# Coverage Metrics for a Scenario Database

## for the Scenario-Based Assessment of Automated Driving Systems

**IEEE IAVVC 2024**
IEEE International Automated Vehicle Validation Conference
October 21–23, 2024 | Pittsburgh, PA

**SUNRISE** — SAFETY ASSURANCE FRAMEWORK FOR CONNECTED, AUTOMATED MOBILITY SYSTEMS

**Maren Buermann, Erwin de Gelder & Olaf Op den Camp**

💼 TNO
📍 The Netherlands
✉ maren.buermann@tno.nl

TNO innovation for life

## (1) Introduction

► Scenarios offer a structured way to describe the large varieties of situations and conditions that Automated Driving Systems (ADS) may encounter on the road. Figure 1 describes how scenarios are extracted and used for testing ADS.

► The trustworthiness of the safety assessment results of ADS depends on the quality of the selection of test scenarios, and consequently depends on how well the underlying data for scenario identification and collection, as well as the selected set of scenarios, cover the Operational Design Domain (ODD).

This poster describes 2 types of Coverage Metrics to quantify how well the ODD is covered
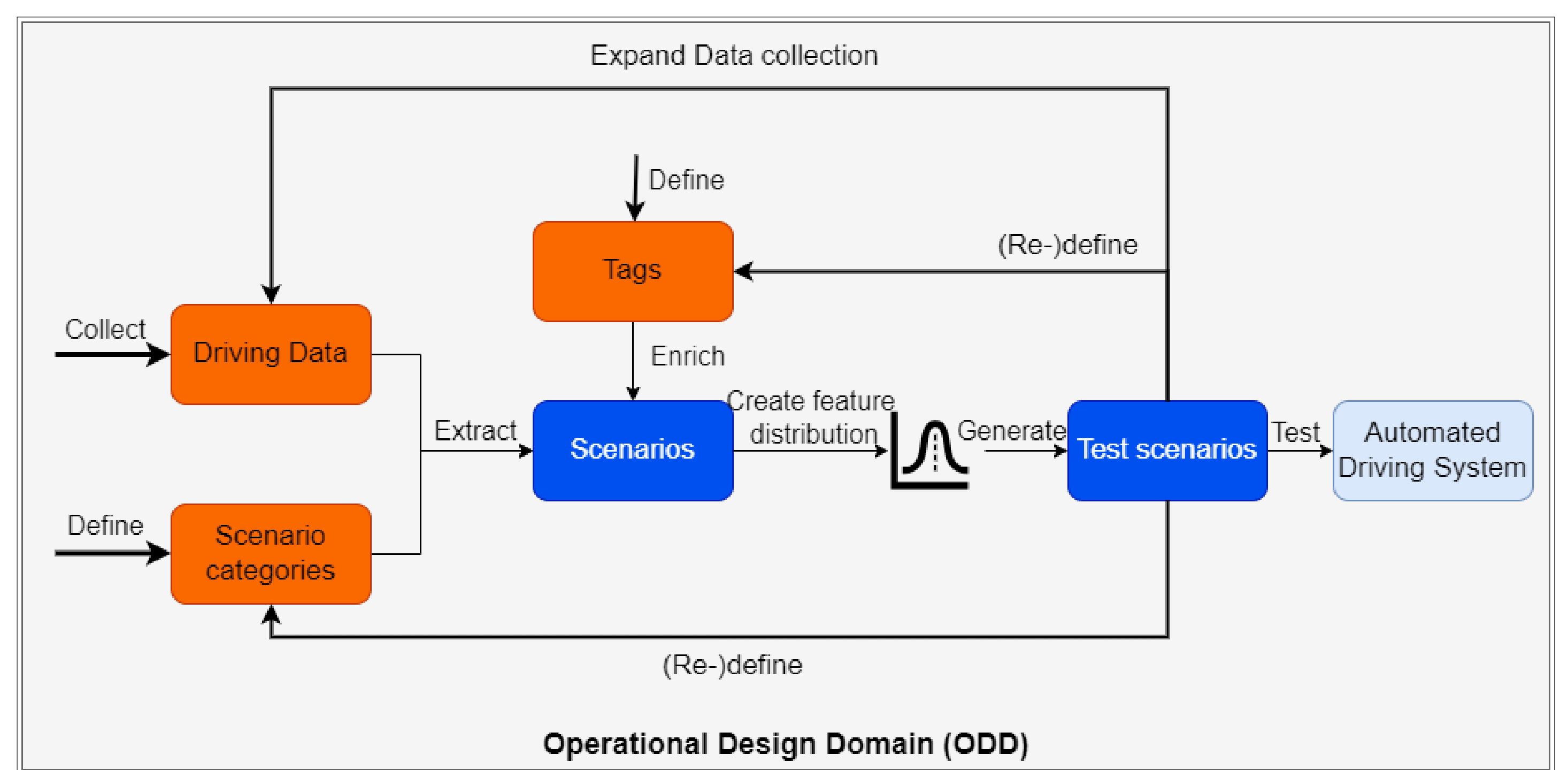


Figure 1: Scenario based safety assessment method

## (2) Coverage metrics

**Type I coverage metric:** Does the data contain scenarios in a wide enough range of conditions to cover the ODD?

**Type II coverage metric:** Do the scenarios from the defined scenario categories cover everything that happens in the data?
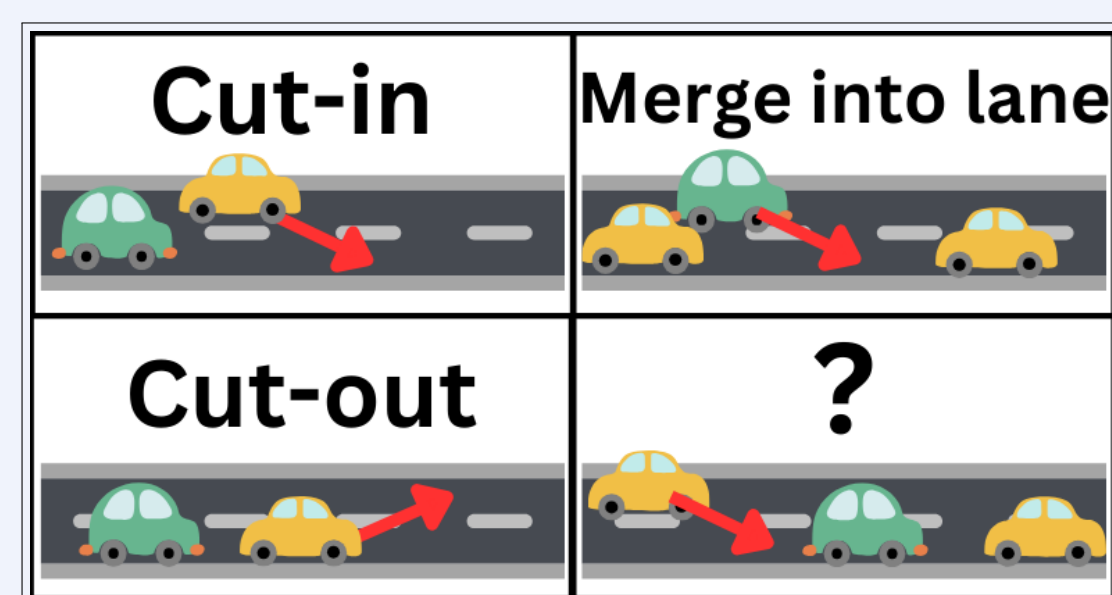


Figure 2: Type I coverage purpose



Cut-in | Merge into lane
Cut-out | ?

Figure 3: Type II coverage purpose

## (3) Coverage metrics

**Tag-based coverage (Type I):**
► Calculates the percentage of scenario categories having at least $n$ scenarios containing the desired tags.

**Time-based coverage (Type II):**
► Calculates the percentage of timestamps in the data, at which at least $n$ scenarios occur.

**Actor-based coverage (Type II):**
► Calculates the percentage of actors that are "relevant" and also part of a scenario.

**Actor-over-time based coverage (Type II):**
► Calculates the percentage of actors that are "relevant" **while** being the target in a scenario.

## (4) Experiment setup

► 10 scenario categories and 18 tags are defined
► The HighD data set is used, with more than 40,000km of driving data and more than 100,000 vehicles
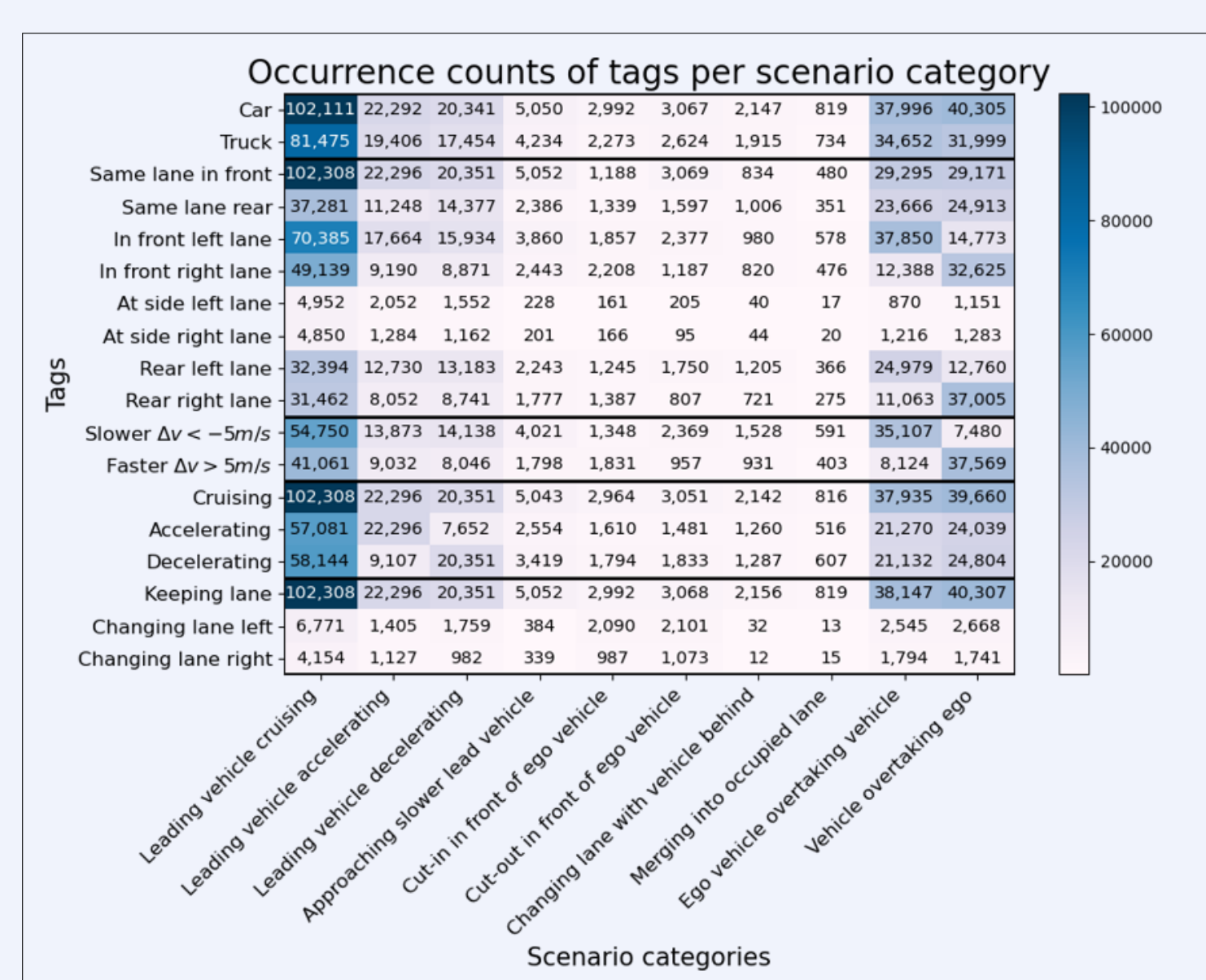► More than 200,000 scenarios were extracted (see Figure 4)



Figure 4: Results of scenario extraction
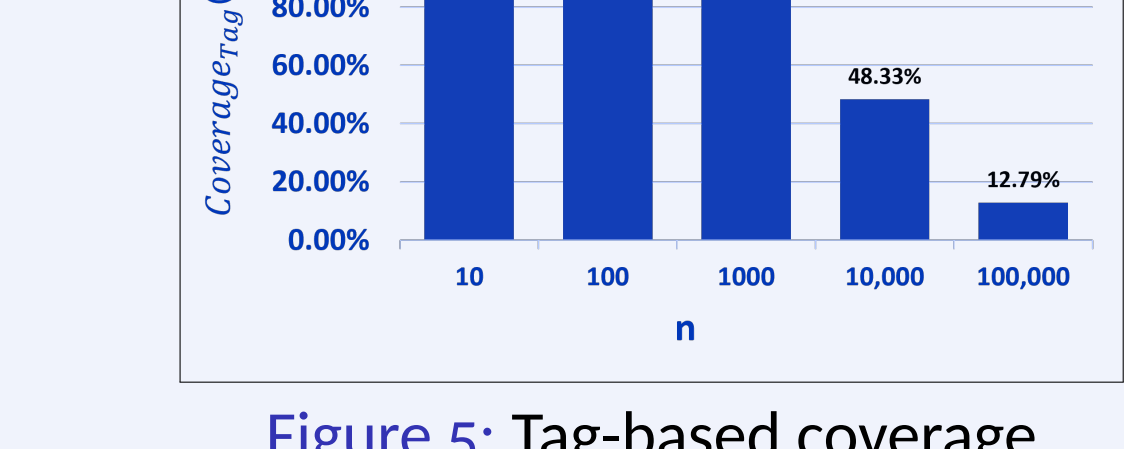
## (5) Results



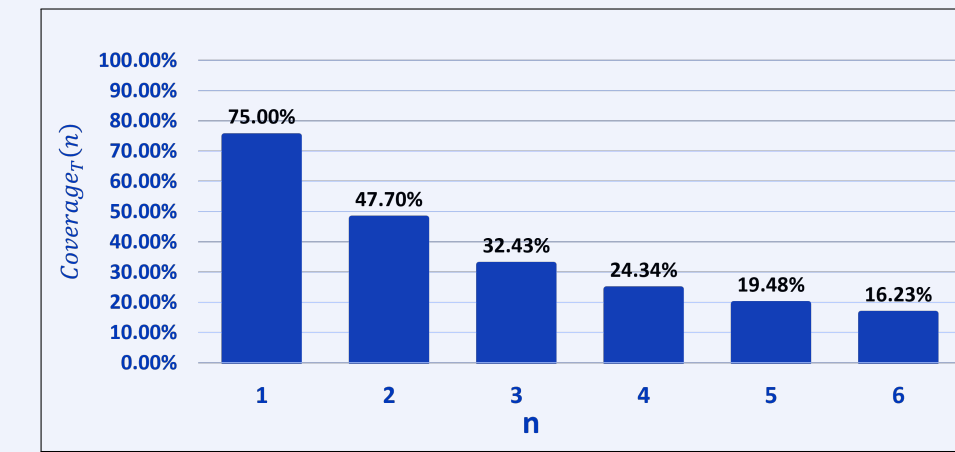Figure 5: Tag-based coverage



Figure 6: Time-based coverage

For Figure 5 and 6, the coverage decreases for larger n.
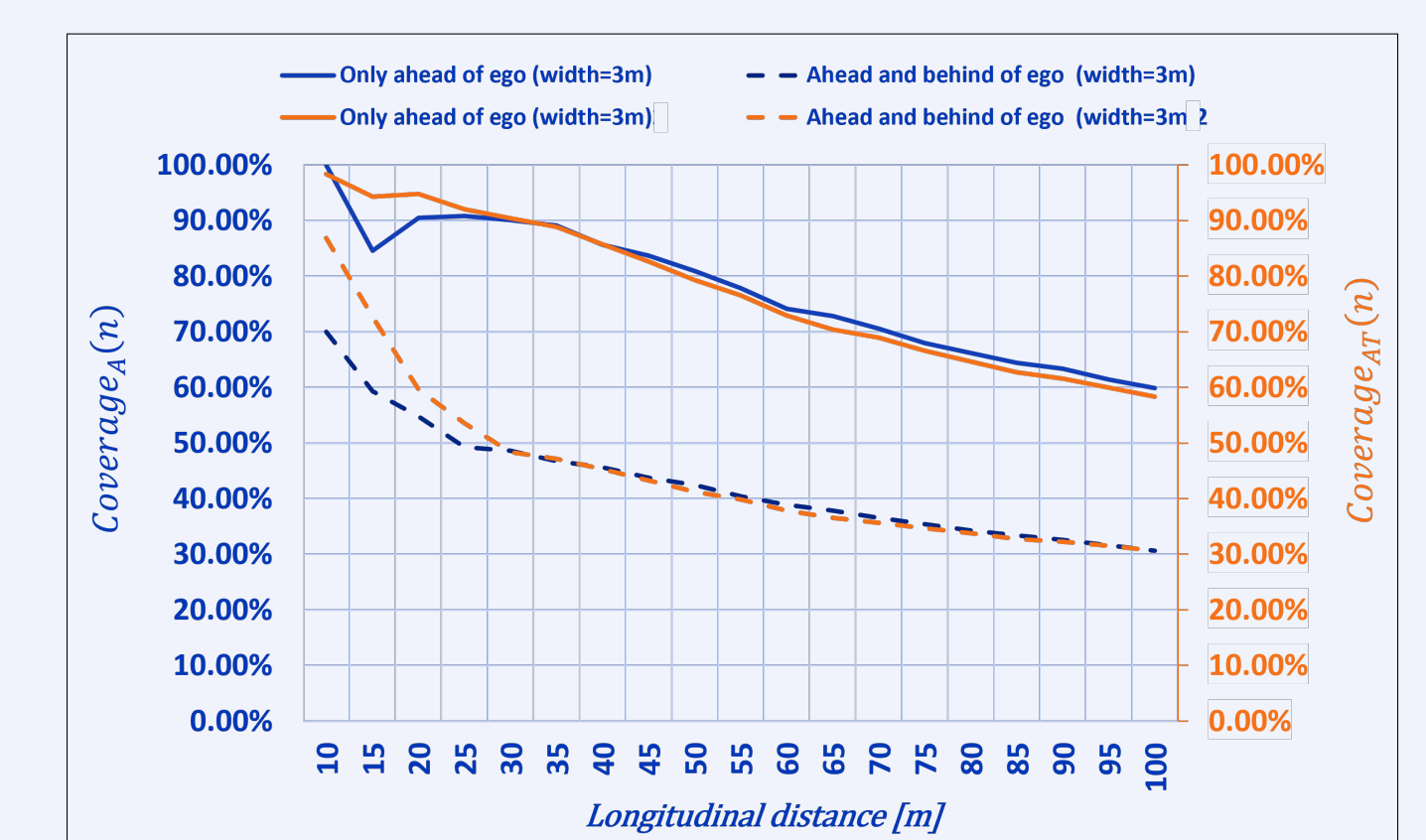


Figure 7: Actor (blue) and Actor-over-time (orange) based coverage

For Figure 7, the coverage shrinks as the longitudinal range in which actors are considered to be "relevant" increases.

## (6) Conclusions

► It is import to evaluate the scenario extractor. False positive/negative detection will affect coverage.

► Achieving high tag-based coverage for larger n is difficult, but might not be necessary. E.g.: Some tags can (theoretically) only occur for certain scenario categories

► For $n = 1$ time-based coverage, a very generic scenario category can achieve a coverage of 1.

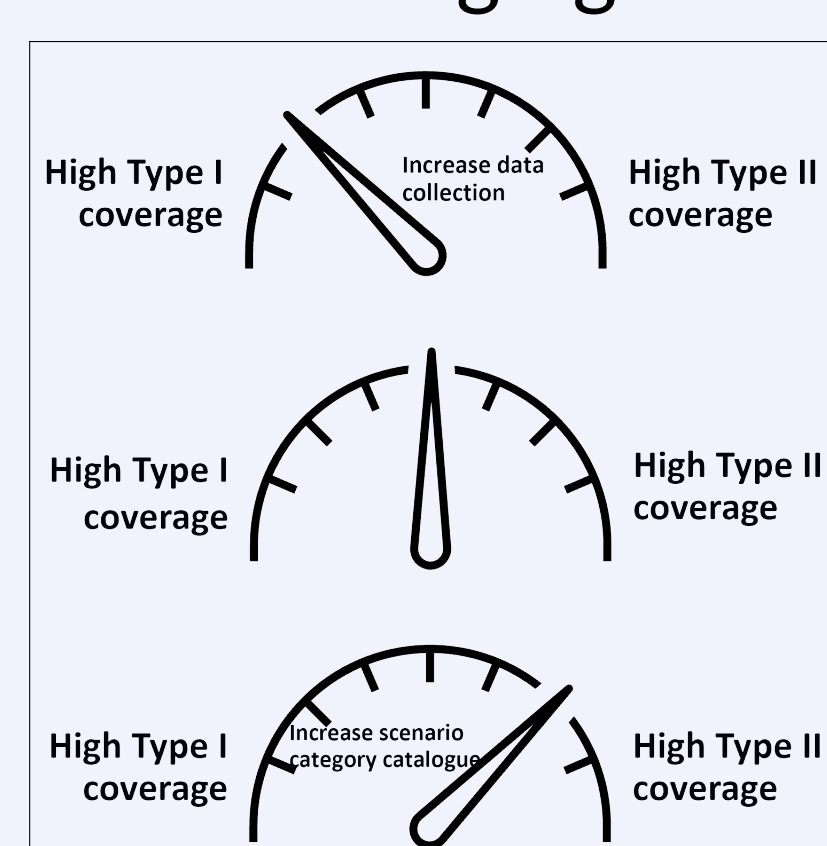► Achieving high Type I and Type II coverage is challenging



Figure 8: Coverage trade-off

## (7) Future Work and Additional Information

► This poster outlined two types of coverage metrics, which can:
  ► quantify the extent to which the data and the scenarios derived from them cover the ADS' ODD
  ► help to identify missing data or scenarios that should also be considered for the safety assessment

► Achieving 100% for all coverage metrics might not always be practical or necessary

► Future work should :
  ► Focus on establishing suitable coverage thresholds
  ► also be dedicated to the completeness of driving data and the identified scenarios

TNO innovation for life