Overview

- Autonomous Vehicle safety standards
  - ISO 26262 & ISO 21448
  - ANSI/UL 4600
  - SAE J3018

- The hard bits beyond that are:
  - Fail operational architecture
  - Building an accurate, predictive world model
  - Safety beyond the driving task
  - How safe is safe enough?
Core AV Design Standards

- **ISO 26262 – Functional Safety**
  - Covers run-time faults & design defects
  - Assume requirements are complete

- **ISO 21448 – SOTIF**
  - SOTIF: “Safety Of The Intended Function”
  - Iteratively discover & mitigate unknowns

- **ANSI/UL 4600: #DidYouThinkofThat?**
  - A technically substantive safety argument
  - Evidence of coverage initially + feedback from surprises
  - Aggressive field feedback based on lessons learned

---

https://bit.ly/3NNwLO1
### Standards-Based Engineering Approach

<table>
<thead>
<tr>
<th>System Safety</th>
<th>ANSI/UL 4600</th>
<th>Safety Beyond Dynamic Driving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic Driving Function</td>
<td>ISO 21448, SaFAD/ISO TR 4804</td>
<td>Environment &amp; Edge Cases</td>
</tr>
<tr>
<td>Functional Safety</td>
<td>ISO 26262</td>
<td>Equipment Faults</td>
</tr>
<tr>
<td>Cyber-Security</td>
<td>SAE J3061, SAE 21434</td>
<td>Computer Security</td>
</tr>
<tr>
<td>Vehicle Safety</td>
<td>FMVSS, NCAP</td>
<td>Basic Vehicle Functions</td>
</tr>
</tbody>
</table>
“Fail Safe” (fail stop) is not enough
- Detect failure
- Switch over to a redundant capability
  - E.g., gracefully terminate mission

Safety architecture challenges
- “Redundancy” is not necessarily enough
- Safety limited by common mode failures across the redundancy
  - “Diversity” is difficult to measure in all dimensions
- If two computations disagree, which do you believe?
  - Disagreement is likely for nondeterministic algorithms

https://bit.ly/3VcFzRs
Perception Limits To Safety

**THE REAL WORLD**

- **VIDEO**
- **LIDAR**
- **RADAR**

**PERCEPTION**

**COMPUTER’S WORLD MODEL:** “Child chasing ball into street 10 meters ahead”

**Path Planning & Motion Control**

**Perception & prediction present a uniquely difficult assurance challenge**
Safety Requires an Accurate World Model

- Good prediction based on the world model
  - Classification accuracy affects prediction
  - Probability cloud for object motion
- Safety limited by heavy tail scenarios (rare, important)
  - Probabilities might be context dependent
  - Rare cases tend to dominate safety
Field Engineering Feedback

- Architectures will need to support lifecycle field feedback
  - Safety Performance Indicators (SPI) data linked to safety case
    - Transition from safety recall model to continuous improvement
SPIs and Lifecycle Feedback

- SPI: direct measurement of safety case claim failure
  - Independent of reasoning ("claim is X ... yet here is ~X")

- A falsified safety case claim:
  - Safety case has some defect

- Root cause analysis might reveal:
  - Product or process defect
  - Invalid safety argument
  - Issue with supporting evidence
  - Assumption error

- Continual Safety case improvement
Role of Humans

- There is no “captain of the ship”
  - Autonomy must assume responsibility

- Interacting with people
  - Occupants, cargo loading
  - Pedestrians & mobility device users
  - Potential abuse, misuse

- Role of humans as drivers?
  - Remote operators and wireless data have their limits
  - Avoid “Moral Crumple Zone” operational concept

- Safety culture for all stakeholders

Is it safe to drive now?

https://bit.ly/2GvDkUN
Safe Behavior & Safe Enough

- Contextual safety for safe vehicle shutdown
  - Is in-lane stop in fast moving highway “safe”?
  - What if stopped AV blocks an emergency vehicle?

- Where is the “safe enough” bar set?
  - Better than human, but...
    - Prediction uncertainty
    - Equity & risk redistribution issues
  - Safety engineering reduces uncertainty
  - Field feedback of SPIs manages uncertainty

- Governance model: who decides to deploy?
  - What basis is used for decision?
Summary

- Follow safety standards for a foundation
  - Identify & mitigate hazards
    - Within vehicle
    - Presented by operational environment
    - At system level, beyond driving task
  - Safety engineering beyond just road testing

- Be prepared to wrestle with these parts:
  - Fail operational architecture
  - Accuracy of building a world model
  - Safety beyond the driving task
  - How safe is safe enough?