


Figure 1: Different Scenarios for the application of Monitors. Scenarios ranging from low uncertainty to high uncertainty (© Safe AI Engineering)

Scenario 1: Known Scenarios

Known scenarios are scenarios that are well represented in the data. The uncertainty should be low and the trust in the system high. The Monitor verifies that the system remains in familiar conditions.

Scenario 2: Anticipated but Hard-to-Cover

The training data can only cover a small fraction of the endless possibilities in an open world, but some variations can be anticipated. This can be different lighting or weather conditions or occlusion and crowd density. The Monitor estimates the uncertainty and indicates whether the prediction remains reliable or not.

Scenario 3: Evolving Environment Over Time

The real world is changing over time, new traffic participants or infrastructure emerges. Field Monitors monitor long-term and identify changes or can tell, when a system or model needs adaptation.

Scenario 4: Unknown and Unexpected

Those scenarios can be the most critical events, because they were not considered during development, are not or underrepresented in the training data and the output of the AI model can be confidently wrong. These are also the scenarios where a monitor can have the greatest impact, as it can identify high-uncertainty situations and alert the system not to rely on the AI model's output.

Monitors in Project Context

We evaluated over **20 different Monitors**, where 11 can be seen at Posters, Demonstrators or integrated in our project-wide App. 7 Monitors are Monitors running at runtime in the car and 3 are Field Monitors. With that we cover the full range of **Scenario 1 to Scenario 4**.

Additionally, some of the monitors can already be applied during testing and give insight into model performance in general.

Monitors to experience at IAA:

Field Monitors	Monitors	
Adversarial Loss Monitor	Information Quality Monitor	Perturbation Detection using Hadamar Output Encoding
Test Time Data Augmentation	Metadata-based Monitoring	Failure Detection via Spatial & Temporal Consistency Check
Statistical Active Learning Search	Adversarial Attack Monitor	Detection Output Monitor
Agentic Safety Verifier	Neural Activation Coverage Monitor	

For more information regarding the poster, contact:

Antje.loyal@aumovio.com,
Xinxing.Wang@bertrandt.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