Safety in Freely-Composed Cyber-Physical Systems
Challenges and Opportunities

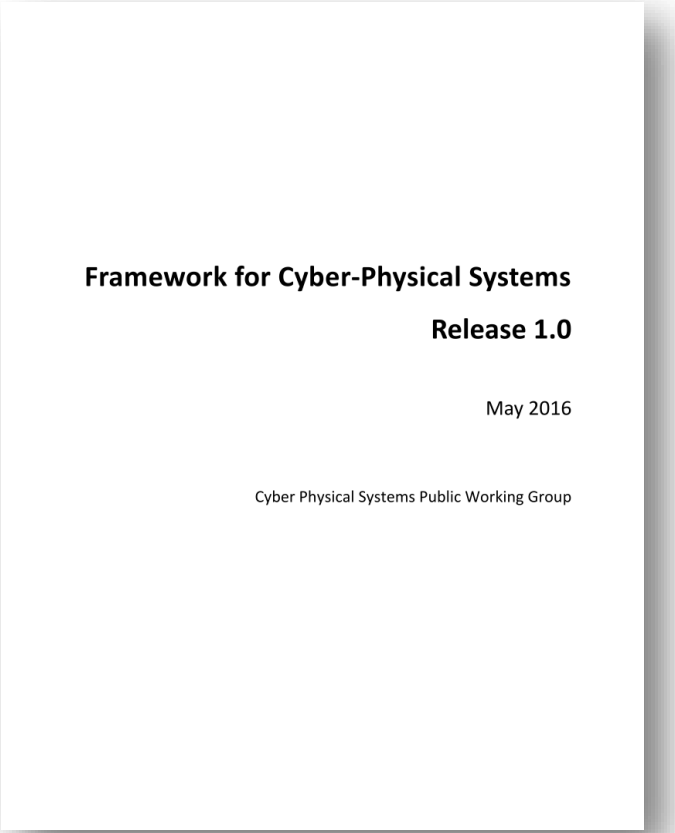
Pieter J. Mosterman and **Akshay Rajhans**
MathWorks Advanced Research & Technology Office (MARTO)

NIST Workshop on Exploring Dimensions of Trustworthiness: Challenges and Opportunities
National Institute of Standards and Technology (NIST), August 30-31, 2016

Facets in the system engineering process

Facets. Facets are views on CPS encompassing identified responsibilities in the system engineering process. They contain well-defined activities and artifacts (outputs) for addressing concerns. There are three identified facets:

- The conceptualization facet captures activities related to the high-level goals, functional requirements, and organization of CPS as they pertain to what a CPS or its components should be and what they are supposed to do. It provides as its overarching output a conceptual model of the CPS.
- The realization facet captures the activities surrounding the detailed engineering design, production, implementation, and operation of the desired systems. These activities include tradeoff analyses, detailed engineering design and simulation(s), and more, that drive towards and are responsible for realization of a CPS.
- The assurance facet deals with obtaining confidence that the CPS built in the realization facet satisfies the model developed in the conceptualization facet. Its activities include evaluating the claims, argumentation, and evidence as required to address important (and sometimes mandatory) requirements of design, policy, law, and regulation.



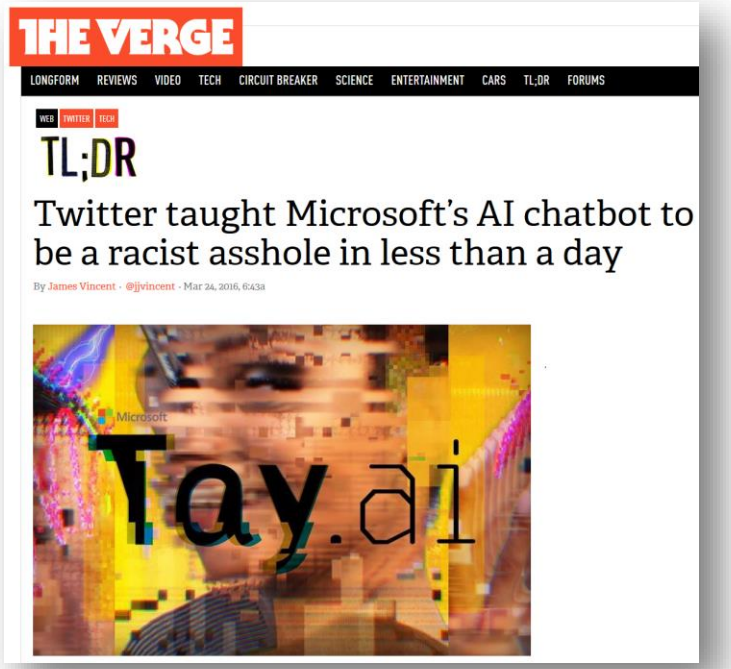
Individual | **Adaptive**

Adaptive

Conceptualize
- How to limit learning to safe behavior?

Realize
- What sensory system has sufficient richness?
  - How to prevent over interpretation?
- Robustness against interpretation edge case?
- Correctly fuse sensor data that is misaligned in time and space?

Assure
- How to test a self-changing artifact?
  - If regimes are not pre enumerated?
- Ensure successful and correct online calibration?



Unfortunately, in the first 24 hours of coming online, a coordinated attack by a subset of people exploited a vulnerability in Tay. Although we had prepared for many types of abuses of the system, we had made a critical oversight for this specific attack.

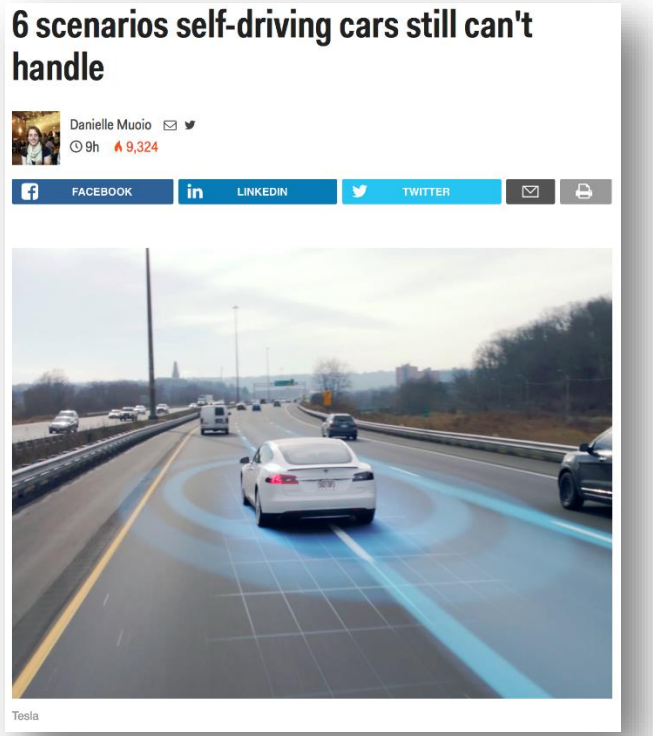
http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction

Adaptive

Conceptualize
- How to limit learning to safe behavior?

Realize
- What sensory system has sufficient richness?
  - How to prevent over interpretation?
- Robustness against interpretation edge case?
- Correctly fuse sensor data that is misaligned in time and space?

Assure
- How to test a self-changing artifact?
  - If regimes are not pre enumerated?
- Ensure successful and correct online calibration?



1. Driverless cars struggle going over bridges
   Because bridges don't have many environmental cues like surrounding buildings, it's hard for the Uber car to figure out where it is. GPS helps the car position itself, but not to the accuracy Uber wants.

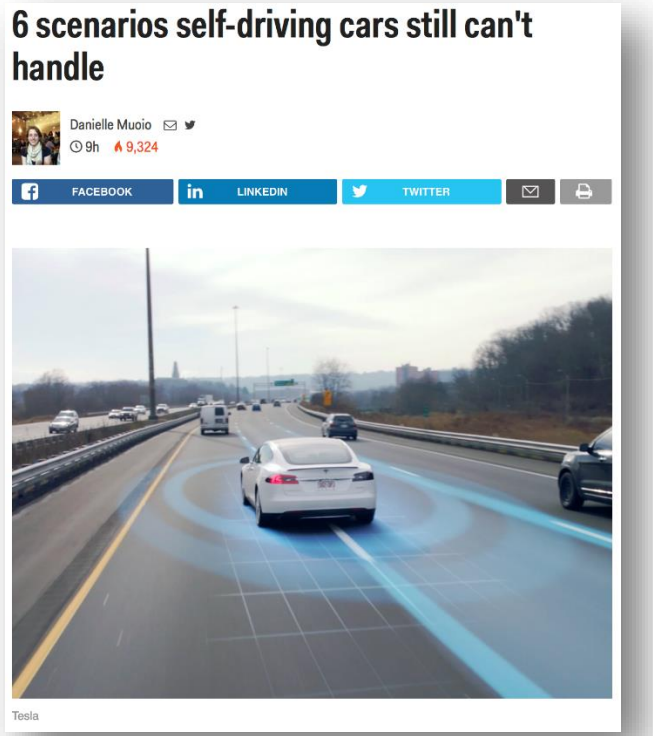
http://www.businessinsider.com/autonomous-car-limitations-2016-8

Adaptive

Conceptualize
- How to limit learning to safe behavior?

Realize
- What sensory system has sufficient richness?
  - How to prevent over interpretation?
- Robustness against interpretation edge case?
- Correctly fuse sensor data that is misaligned in time and space?

Assure
- How to test a self-changing artifact?
  - If regimes are not pre enumerated?
- Ensure successful and correct online calibration?



2. Cars struggle in inclement weather

   "Heavy snow and rain tend to confuse LiDAR sensors and also cameras," John Dolan, principle systems scientist at Carnegie Mellon's Robotics Institute, told Business Insider. "So you end up having some problems."

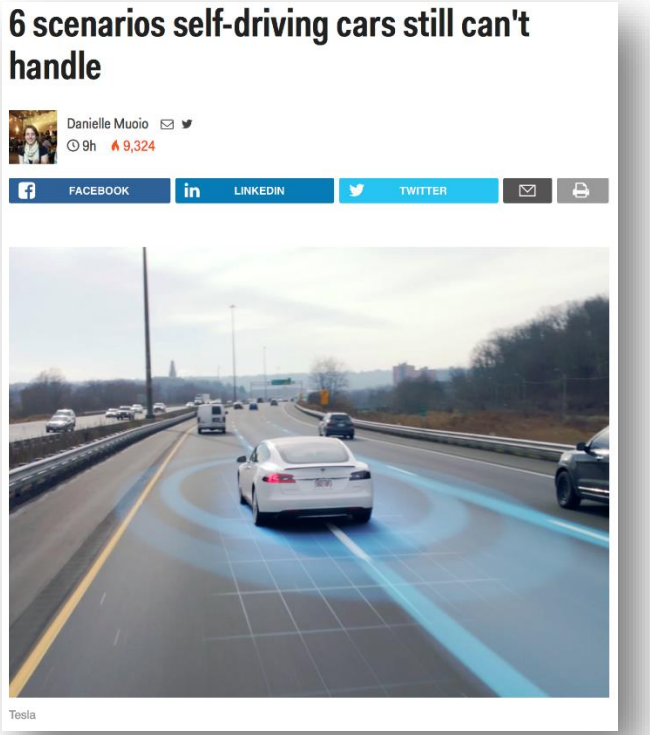
http://www.businessinsider.com/autonomous-car-limitations-2016-8

Adaptive

Conceptualize
- How to limit learning to safe behavior?

Realize
- What sensory system has sufficient richness?
  - How to prevent over interpretation?
- Robustness against interpretation edge case?
- Correctly fuse sensor data that is misaligned in time and space?

Assure
- How to test a self-changing artifact?
  - If regimes are not pre enumerated?
- Ensure successful and correct online calibration?



3. Cars struggle without clear lane markings
   When driverless cars can't distinguish the lanes, it makes it nearly impossible for them to drive or change lanes safely.

http://www.businessinsider.com/autonomous-car-limitations-2016-8

Adaptive

Conceptualize
- How to limit learning to safe behavior?

Realize
- What sensory system has sufficient richness?
  - How to prevent over interpretation?
- Robustness against interpretation edge case?
- Correctly fuse sensor data that is misaligned in time and space?

Assure
- How to test a self-changing artifact?
  - If regimes are not pre enumerated?
- Ensure successful and correct online calibration?

http://www.metatube.com/en/videos/150541/Cruel-parents-Oncoming-Truck-Prank/

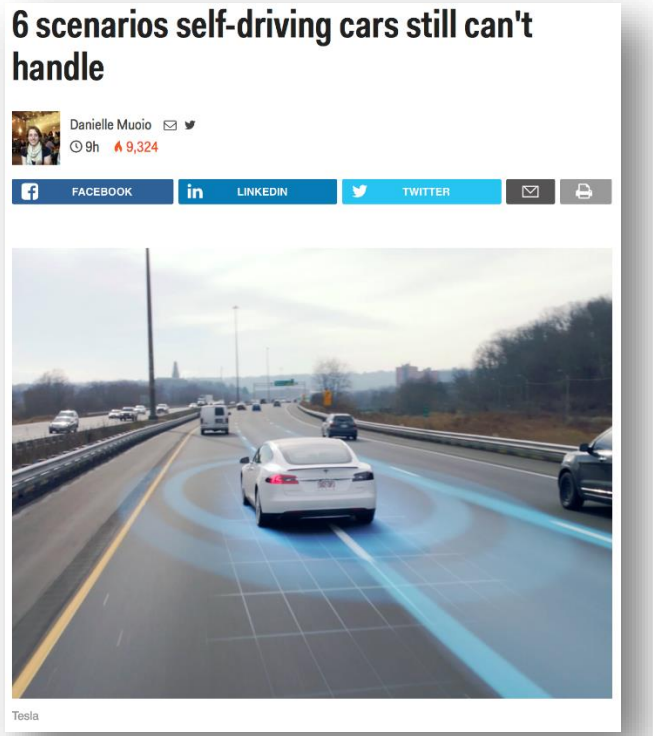
Individual | **Autonomous**

Autonomous

Conceptualize
- Sufficiently predictive environment models?
- Safe but nontrivial interaction with humans?
  - What are safe level of aggressiveness?

Realize
- Robust operation in an exceedingly complex environment?
- Fail safely with loss of minimum information?
  - Know a planned action is safe?
- Assess risk online?

Assure
- Degraded safety (there is no perfect safety)?
- Ensure reasoning is always safe?
- Turing test for cars?



4. City driving is much harder than highway
   Cities are a mess of pedestrians, cars, potholes, traffic cones — you get the point.

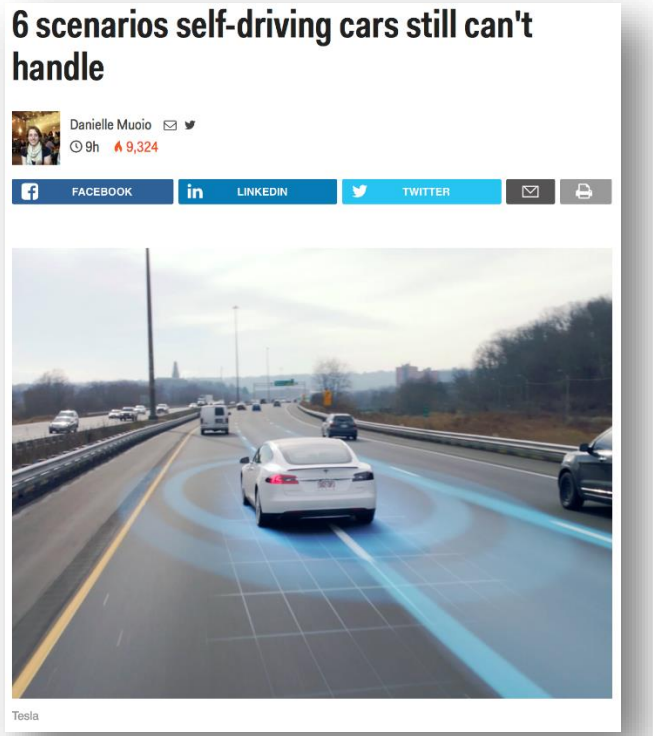
   If you're trying to do urban driving and depending on GPS to a large extent, then when you get into areas where there are a lot of tall buildings it's hard to receive the GPS signal and you'll have drop outs
   http://www.businessinsider.com/autonomous-car-limitations-2016-8

Autonomous

Conceptualize
- Sufficiently predictive environment models?
- Safe but nontrivial interaction with humans?
  - What are safe level of aggressiveness?

Realize
- Robust operation in an exceedingly complex environment?
- Fail safely with loss of minimum information?
  - Know a planned action is safe?
- Assess risk online?

Assure
- Degraded safety (there is no perfect safety)?
- Ensure reasoning is always safe?
- Turing test for cars?



5. Robot cars can't interact like humans can
   … more times than not, we rely on waving to let someone know it's ok to go. Driverless cars don't have that luxury

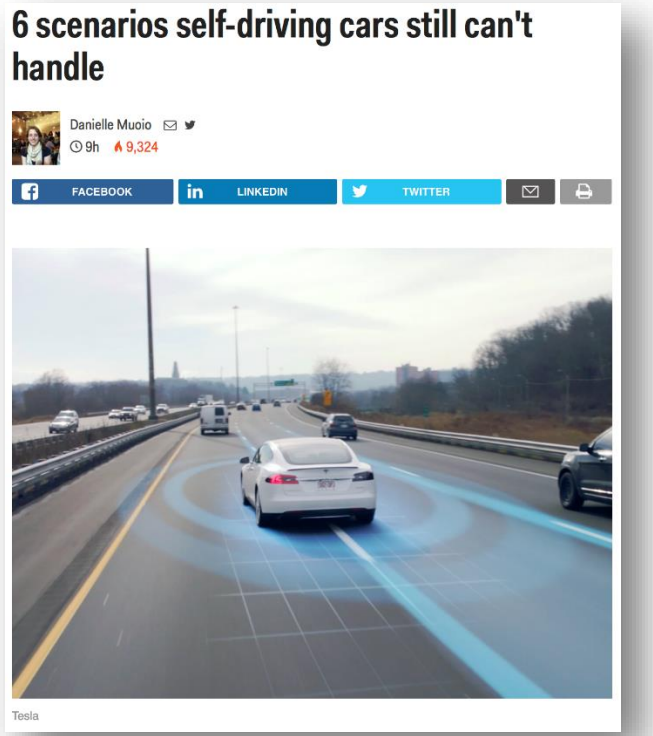
http://www.businessinsider.com/autonomous-car-limitations-2016-8

Autonomous

Conceptualize
- Sufficiently predictive environment models?
- Safe but nontrivial interaction with humans?
  - What are safe level of aggressiveness?

Realize
- Robust operation in an exceedingly complex environment?
- Fail safely with loss of minimum information?
  - Know a planned action is safe?
- Assess risk online?

Assure
- Degraded safety (there is no perfect safety)?
- Ensure reasoning is always safe?
- Turing test for cars?



6. High-speed situations may be trouble

   … when human drivers try to merge onto roads with cars traveling at higher speeds, they tend to inch forward to make sure it's ok.

   But a driverless car probably wouldn't take that risk

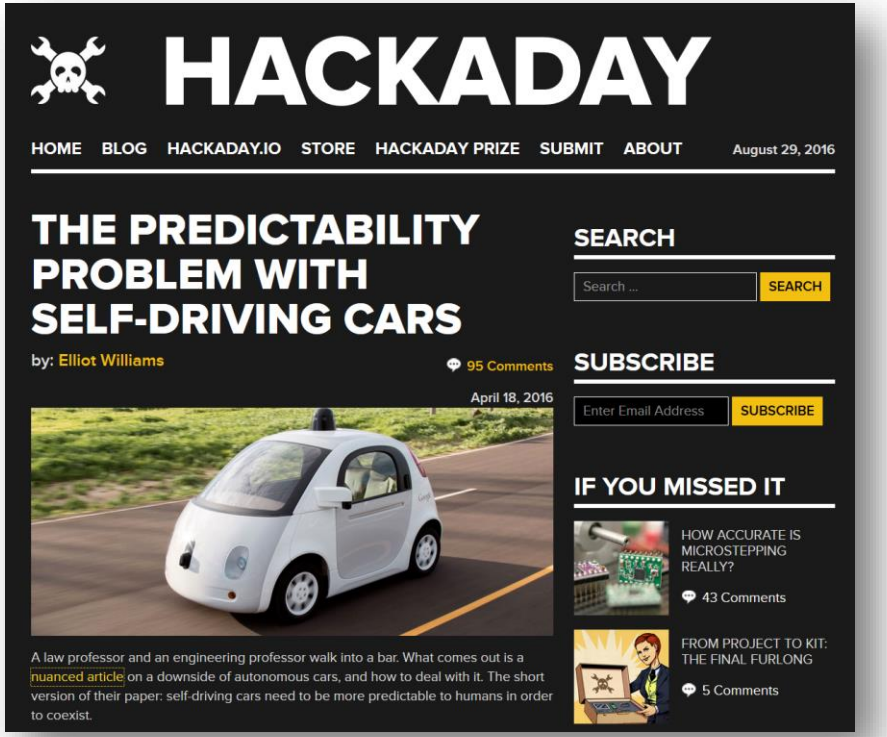
http://www.businessinsider.com/autonomous-car-limitations-2016-8

Autonomous

Conceptualize
- Sufficiently predictive environment models?
- Safe but nontrivial interaction with humans?
  - What are safe level of aggressiveness?

Realize
- Robust operation in an exceedingly complex environment?
- Fail safely with loss of minimum information?
  - Know a planned action is safe?
- Assess risk online?

Assure
- Degraded safety (there is no perfect safety)?
- Ensure reasoning is always safe?
- Turing test for cars?



When we want to know what another car is going to do, we think about the driver of the car, [..] We then think about what we'd do in their place,

If people can't read your car's AI's mind, you're gonna get your fender bent.

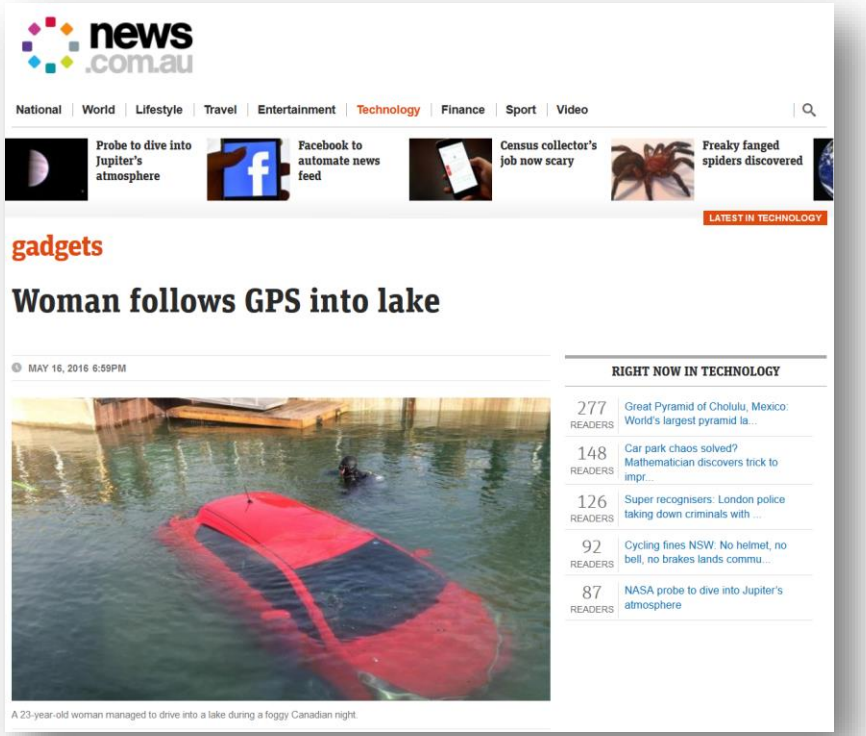
Ensemble | Connected

Connected

Conceptualize
- How to interpret data safely
  - Which data to corroborate information?

Realize
- Safely operate in the face of communication challenges
  - Degradation, loss
  - Corruption
- Timeliness and responsiveness guarantees?
  - Denial of Service (DoS), service discovery time out

Assure
- How do you obtain failure probabilities?
- Is closed loop verification possible?



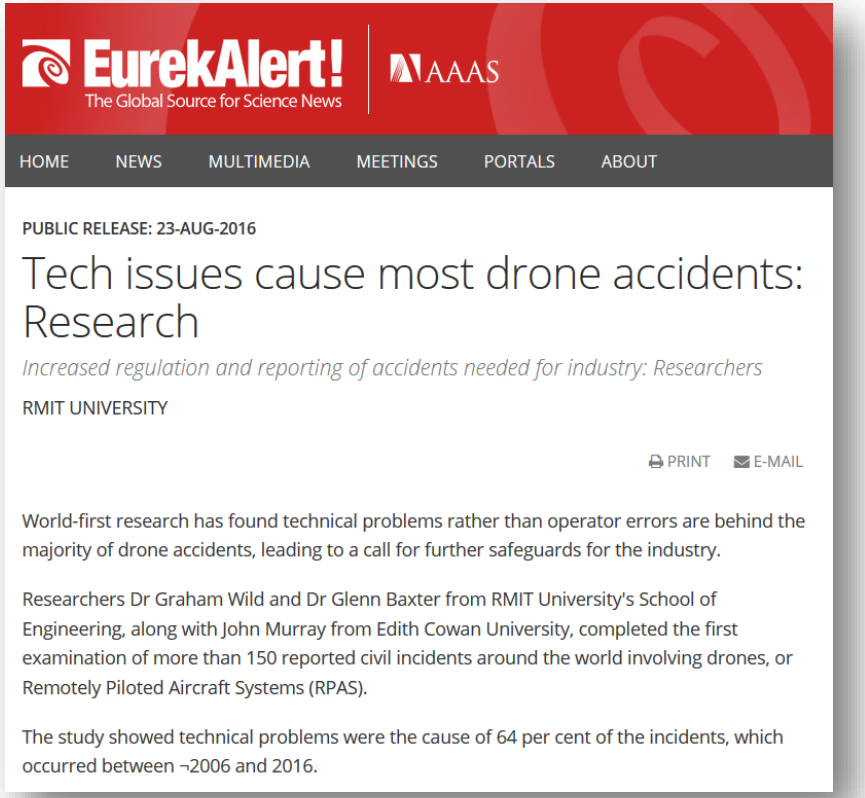
the woman was following a route on her car's GPS while driving in the dark on a foggy night in Ontario when it directed her to drive onto a boat launch, and she ended up in a lake.

Connected

Conceptualize
- How to interpret data safely
  - Which data to corroborate information?

Realize
- Safely operate in the face of communication challenges
  - Degradation, loss
  - Corruption
- Timeliness and responsiveness guarantees?
  - Denial of Service (DoS), service discovery time out

Assure
- How do you obtain failure probabilities?
- Is closed loop verification possible?


EurekAlert! The Global Source for Science News | AAAS

HOME NEWS MULTIMEDIA MEETINGS PORTALS ABOUT

PUBLIC RELEASE: 23-AUG-2016

Tech issues cause most drone accidents: Research

Increased regulation and reporting of accidents needed for industry: Researchers

RMIT UNIVERSITY

PRINT E-MAIL

World-first research has found technical problems rather than operator errors are behind the majority of drone accidents, leading to a call for further safeguards for the industry.

Researchers Dr Graham Wild and Dr Glenn Baxter from RMIT University's School of Engineering, along with John Murray from Edith Cowan University, completed the first examination of more than 150 reported civil incidents around the world involving drones, or Remotely Piloted Aircraft Systems (RPAS).

The study showed technical problems were the cause of 64 per cent of the incidents, which occurred between ¬2006 and 2016.

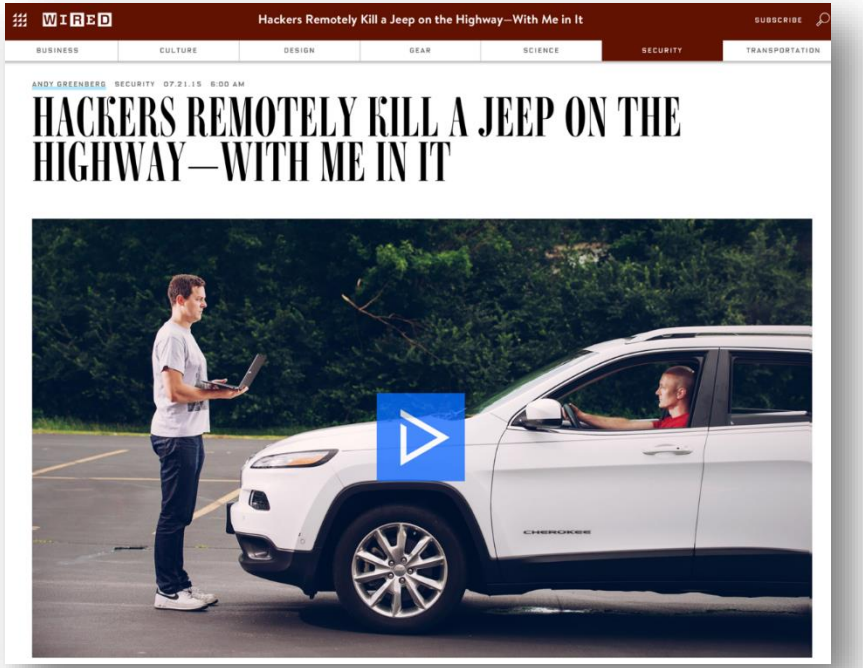
Recently published in the journal Aerospace, the study found that in most cases, broken communications links between the pilot and the Remotely Piloted Aircraft Systems (RPAS) were the cause of the incident,

Connected

Conceptualize
- How to interpret data safely
  - Which data to corroborate information?

Realize
- Safely operate in the face of communication challenges
  - Degradation, loss
  - Corruption
- Timeliness and responsiveness guarantees?
  - Denial of Service (DoS), service discovery time out

Assure
- How do you obtain failure probabilities?
- Is closed loop verification possible?



Uconnect's cellular connection also lets anyone who knows the car's IP address gain access from anywhere in the country.

From that entry point, the attack pivots to an adjacent chip [..] rewriting the firmware [..] capable of sending commands through the CAN bus, to its physical components like the engine and wheels.

Ensemble | **Collaborative**

Collaborative

Conceptualize
- Cross-organization failure effect analysis?
- How to identify and prevent race conditions?
- Robust conflict resolution across an ensemble?
- How to trade off system vs. ensemble safety?

Realize
- How to perform online safety analysis?
  - Safety of ad hoc rules in collaboration?
  - How to gracefully enter/exit a collaboration?
- How to assign risk to a collaboration?
- How to ensure ample resources to be safe?
- Can you assign probability to reliance?

Assure
- How do you test? Measure coverage?
- Work outside nominal regions (online derating)?
- Assumptions about collaborating systems?

Swiss regional air traffic chief Anton Maag said both aircraft were diving to avoid a crash when they flew into each other.

And he added that the Russian pilot had started a steep dive only after controllers had repeatedly instructed him to do so.

Collaborative

Conceptualize
- Cross-organization failure effect analysis?
- How to identify and prevent race conditions?
- Robust conflict resolution across an ensemble?
- How to trade off system vs. ensemble safety?

Realize
- How to perform online safety analysis?
  - Safety of ad hoc rules in collaboration?
  - How to gracefully enter/exit a collaboration?
- How to assign risk to a collaboration?
- How to ensure ample resources to be safe?
- Can you assign probability to reliance?

Assure
- How do you test? Measure coverage?
- Work outside nominal regions (online derating)?
- Assumptions about collaborating systems?


SECTIONS HOME SEARCH The New York Times LOG IN

BOYDTON JOURNAL
Cloud Computing Brings Sprawling Centers, but Few Jobs, to Small Towns

FEATURE
Inside Facebook's (Totally Insane, Unintentionally Gigantic, Hyperpartisan)...

Cycling Matches the Pace and Pitches of Tech

Adopted Koreans, Stymied in Search of Birth Parents, Find Hope in a Cotton Swab

TECHNOLOGY

TECHNOLOGY

Google's Driverless Cars Run Into Problem: Cars With Drivers

By MATT RICHTEL and CONOR DOUGHERTY SEPT. 1, 2015

855

A Google self-driving car in Mountain View, Calif. Google cars regularly take the most cautious approach, but that can put them out of step with the other vehicles on the road.
Gordon De Los Santos/Google

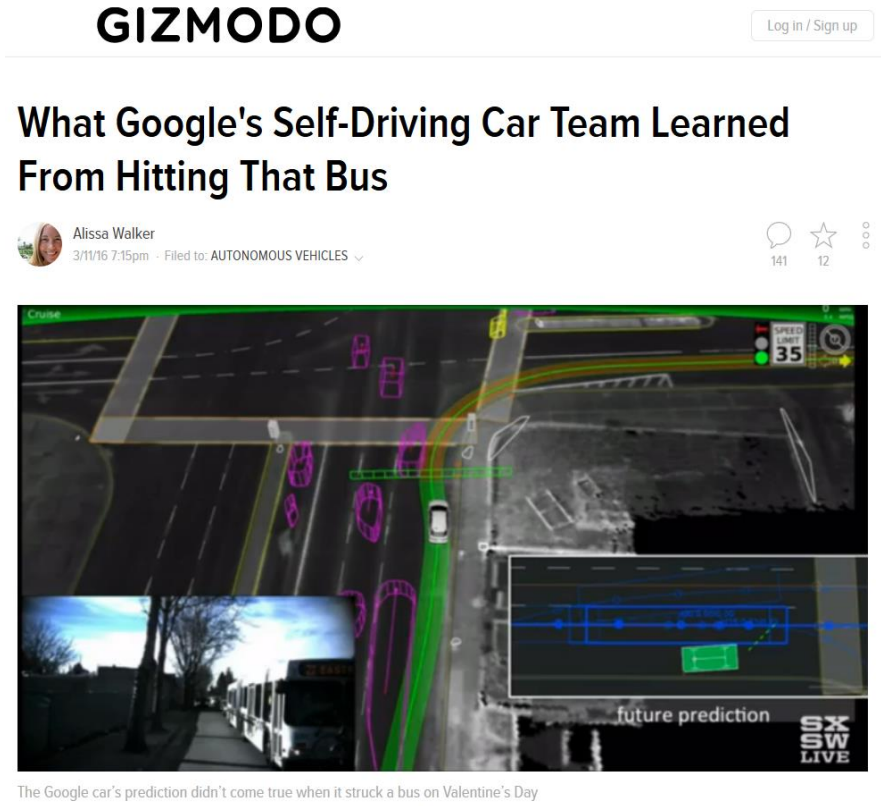
The way humans often deal with these situations is that "they make eye contact. On the fly, they make agreements about who has the right of way," said John Lee, a professor of industrial and systems engineering and expert in driver safety and automation at the University of Wisconsin.

Collaborative

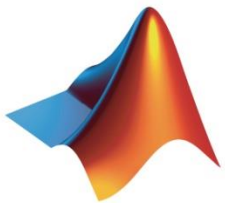
Conceptualize
- Cross-organization failure effect analysis?
- How to identify and prevent race conditions?
- Robust conflict resolution across an ensemble?
- How to trade off system vs. ensemble safety?

Realize
- How to perform online safety analysis?
  - Safety of ad hoc rules in collaboration?
  - How to gracefully enter/exit a collaboration?
- How to assign risk to a collaboration?
- How to ensure ample resources to be safe?
- Can you assign probability to reliance?

Assure
- How do you test? Measure coverage?
- Work outside nominal regions (online derating)?
- Assumptions about collaborating systems?



"Our car was making an assumption about what the other car was going to do," said Chris Urmson



MathWorks®

Accelerating the pace of engineering and science