The Big Picture for Self-Driving Car Safety

Safe Systems Summit, April 2019

Prof. Philip Koopman
Overview

- Conventional vehicle software safety
- Driver assistance safety
  - “ADAS”/“SOTIF”
- Fully autonomous safety
  - Perception
  - Operations
  - System life cycle
  - Societal & ethical issues

[General Motors]
Brute Force Road Testing

- Good for identifying “easy” cases
  - Expensive and potentially dangerous

Unrealistic to Brute Force Safety

- If 100M miles/critical mishap...
  - Test 3x–10x longer than mishap rate
  - Need 1 Billion miles of testing

- That’s ~25 round trips on every road in the world
  - With fewer than 10 critical mishaps

- Instead, use standards-based safety engineering
Conventional Vehicle Safety

- Human driver ultimately responsible for safety
  - "Functional Safety" – respond to equipment; avoid design faults
  - Human does the right thing for malfunctions ("Controllability")

ISO 26262:
- Functional Safety
- Equipment Faults
- Design Faults
- Controllability

Perfom turns, etc.
Vehicle motion
Example Automotive Software Defects

Small sampling of NHTSA recalls (i.e., confirmed bugs)
- 17V-713: Engine does not reduce power due to ESP software defect
- 17V-686 and MANY others: Airbags disabled
- 15V-569: Unexpected steering motion causes loss of control
- 15V-460 and others: Airbags deploy when they should not
- 15V-145: Unattended vehicle starts engine → carbon monoxide poisoning
- 14V-370: Turns off headlights when driving
- 14V-204: 1.5 seconds reverse while displaying Drive

Voluntary Recalls:
- 2018 hybrid engine stall at high speeds (https://bloom.bg/2y21T71)
- 2014 sudden unintended acceleration (https://goo.gl/R9zgL1)
Advanced Driver Assistance Systems (ADAS)

- **Lane keeping, forward collision avoidance**
  - Driver still responsible for safety, but these help out

- **Safety Of Intended Function (SOTIF)**
  - Malfunctions can occur even with defect-free equipment
    - Example: camera does not detect white vehicle on white background

**ISO 21448 (SOTIF):**
- Enumerate Scenarios
- Triggering Events
- Historically for ADAS

**ISO 26262:**
- Functional Safety
- Equipment Faults
- Design Faults
- Controllability

- Radar, Lidar, etc.
  - Don’t hit things
SOTIF Is All About the Scenarios

- Will the sensors detect objects?
  - Impaired by rain, snow, glare

- Is this something the car knows how to do?
  - Road geometries
  - Other vehicles & pedestrians
  - Traffic lights & road signs
  - Road surface conditions
  - Weather

- ISO 21448 SOTIF Strategy:
  - Giant scenario database
  - Push everything into “SAFE” bin
Closed Course Testing

- Safer than road testing, but expensive
  - Not scalable
  - Only tests things you have thought of!

Volvo / Motor Trend
Simulation

- Highly scalable; less expensive
  - Only tests things you have thought of!
  - Need higher coverage for full autonomy compared to ADAS

[Simulation Link]

Udacity

[Simulation Link]

Apollo
But Wait – There’s More!

Drivers do more than drive
Object Classification

ISO 21448 (SOTIF):
- Enumerate Scenarios
- Triggering Events
- Historically for ADAS

ISO 26262:
- Functional Safety
- Equipment Faults
- Design Faults
- Controllability

How safe is enough?
Maintenance, Updates
Perception Use of Machine Learning

Learn by example
- Training data instead of design
- E.g., 1 million pictures of people

Inductive training
- No requirements
- No design insight

Traditional safety includes validation against requirement
What About Edge Cases?

- Gaps in training data can lead to perception failure
  - Safety needs to know: “Is that a person?”
  - Machine learning provides: “Is that thing like the people in my training data?”

- Edge Case are surprises
  - You won’t see these in training or testing
    ➔ Edge cases are the stuff you didn’t think of!

PREDICTED CONCEPT | PROBABILITY
--- | ---
bird | 0.997
**no person** | **0.990**
one | 0.975
feather | 0.970
nature | 0.963
poultry | 0.954
outdoors | 0.936
color | 0.910
animal | 0.908

https://www.clarifai.com/demo
http://bit.ly/2In4rzj
Need An Edge Case “Zoo”

Novel objects (missing from zoo) are triggering events


https://goo.gl/J3SSyu
Stress Testing Perception

- Augmenting images with noise highlights perception issues
  - *Identifies systemic weaknesses even in absence of noise*

- False negative when in front of dark vehicle
- False negative when person next to light pole
- False positive on lane marking
  False negative real bicyclist
Brittle perception behavior indicates Edge Cases

- Data augmentation reveals triggering events

Edge Cases:
- False negatives
- Novel objects
- False positives
- Combinations
Mask-R CNN: examples of systemic problems

Baseline, un-augmented images. (Your mileage will vary.)
Drivers do more than just drive
- Occupant behavior, passenger safety
- Detecting and managing equipment faults

Operational limitations & situations
- System exits Operational Design Domain
- Vehicle fire or catastrophic failure
- Post-crash response

Interacting with non-drivers
- Pedestrians, passengers
- Police, emergency responders
Handling updates
- Fully recertify after every weekly update?
- Security in general

Vehicle maintenance
- Pre-flight checks, cleaning
- Corrective maintenance

Supply chain issues
- Quality fade
- Supply chain faults

Is windshield cleaning fluid life critical?
Safety Standard Landscape

IEEE P700x - Ethical Life-Cycle

ISO 21448 (SOTIF): - Enumerate Scenarios - Triggering Events - Historically for ADAS

ISO 26262: - Functional Safety - Equipment Faults - Design Faults - Controllability
Note: ISO 26262 covers functional safety lifecycle activities.

Note: Provides significantly more detailed treatment regarding autonomous systems for those topics also mentioned in ISO 21448, ISO 26262. Intended to supplement, not replace these.

© 2019 Philip Koopman
Where did “94%” come from?

- “The critical reason was assigned to drivers in an estimated 2,046,000 crashes that comprise 94 percent of the NMVCCS crashes at the national level. However, in none of these cases was the assignment intended to blame the driver for causing the crash.”

[DOT HS 812 115]

Practical improvement perhaps only 50%

It’s Not All Cell Phone Distraction

- 37,133 Fatalities
  - 63% Passenger vehicles
  - 14% Motorcycles
  - 19% Pedestrians, bikes, etc.
  - 2% Large trucks  (+ 2% other = 100%)

- One fatality per 86 million miles, including impaired drivers
  - 29% Alcohol Impairment
  - 27% Seat belt not used in passenger vehicle
  - 26% Speed (can be too fast for conditions)
  - 9% Distracted driving  (total of all sources was more than 100% due to overlap)

- The bar is set high for unimpaired humans
  - 99.999999%+ fatality-free miles
True self-driving cars will take a long time
- “Disengagement” is the wrong metric to use for road testing
- Start with very limited operational design domains (ODDs)
- “Self driving cars will be safer” is aspirational

Driver assistance (ADAS) offer low hanging fruit
- Safety-net functions such as AEB can help, even if imperfect
- BUT:
  - False activations can be a problem (e.g., false alarm panic braking)
  - Pretty good autonomy results in bad human supervision

If you blame the human operator for a crash, it’s not a self-driving car
Ways To Improve AV Safety

- **More safety transparency**
  - Independent safety assessments
  - Industry collaboration on safety

- **Minimum performance standards**
  - “Driver test” is necessary -- but not sufficient
    - How do you measure maturity?

- **Autonomy software safety standards**
  - ISO 26262/21448 + UL 4600 + IEEE P700x
  - Dealing with uncertainty and brittleness

Thanks!