Building Trust

Stakeholder Trust

Carnegie Mellon University

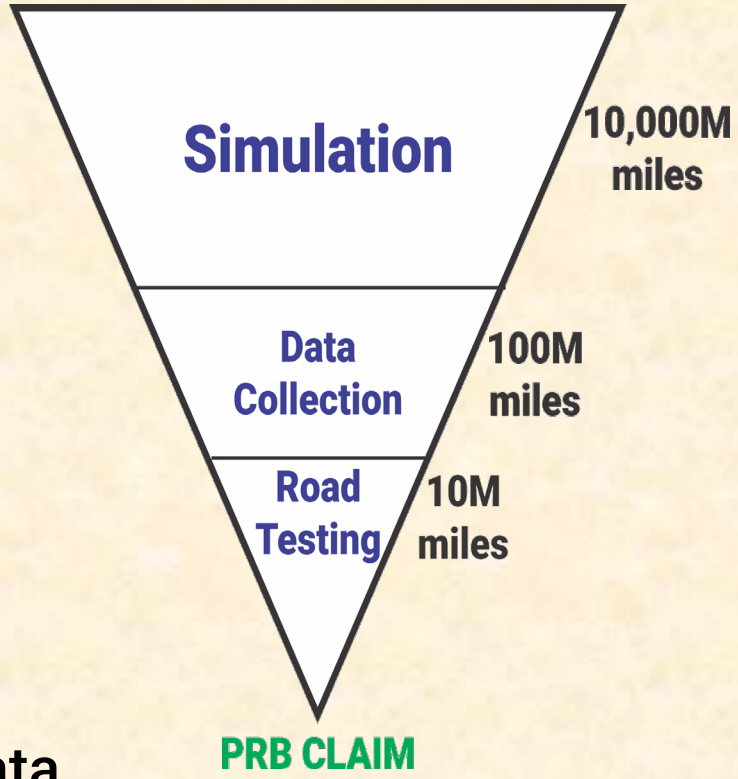
- Would you trust your life to an automated vehicle?

Ford VSSA https://bit.ly/3njionT

Hypothetical Validation Campaign

Carnegie Mellon University

- 10,000M mile simulation campaign
  - Goal: under 1 fatality/billion miles
  - Claim ~5-10x better than human
- 100M mile collected data/scenarios
  - Claim simulating this is representative
- 10M road testing of final software
  - Claim this validates simulation



- Is this statistically valid?
  - Questionable confidence in collected data
  - Road testing useful, but insufficient on its own

How Much Do You Trust Validation?

- Would you put a child in front of this self driving car?
  - 10,000M mile sims
    … perhaps with a simulator error?
  - 100M miles data collected
    … perhaps with scenario analysis errors?
  - 10M of road testing
    … that missed the above errors?
  - 10K repetitions of closed course testing
    … with standard dummies instead of people
  - With biased perception training data?
  - Built from software binaries & tools
    … with no safety qualification?

Engineering Rigor

- Testing alone is insufficient for life-critical systems
  - So we use also use engineering rigor
- Can you trust the system itself?
  - Is it engineered for safety?
  - Were standards and best practices used?
  - Is there a safety case documenting all this?
- Can you trust your validation process?
  - Did you engineer the simulations properly?
  - Did you design the validation campaign properly?

Field Engineering Feedback

- Expected risk has a mean + uncertainty
  - You should deploy only when mean is acceptable
  - But there will be uncertainty
    - Missed edge cases during road testing
    - Unknown gaps in validation plan
    - Unknown unknowns in general
- Solution: continuous field monitoring
  - Monitor Safety Performance Indicators (SPIs)
    - SPI violation means safety argument has a defect
    - Investigate and fix root causes before loss events
  - Start during validation; continue after deployment

Safety Culture

Carnegie Mellon University

- Did you do what you said you did?
  - Did your validation skip over known problems?
  - Did your engineering team skip process steps?
  - Is your field monitoring ignoring SPI violations?
- Good safety culture mitigates risk
  - Having a Safety Management System is ***a start***
  - Safety culture involves everyone in the lifecycle
- Safety culture simplified:
  - Are you incentivized to do the right thing?
  - Is it OK to tell your boss bad news? Will your boss fix it?

https://bit.ly/3i5wl57

Investigation finds Uber's 'ineffective safety culture' to blame for its self-driving car killing a pedestrian last year

https://bit.ly/3epKmdy

David Shepardson , Reuters Nov 20, 2019, 8:48 AM

National Transportation Safety Board (NTSB) investigators examine a self-driving Uber vehicle involved in a fatal accident in Tempe, Arizona, U.S., March 20, 2018. National Transportation Safety Board/Handout via REUTERS

Positive Trust Balance

Carnegie Mellon University

- Positive Trust Balance:
  - Stakeholders trust that lifecycle risk will be acceptable

Building Trust with Stakeholders

- Safety transparency
- Beyond testing to Positive Trust Balance:
  Engineering, Validation, Feedback, Culture
- Robust safety culture required to succeed

Carnegie Mellon University