Autonomous Vehicles: Safety Measures and Benchmarks for Perception & Cognition Functions*

AAAI 2021 FALL SYMPOSIUM SERIES
S3: Cognitive Systems for Anticipatory Thinking

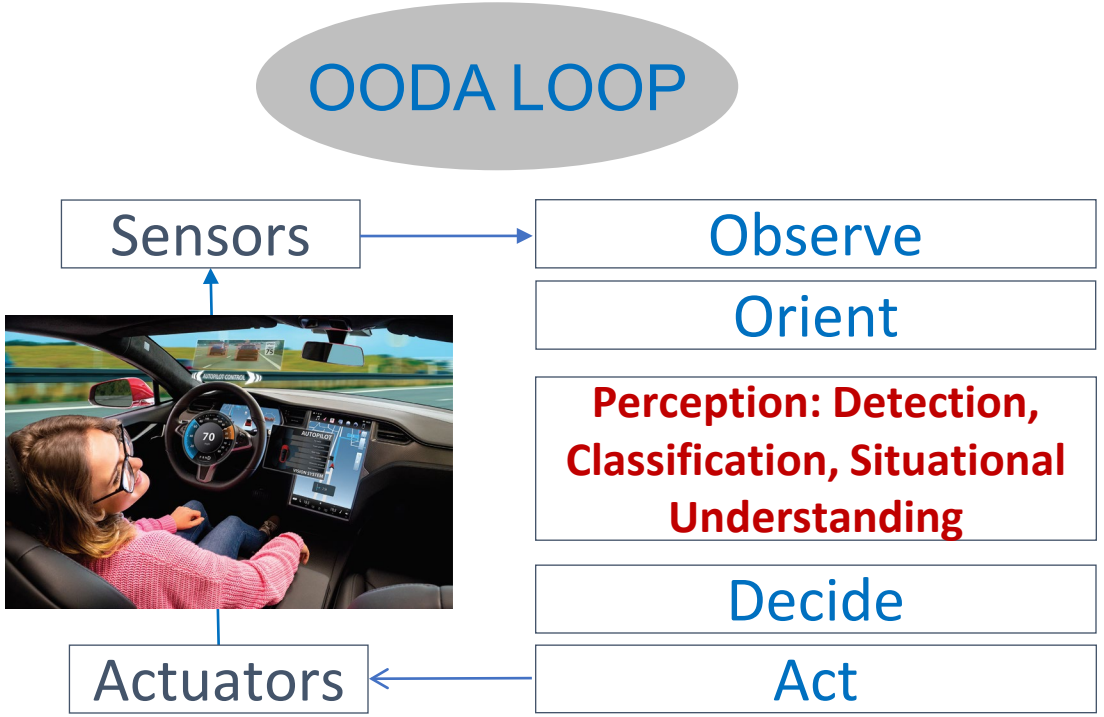
Dr. Jay Lala
Project Lead
Intelligent Vehicle Dependability & Security (IVDS)
https://www.dependability.org/wg10.4/ivds/index.html

6 November 2021

*Work performed under the auspices of the IVDS Project of the IFIP WG10.4 on Fault-Tolerant and Dependable Computing

Autonomous Vehicles: Role of Perception

- Sensors (Observe): Electro-Optical, Infrared, Radar, LIDAR, GPS, MEMS, Vehicle subsystems (Engine/Brakes/etc) performance, health & status sensors
- Algorithms (Orient & Decide): Catch-all for all the Feedback Control System Functions, incl. sensor processing and correlation, situational awareness, decision making, collision avoidance, etc.
- Actuators (Act): Commands to Engine, Brakes, Steering
- Processors: CPUs, GPUs, Software
- Communication: Links to other cars and Traffic Signaling Systems



Perception is key to situational awareness and safe operation:
How good does it have to be to equal a human driver's performance?

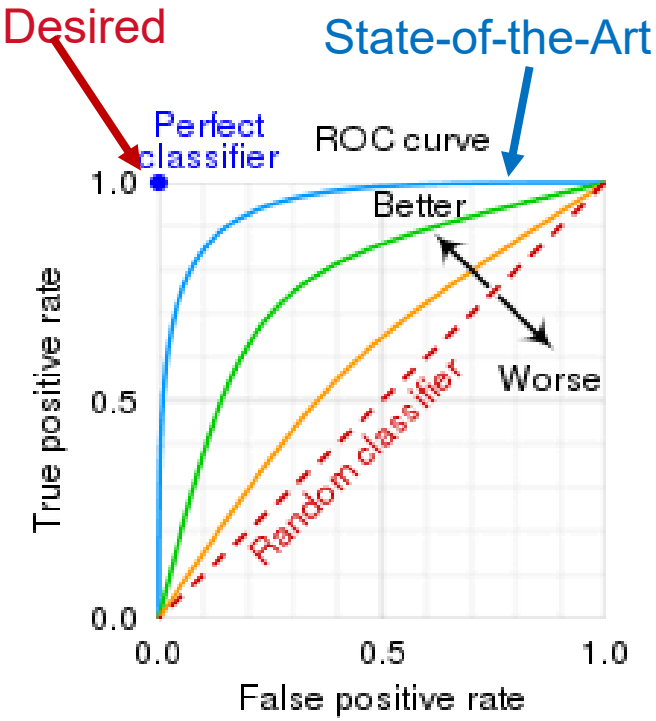
Perception in AVs: Risk Analysis

- Just like a human driver, AV Perception function is constantly making many safety-critical decisions
- Let's use a specific scenario as a representative:
  - Is this a plastic bag to drive thru or a human being?
- Undesired outcomes and consequences of Perception function:
  - False Positive (perception decides it's a human being when it's not)
    - AV brakes unnecessarily, creating a rough ride
  - False Negative (perception decides it's not a human being when it is)
    - AV does not stop, causing injury or death
- For most AI/ML algorithms, about 0.10 – 0.01% False Negative rate is considered excellent performance

What's an acceptable False Negative rate for AV Machine Perception that would be no riskier than a human driver ?

AV Machine Perception: Acceptable False Negative Rates to Equal Human Performance

- Current human-caused fatalities per hour of driving (2020 NHSTA):
  - $5.5X10^{-7}$ fatalities/hr
- False Negative rate must not exceed
  - $5.5X10^{-8}$ in ideal conditions: e.g., highway, daytime, good visibility
    - assumes 10 safety-critical decisions/hr (every 6 minutes)
  - $5.5X10^{-9}$ in uncertain environment: e.g., urban traffic, nighttime, snow/rain/fog
    - assumes 100 safety-critical decisions/hr (every 36 seconds)
- Perception function can err only about once out of 10 to 100 million critical decisions, to equal human performance!



This simple benchmark exercise answers Heilmeier question: What's the final "exam" to check for Perception success?

Potential Solutions

- Previous model is based on a single sensor
- AVs typically have multiple sensors with complementary sensing capabilities:
  - e.g., LIDAR, Radar, EO/Optical, etc are great for obstacle/object detection
  - Sensor fusion has been used successfully in many other disciplines and mathematically modeled
    - e.g., Inertial Measurement Unit and GPS
- Next step in perception, object classification (most reliant on AI/ML), is not as amenable to dissimilar redundancy
- However, some promising new research in Ensemble Learning from the Dependable and Fault-Tolerant community maybe a good starting point:
  - applying N-version programming to NN models used for autonomous steering control in AVs: https://homa-alem.github.io/papers/ISSRE2018_FA.pdf
  - N-Version Machine Learning Models for Safety Critical Systems https://ieeexplore.ieee.org/document/8806018
  - New Wine in an Old Bottle: N-Version Programming for Machine Learning Components | Karthik Pattabiraman (ubc.ca)

Getting machine perception to be as good as a human's is definitely a DARPA-hard problem

BACKUP

Assumptions

- Current human-caused fatalities per hour of driving (2020 NHSTA):
  - $5.5X10^{-7}$ fatalities/hr
  - We are assuming that all fatalities are caused by a failure of human perception
  - If only a fraction of these can be attributed to perception, the bar for the machine perception becomes even higher.
- False Negative rate must not exceed
  - $5.5X10^{-9}$ in uncertain environment: e.g., urban traffic, nighttime, snow/rain/fog
    - assumes 100 critical decisions/hr (every 36 seconds)
  - This rate is probably much lower than what a human driver does in white-knuckle driving. That is, the actual rate of critical decision making is prob every few seconds rather than a couple of time per minute.
- We define critical decisions as those having an impact on safe driving. Of course, not every decision is safety-critical. How fast to drive, when to change lanes, how closely to follow vehicles, etc are usually not impactful.