Assurance arguments for AI systems [27]

Evaluation of the assurance argument and recommendation for release [4]

Using GSN (Goal Structuring Notation)
Include **Assurance Claim Points (ACP)** to highlight areas of:
•Weak evidence
•Uncertain assumptions
•Incomplete inference
🛠 **Tool Support**: Use tools like GoalStruct or Safety Argument Editors to maintain traceability.

Exercise:
You are part of a release review board. An AI-based lane departure warning system has been tested in dry conditions but not yet in snow or heavy rain.

Question:
Would you approve a release? Under what conditions? What would you require as conditional acceptance?

•Identify missing evidence
•Classify assertion types (context/evidence/inference)
•Propose a release strategy and enhancement plan

Assurance arguments for AI systems [28]

Evaluation of the assurance argument and recommendation for release [5]

Answer to the exercise to the AI Lane Departure Warning System

Scenario Recap: An AI-based lane departure warning system has been tested under dry conditions but not yet in snow or heavy rain.

Step 1: Identify Missing Evidence

•**Context Assumption at Risk**: The operational domain includes weather variations (snow, heavy rain), but current evidence is limited to dry conditions.

•**Evidence Gaps**:

- No test results for adverse weather.
- No robustness analysis across weather-induced perception changes (e.g. obscured lane markings).

•**Risk**: Safety requirements might not be met under snow or rain conditions → Potential *functional insufficiencies* and *SOTIF risks*.

Assurance arguments for AI systems [29]

Evaluation of the assurance argument and recommendation for release [6]

Step 2: Classify Assertion Types

Assertion Type	Issue
Asserted Context	Assumption: AI performs similarly in snow/rain as in dry conditions. → **Unverified**.
Asserted Evidence	Test dataset lacks distribution over full environmental conditions.
Asserted Inference	Safety argument doesn't address weather-based degradation or mitigation measures.

Assurance arguments for AI systems [30]

Evaluation of the assurance argument and recommendation for release [7]

Step 3: Conditional Release Proposal
Release could be **conditionally approved** with the following:
📌 **Conditions for Release**

1.Restricted ODD:
➤ Operational Design Domain limited to dry-weather use.
➤ Feature disabled automatically when rain/snow is detected (via weather sensors or V2X).

2.Monitoring Plan:
➤ Log environmental data and system outputs during real-world operation.
➤ Identify performance issues in adverse conditions using fleet data.

3. Action Plan:
➤ Collect and label snow/rain data.
➤ Retrain and validate AI with augmented dataset (synthetic + real).
➤ Update assurance argument including:
1. Confidence intervals for degraded conditions
2. Safety margin justifications
3. Out-of-distribution detection or fallback mechanisms

4. Documentation Update:
➤ Assurance argument must clearly state:
1. Contextual limitation (asserted context)
2. Evidence limitation (asserted evidence)
3. Planned closure strategy

Final Assessment
✔ Conditional acceptance is possible.
❗ But it **must include traceable mitigation and enhancement measures**, and **must not allow false safety perception** in unverified environments