



Phil Koopman

Solving Edge Cases with Remote Assistance

PhilKoopman.Substack.com
AVsafety.Substack.com

PHILIP KOOPMAN

EMBODIED AI SAFETY

Reimagining
Safety Engineering for
Artificial Intelligence
in Physical Systems

RISK

**MACHINE LEARNING
SAFETY ENGINEERING
CYBERSECURITY
HUMAN INTERACTION**

TRUST

CHALLENGES

Edge Cases Solved – with Remote Assist

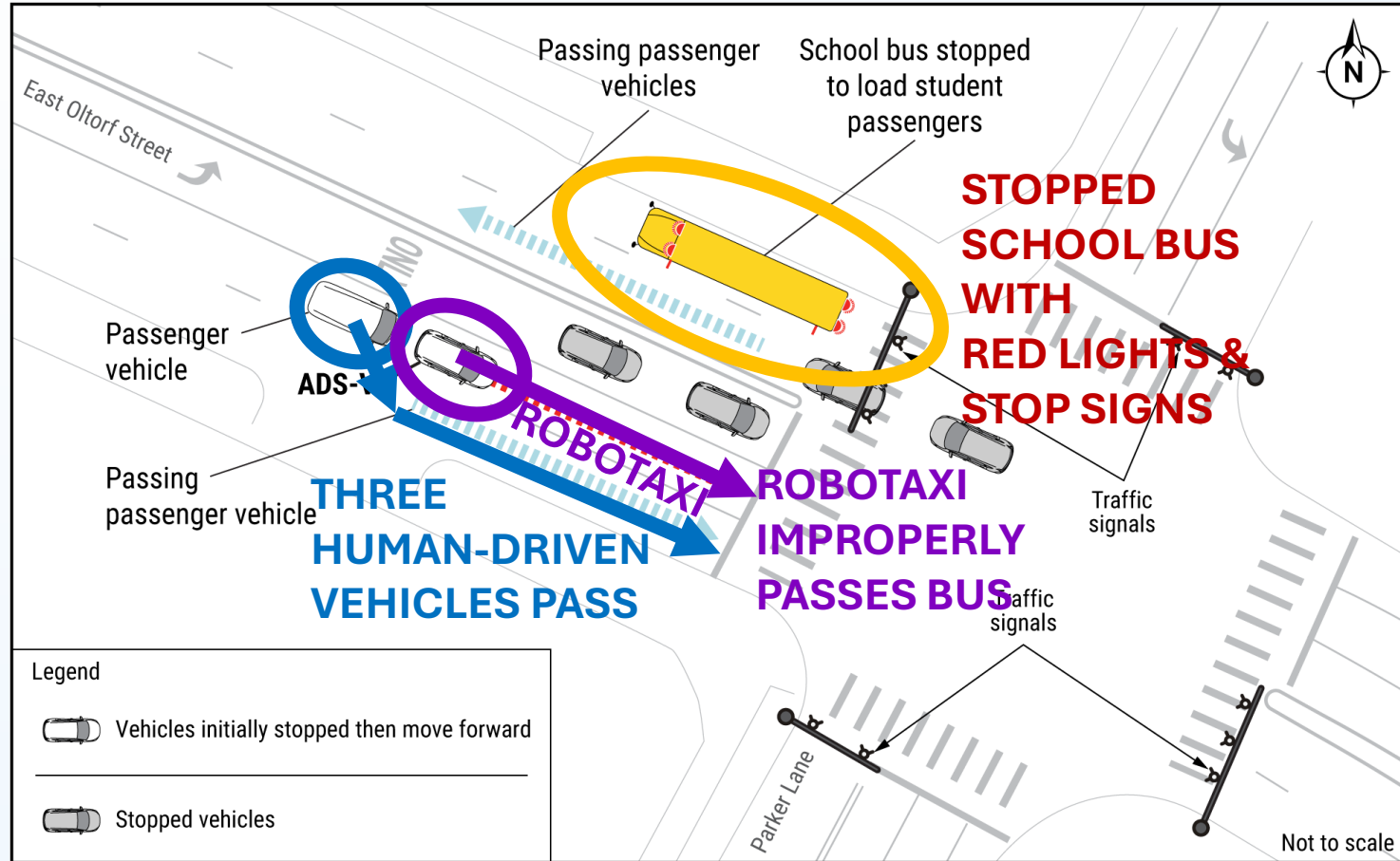
- **Edge cases control AI safety outcomes**
 - Too many to train on them all
 - Edge cases might not be what you think
- **Impracticable to solve them all**
- **Remote assistants to the rescue!**
 - Many different roles...
 - ... all of which have safety implications



Remote Assistant: Driving Past School Bus

- Jan 12, 2026
- Austin TX

[NTSB]
HWY26FH007



Remote Assistant Triggered Incident

- **“Is this a school bus with active signals?”**
 - Michigan Remote Assistant (RA) said “no”
 - Robotaxi proceeded past school bus
- **Significant controversy**
 - Confusion about whether RA is “driving”
 - Does RA invalidate autonomy claim?
- ***Reveal:* RAs are the plan for edge cases**





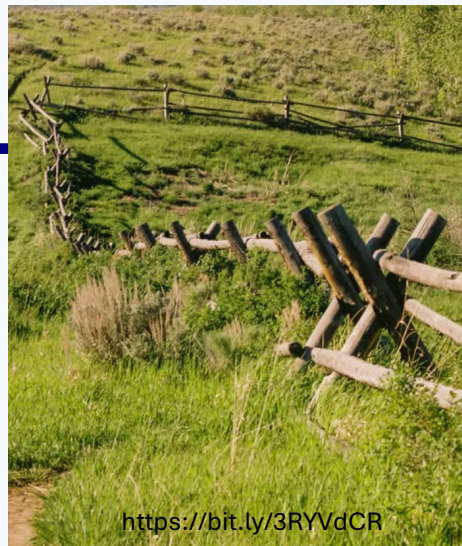
CAUTION: REMOTE ASSISTANT AT WORK

BUT ... Not the whole story!

Wizard of Oz

What's an Edge Case?

- **Historically:**
 - An extreme value
 - A boundary condition
- **For embodied/physical AI:**
 - Something insufficiently addressed in training
- **What you think is rare or extreme – is irrelevant**
 - If the AI is not trained on it, it's an edge case
- **Edge case failures → embarrassing incidents**



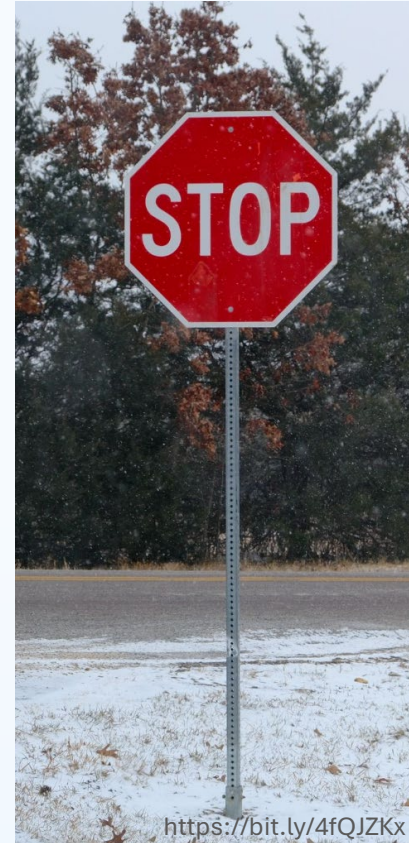
More than Just Strange Objects

- **Novel objects or situations**
 - Elephant on a US city street
 - Post-earthquake road damage
- **Novel attributes of common objects**
 - People dressed in yellow
 - Standing pedestrian traveling 20 mph
- **Context that changes required action**
- **Too many to catch them all... especially contexts**



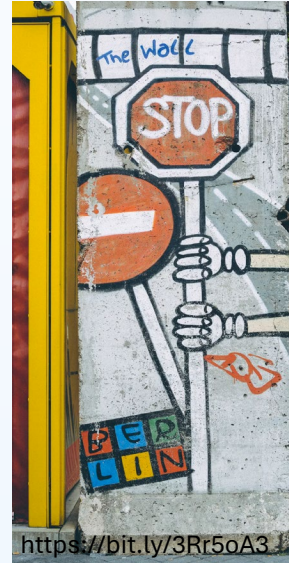
Context and Stop Signs

- **What do you do at a stop sign?**
 - Then what do you do?
- **Does that rule always apply?**



Stop Sign On A Stick

- What do you do in this situation?
- What about these situations?



School Buses & Stop Signs

- What if stop sign is on a school bus?
- What about this stop sign on this school bus?



CONTEXT MATTERS!

Remote Assistants → Human Judgment

- **RAs offer advice when asked by AV**
 - Some RAs are not classical “drivers”
 - AV companies insist RAs don’t do safety
- **“Phone a friend” analogy breaks down**
 - A person asking for help is accountable
 - But a machine cannot be accountable
- **“But, the computer rejects bad advice!”**
 - Apparently not for school buses



Accountability Laundering

- **“RAs do not have a safety role”**
 - If RA affects behavior, they are involved
- **What about hiding behind:**
“the AI is responsible”?
- **Imagine if:**
 - Surgeon + partially automated robot
 - Something goes wrong
 - Dangerous input from remote operator
 - Robot arm AI is said to be “responsible”
 - Who or what is accountable?

<https://bit.ly/4d5lAxB>



All Roles Have Safety Responsibilities

- **Operate:** Normal safety
- **Supervise:** Intervene to ensure safety
- **Decision:** Cast a vote in determining system behavior
- **Response:** OK to move vehicle after a collision?
- **Standby:** Unimpaired, available
- **Maintenance:** Correct problems and certify safe to operate



REMOTE OPERATION
REMOTE SUPERVISION
REMOTE DECISION
REMOTE RESPONSE
REMOTE STANDBY
REMOTE MAINTENANCE

Steps For Safe RA Handling of Edge Cases

- RAs provide human expertise
- Important steps for this to work
 1. AI detects an edge case and de-risks
 2. AI calls for help (reliable communications)
 3. Skilled RA available per triage process
 4. RA gains situational awareness
 5. RA has authority to take action
 - Or escalate to someone who can take action
- Embarrassing incidents tend to be edge case failures



RA Lessons from Robotaxi Operations

- **Staffing size**
 - How are mass disruptions handled?
- **Operator location**
 - Can police give a sobriety test?
 - Is extradition needed for accountability?
- **Operator licensing**
 - Would operator be qualified if local?
- **Control lag**
 - How fast depends on concept of operations
 - Safety must be assured even if comms blackout



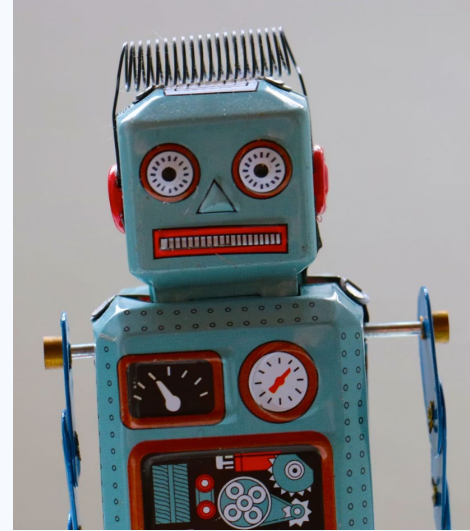
Implications for Physical/Embodied AI

- **Pure-AI challenges:**

- Human intervention seen as a technology failure
- Impracticable to train on enough edge cases for real-world operations

- **Remote assistant challenges:**

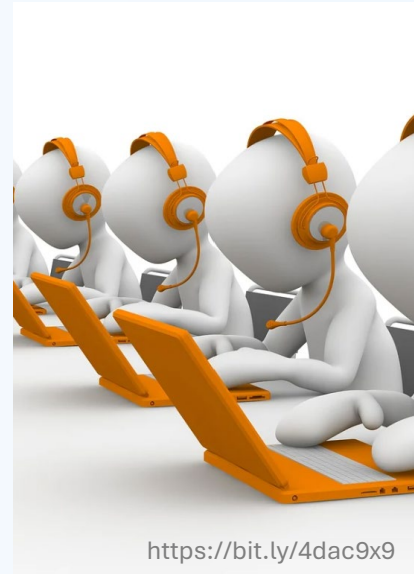
- Human intervention is normalized
- Detecting high-consequence edge cases
- Ensuring safety while waiting for remote help
- Proper staffing & human interface for remote assistants



<https://bit.ly/4wcp4GS>

Where Does This End Up?

- **Remote assistants are here to stay**
 - They make edge case judgment calls
- **Incentives to do the wrong thing:**
 - Pressure to go cheap with call centers
 - Force product liability law to limit legal exposure
 - Temptation to launder accountability via AI
- **If a human operator would be accountable...**
 - ... the human assistant (and employer) should be accountable**
 - ... but only to the degree they avoidably contributed to harm**



<https://bit.ly/4dac9x9>

Questions?

PhilKoopman.Substack.com
AVsafety.Substack.com