

# Detecting the Unexpected: Monitoring for Autonomous Driving

Anna Bieberstein, Xinyi Li , Riken Mangukiya,  
Christian Freystein, Dr. Xinxing Wang,  
Jochen Schwenninger (Bertrandt Group)



**SAFE AI**  
ENGINEERING

### Metadata-Based Input Monitor

The metadata monitor continuously analyzes contextual information of the driving scene, such as vehicle state, camera settings, image quality, and environmental conditions. Using a model trained only on normal operating metadata, it assigns an anomaly score to each frame that indicates how strongly the current situation deviates from known conditions. High scores suggest potentially out-of-distribution (OOD) scenarios.

**Input:** Scene metadata feature vector

**Model:** Feature Preprocessing, Autoencoder (AE), Gaussian Mixture Model (GMM)

**Output:** Anomaly score, ID / OOD decision

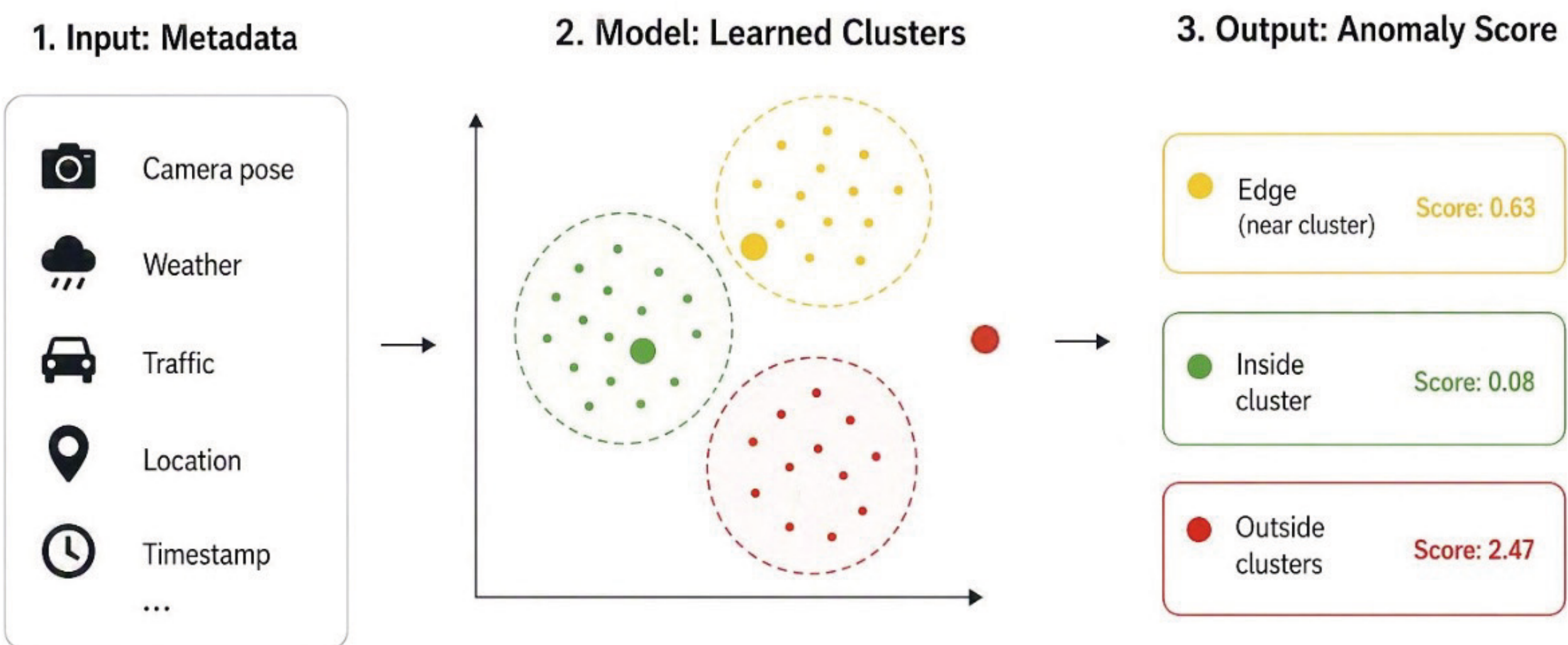


Figure 1: Metadata-Based Monitor – From metadata vector to anomaly score (© Bertrandt Group) \*

### Safety-Related Impact

The monitor can detect previously unseen or unusual operating conditions in real time, providing an additional safety signal whenever perception performance may become less reliable.

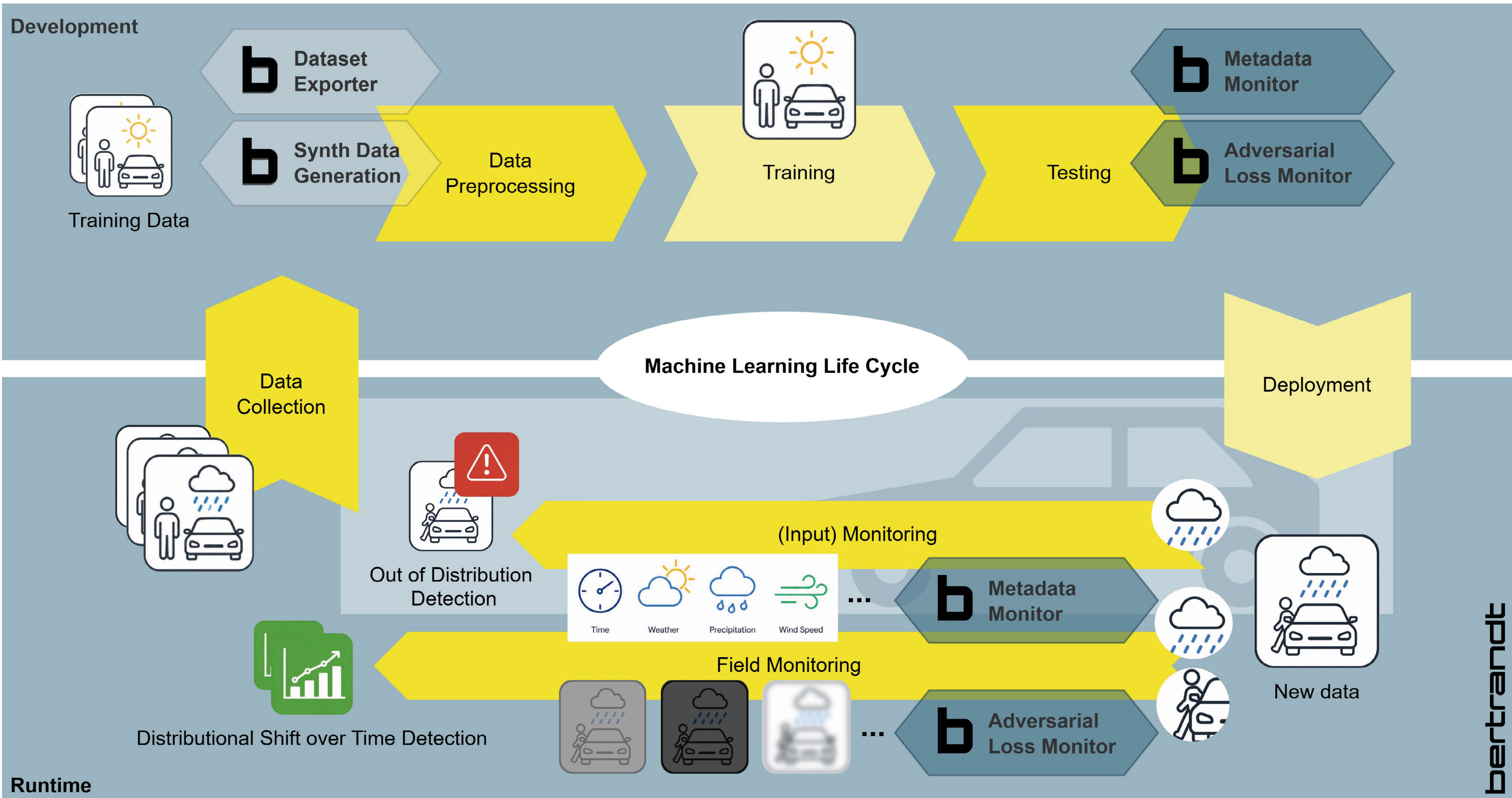


Figure 2: Machine Learning Life Cycle (© Bertrandt Group) \*

### Adversarial Loss Field Monitor

This monitor is a post-hoc uncertainty estimation method that assesses the reliability of object detector predictions without modifying the detector itself.

The key idea is that reliable detections remain stable under small, semantically preserving image perturbations, while unreliable detections

show reduced confidence, degraded localization, or disappear.

Multiple perturbed image variants are analyzed, and changes in confidence and bounding box positions are combined into a single uncertainty score. High uncertainty indicates that a prediction is sensitive to small input variations and may therefore be less trustworthy.

### Safety-Related Impact

By identifying detections whose confidence and localization are unstable under small input perturbations the monitor enables identification of safety-critical performance degradation.

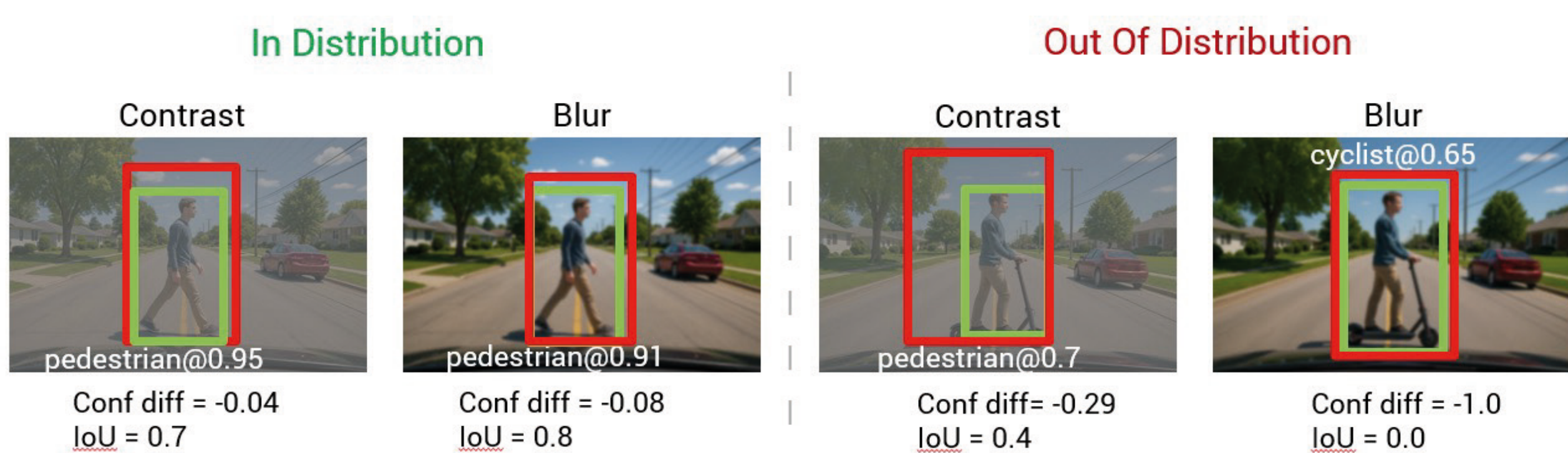


Figure 3: Conceptual illustration of confidence and localization degradation under OOD conditions (© Bertrandt Group) \*

\* Images are partly generated with AI

For more information regarding  
the poster, contact:

anna.bieberstein@bertrandt.com,  
xinyi.li@bertrandt.com,

**bertrandt**

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](https://www.safe-ai-engineering.de)