“AI” and Autonomous Vehicle Safety

Prof. Philip Koopman

ARTS 2023
July 10, 2023
Artificial Intelligence ("AI")
- Capability to which people attribute intelligence
- The best AI can simulate narrowly intelligent behavior

Machine Learning ("ML")
- A statistical technique to implement AI capabilities
- When people say "AI" they usually mean "ML"
- TRAINING: show the system lots and lots of data
- DEPLOYMENT: outputs are based on statistics

https://en.wikipedia.org/wiki/HAL_9000
Classification Via ML

1. “Train” on lots of data with labels
   - E.g.: {person, taxi}

2. Examine a new piece of input
   - E.g.: some image while driving

3. Which label is statistically closer?
   - Classify as either person or taxi

- Crucial points
  - Self-taught statistical correlations
  - Might train on unexpected features
  - Very confident when clueless

"96% TAXI"
Generative AI

- Synthesize something statistically plausible
- Example 1: photos
  - “Deer at side of road standing still”
- Example 2: chat

I see a deer standing still at the side of the road. Will the deer run in front of my car? Yes/No answers only.
ML Advantages for AVs

- Train based on examples
  - *Old school*: mathematical description of “a person”
  - *Old school*: physics equations of motion
  - *ML*: train on millions of pictures of people
  - *ML*: train on millions of traffic data sets

- Simpler, scalable development
  - Collecting data seen as easier than writing code

- Impressive effectiveness
  - Might get 90% - 99% accuracy...
    ... often *much* better than previous methods
  - Viable technology for many perception tasks

ML Challenges for AVs

- Does not “understand” in the deep sense
  - Correlative rather than causal connections
- Vulnerable to surprises
  - Struggles when detecting something unexpected
  - Often falsely confident when it is just guessing
  - Can miss small clues that flip interpretation
- Safety is engineering process, not just testing
  - Good ML is 99%; Safety is 99.99999999%+
  - Testing does not prove safety.
  - Testing validates good safety engineering
    - How do we validate engineering of an ML-based system?
Safety Questions To Ask:

- What exactly do you mean by “safe”?
  - How can we measure your safety outcomes?

- How safe is your un-crewed vehicle right now?
  - Need 100M+ miles if based only on road experience

- Do you follow industry-written safety standards?
  - ISO 26262, ISO 21448, ANSI/UL 4600, AVSC guidelines
  - Which do you actually conform to? (Not just cherry picked some ideas)

- Do you believe that safety requires transparency?
  - Are your NHTSA crash reports 100% transparent?

https://on.gei.co/2r2rjzg