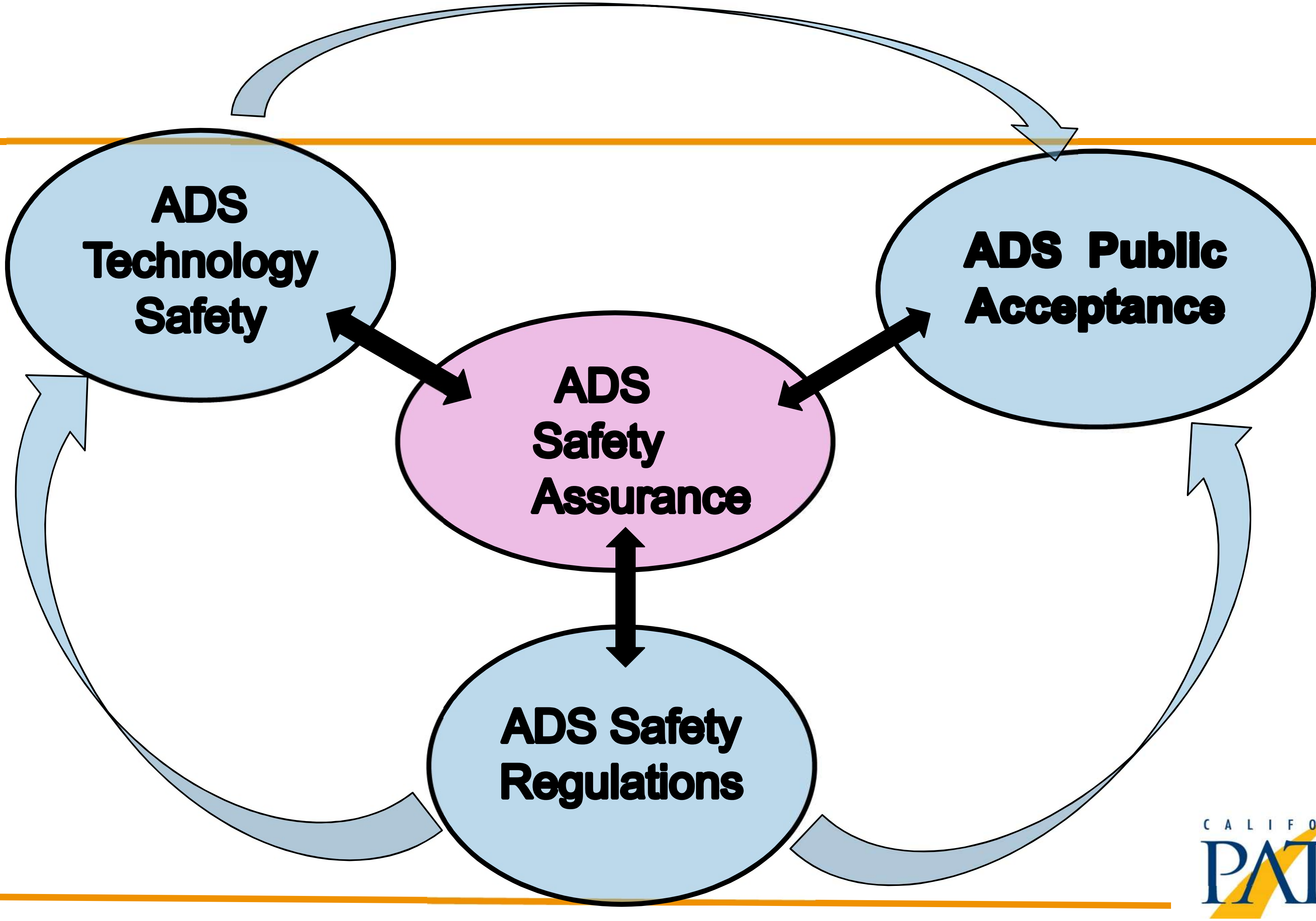
Safety Assurance to Earn Public Trust: Formalizing the Safety Case for ADS (Automated Driving Systems)

Steven E. Shladover, Sc.D.
California PATH Program
University of California, Berkeley
V&V Methods Mid-Term Meeting
March 16, 2022


ADS Technology Safety
ADS Public Acceptance
ADS Safety Assurance
ADS Safety Regulations

The Safety Case Context

Societal Inputs :
Safety benchmark
What metrics to apply to ADS to compare?
How much safer does it need to be?
What stakeholders must be engaged?

Safety Case Development
(Technical analyses, prioritized)
- Functional safety analyses
- SOTIF analyses
- Safety Management Systems
- Proving ground test results
- Public road test results
- Simulation results

How to explain outputs accurately and convincingly ?

Earning stakeholder trust
- Corporate risk managers
- Safety regulators
- General public and traffic safety advocates

Need to Define Safety Benchmark up Front

- **Start from today's traffic safety**
  - **Well documented, large data sample (statistically valid)**
- **Easy to explain to regulators and general public**
- **Good basis for starting discussions about how much safer ADS need to be**

- **Central challenge: How to estimate safety of ADS for comparison with the baseline?**

Desired Outcomes from Safety Case

- **Goal: <u>Earn</u> the trust of safety regulators and the general public so that they can be legitimately assured of ADS safety before deployment.**
- **Objectives:**
  - **Demonstrate due diligence applied to ADS development and deployment by following best safety practices (UL4600, ISO 26262, ISO 21448)**
  - **Produce quantitative evidence of safety case credibility**
  - **Use leading measures to show expected traffic safety improvement from ADS deployment**

CALIFORNIA PATH

Need for leading measures of effectiveness

- **Testing of prototype ADS cannot produce sufficient data within reasonable time and cost (RAND study)**
- **Direct comparison of ADS performance with human performance in specific safety scenarios is not viable**
  - Cannot represent huge diversity of human performance realistically in models or tests
  - Safety-critical scenarios amplify randomness and diversity in human behavior
  - Driving simulators lack realism in extreme conditions
  - Ethical constraints on use of human test subjects

CALIFORNIA PATH

Potential leading measures of effectiveness

- **Demonstrated ability of ADS to *avoid* crashes in specific challenging scenarios**
  - Proving ground tests of ADS
  - Simulations (if simulation can be validated)
- **Demonstrated ability of ADS to *significantly mitigate severity* of crashes in specific *very* challenging scenarios**
  - Proving ground tests of ADS
  - Simulations (if simulation can be validated)

Leading and Trailing Measures – Trade-offs

	Leading (Pre-deployment scenario-based assessments)	Trailing (Post-deployment real-world experience)
Baseline (Human driving)	- Human driving in hazard scenarios is too diverse and complex to model realistically - Realistic experiments would be too dangerous and costly	Current aggregate traffic safety statistics: - Well documented and understood - Huge sample (statistically robust)
Automated Driving	Predicting ability to respond to hazardous scenarios: - How to identify scenario set that can adequately represent real-world hazards? - How to develop and validate sufficiently realistic simulations?	- Too late to be useful for deciding on deployability - Very limited samples, under limited conditions, - Data not open to public scrutiny

Summary of KeyTechnical Challenges

- **How to produce real data to show (quantitatively) that a prototype/design ADS will improve traffic safety, so it should be deployed?**
  - Selecting the most relevant *leading measures* of effectiveness to compare to the baseline *trailing measures* of crash rates of different severities?
  - What range of scenarios will need to be simulated and tested to produce sufficient data?
  - What mix of testing and simulation is needed?
  - How can simulations be validated to a sufficient level that their results can be trusted?

CALIFORNIA PATH

Start as simple as possible

- **Limited ADS functionality within limited ODD conditions to bound complexity of relevant scenarios**
  - Start with scenarios from current crash data
  - Add scenarios based on available information about near-misses under current conditions
  - Add scenarios based on ADS fault conditions from functional safety assessments
  - Add scenarios based on potential external hazards from SOTIF assessments
  - For all scenarios, do parameter variations

CALIFORNIA PATH

Parameter Variations in Scenarios

- Crashes are rarely under "mean value" conditions
- Assessments must account for wide variations in:
  - Initial location and velocity of every mobile object
  - Condition of road markings and signage
  - Presence of static objects on and near the road
  - Weather, lighting and electromagnetic environment
- How many combinations of these variations and how far out on the tails of the distributions?
- How many to deter gaming by "design to the test"?
- What success percentage needed to "pass"?

CALIFORNIA PATH

If using simulation, how to validate it?

- **Crash-imminent situations stretch simulations beyond their normal validity (extreme conditions, nonlinear performance)**
- **What tests are needed to produce a validation data set containing those extreme combinations of conditions?**
  - How can they be done safely?
  - Can validation be done at component or subsystem level?
- **How closely do simulations need to match test data to be considered "valid" for safety assurance?**

CALIFORNIA PATH

Limitations in Realism of Simulations

- **Sensor phenomenology – anomalies based on noise, EMI, bad lighting (low sun angle, specular reflections), poor target resolution,…**
- **Vision-specific errors – shadows, foreign objects on road, reflections, glare, worn or occluded signs and markings**
- **Actions of other road users to try to avoid crash**
- **Vehicle imperfections – worn components, tire contact friction, suspension bottoming…**
- **Road geometry and surface condition imperfections**
- **Driver override interventions**

CALIFORNIA PATH

Plenty of efforts still needed…

- **Developing processes for engaging stakeholders to agree on safety criteria**
- **Extrapolating to predict real-world ADS safety based on results for limited (affordable) scenarios**
- **Methods for simulating ADS safety-critical scenarios and validating them to an acceptable level of fidelity**
- **Methods for combining simulation and testing to produce believable real-world ADS safety estimates**
- **Methods for explaining ADS safety case findings to regulators and the general public**

CALIFORNIA PATH

International Harmonization Topics

- Safety baseline(s) – variables (not numerical values)
- Relevant *leading measures* of effectiveness of ADS safety (and how to estimate them)
- Standards on validation of ADS safety simulation models
  - Validation methods
  - Validation measures of effectiveness and passing criteria
- Standards for selection of ADS scenarios
  - Criteria for prioritizing relevance to real-world safety
  - Criteria for determining sufficient variety and number of scenarios to support the safety case