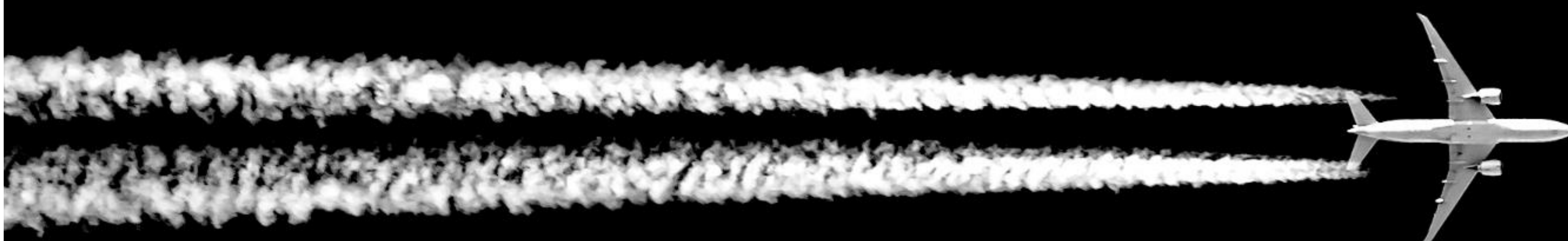
SO YOU WANT TO PUT A NEURAL NETWORK ON AN AIRPLANE...



DR. DARREN COFER / PRINCIPAL FELLOW

NRC REGULATORY INFORMATION CONFERENCE
THE FUTURE OF NUCLEAR: ADAPTING TO AI-ENABLED AUTONOMY
13 MARCH 2024

Collins Aerospace | APPLIED RESEARCH & TECHNOLOGY

© 2024 Collins Aerospace
This document does not include any export controlled technical data.

BUT WHY?

- Increasing demand for commercial air travel, cargo/supply chain, service to high-density urban areas
- Shortage of trained pilots for the foreseeable future
- Proposals for reduced crew, single pilot operations… full autonomy?
- Requires automation of tasks currently done by pilots (some safety-critical)
- *Many will be implemented using machine learning and AI technologies*

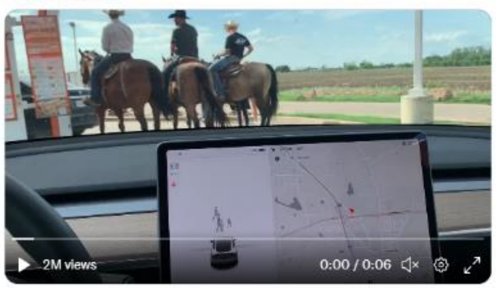
- There are also many use cases for AI/ML that don’t have anything to do with autonomy

UNANTICIPATED INPUTS / UNINTENDED BEHAVIORS

Tesla car was on Autopilot when it hit a Culver City firetruck, NTSB finds

https://www.latimes.com/business/story/2019-09-03/tesla-was-on-autopilot-when-it-hit-culver-city-fire-truck-ntsb-finds

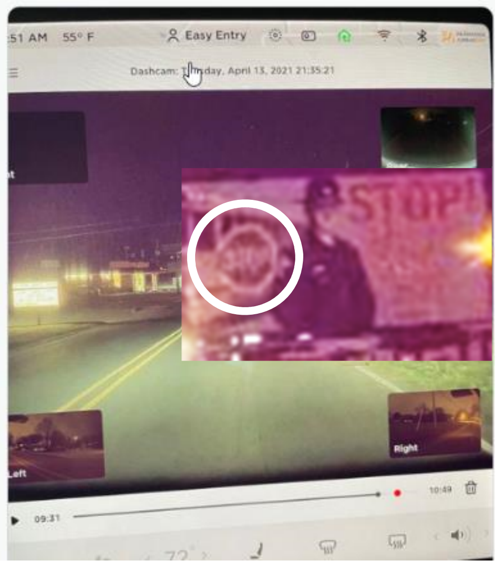
Could we get the Tesla software to better recognize horses in the drive thru..at Whataburger? This is Texas you know. jk. @Tesla @elonmusk @Teslarati @JaneidyEve



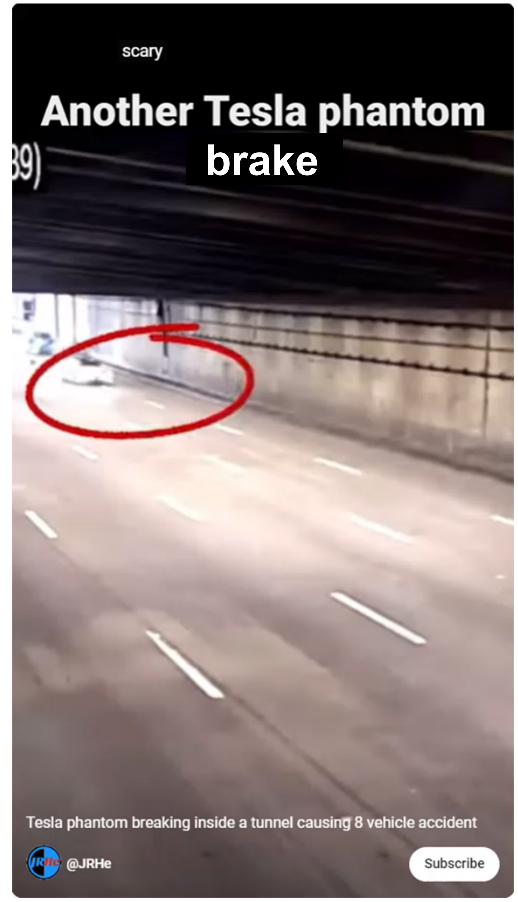
https://twitter.com/Cbiggs90/status/1519821916189741057

Andy Weedman
@andyweedman

@karpathy @elonmusk @DirtyTesla here is a fun edge case. My car kept slamming on the brakes in this area with no stop sign. After a few drives I noticed the billboard.



https://twitter.com/andyweedman/status/1382459653863378944



This document does not include any export controlled technical data.

CERTIFICATION CHALLENGES FOR AI/ML

14 CFR Part 25 Airworthiness for Transport Category Airplanes

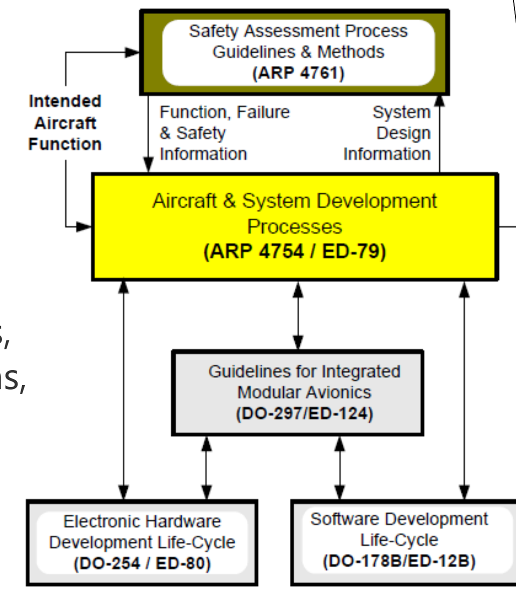
§ 25.1309 Equipment, systems, and installations.

(a) The equipment, systems, and installations whose functioning is required by this subchapter, must be designed to ensure that they *perform their intended functions under any foreseeable operating condition*.

(b) The airplane systems and associated components, considered separately and in relation to other systems, must be designed so that—

(1) The occurrence of any failure condition which would prevent the continued safe flight and landing of the airplane is *extremely improbable*

GUIDANCE STANDARDS : MEANS OF COMPLIANCE



ARP4754A FIGURE 1

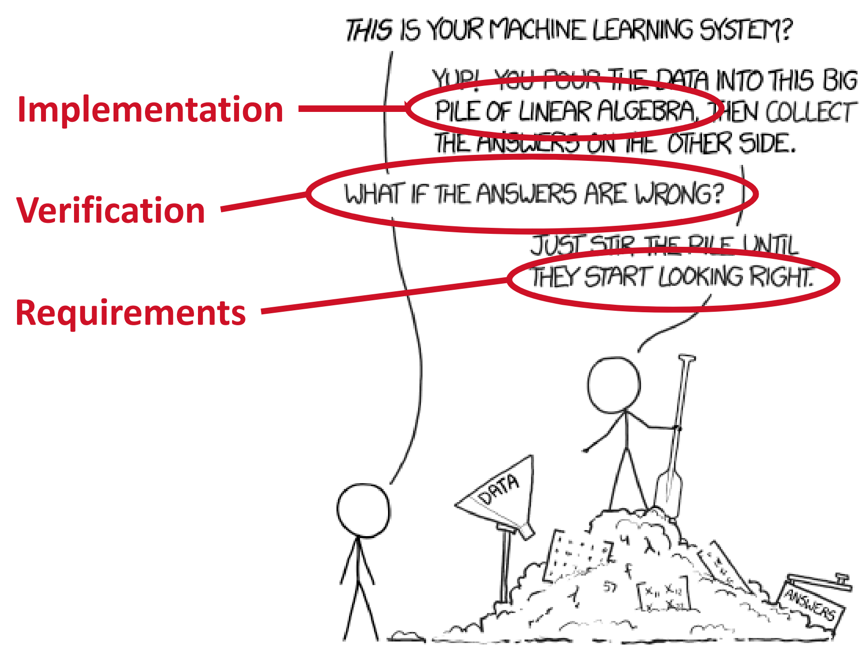
ARTIFICIAL INTELLIGENCE
IN AERONAUTICAL SYSTEMS
STATEMENT OF CONCERNS

1. Aircraft/system satisfies requirements 2. No unintended behavior

This document does not include any export controlled technical data.

ASSURANCE CHALLENGES FOR AI/ML

- Requirements
  - Large training/test datasets define the desired behaviors
- Verification
  - Current assurance methods rely on requirements-based test cases with structural coverage metrics to assess completeness
  - Structural coverage metrics are unable to detect **unintended behavior** in neural networks
- Implementation
  - Design intent cannot be inferred from examination of design artifacts
  - Design elements and implementation code are **not traceable** to requirements
  - AI/ML development languages, libraries, and tools incompatible with certification processes



https://xkcd.com/1838/

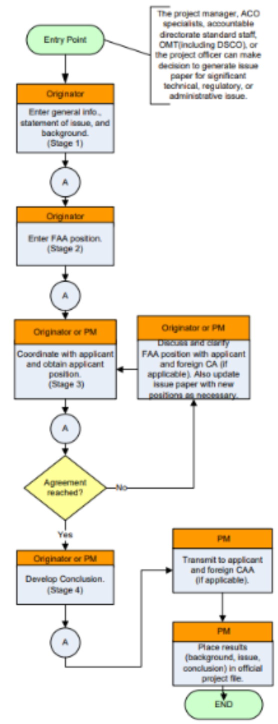
How can we demonstrate absence of unintended behavior?

This document does not include any export controlled technical data.

AUTHORITIES ARE ENGAGING IN DIFFERENT WAYS

Anticipating many applications of ML technologies in aviation, EASA has produced a roadmap, concept papers, and studies addressing the associated assurance issues

ARP6983 : Process Standard for Development and Certification / Approval of Aeronautical Safety-Related Products Implementing AI



AC 20-166A Fig. 1

FAA has determined that ML technologies are "new and novel" so approval must be obtained using the Issue Paper process in accordance with AC 20-166A (can't use existing guidance as *method of compliance*)

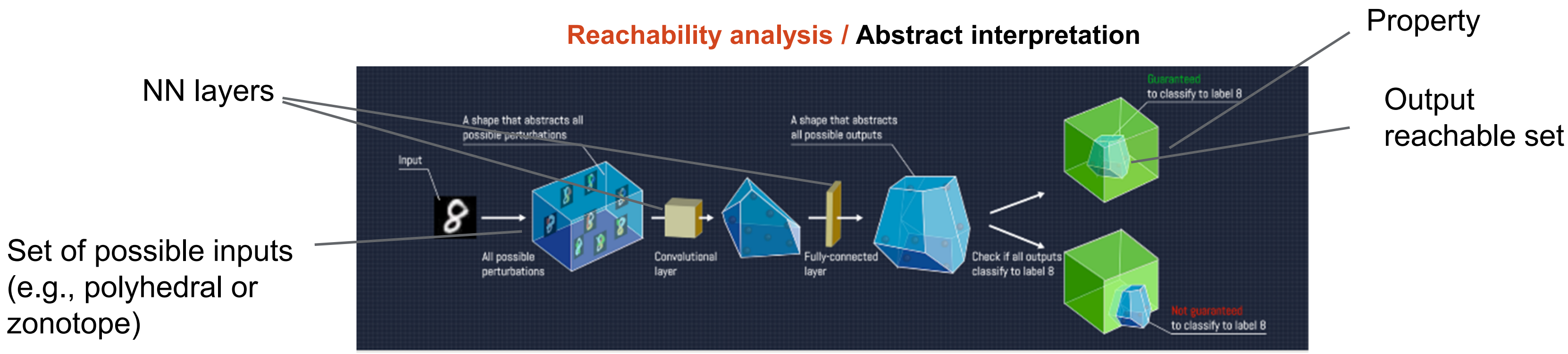
New and Novel technologies require new methods of compliance

This document does not include any export controlled technical data.

FORMAL METHODS TOOLS FOR NN

COMPREHENSIVE ANALYSIS TO DETECT/ELIMINATE UNINTENDED BEHAVIOR

Reachability analysis / Abstract interpretation



Source: SRI Lab, ETH Zurich

EXACT METHOD

- Calculates a separate output reachable set for each distinct combination of internal NN activations. Total reachable set is a union.
- **Lots of computation and memory required – may not scale**
- Provides formal guarantees both on satisfaction and on violation of the property (SAT or UNSAT + counterexample)
- Tools: NNV, Marabou

APPROXIMATE METHOD

- Calculates a single output reachable set. The output set of each NN layer is over-approximated
- **Significantly reduces computation overhead**
- **May return UNKNOWN answer**
- Counterexamples may be spurious
- Tools: NNV, ERAN

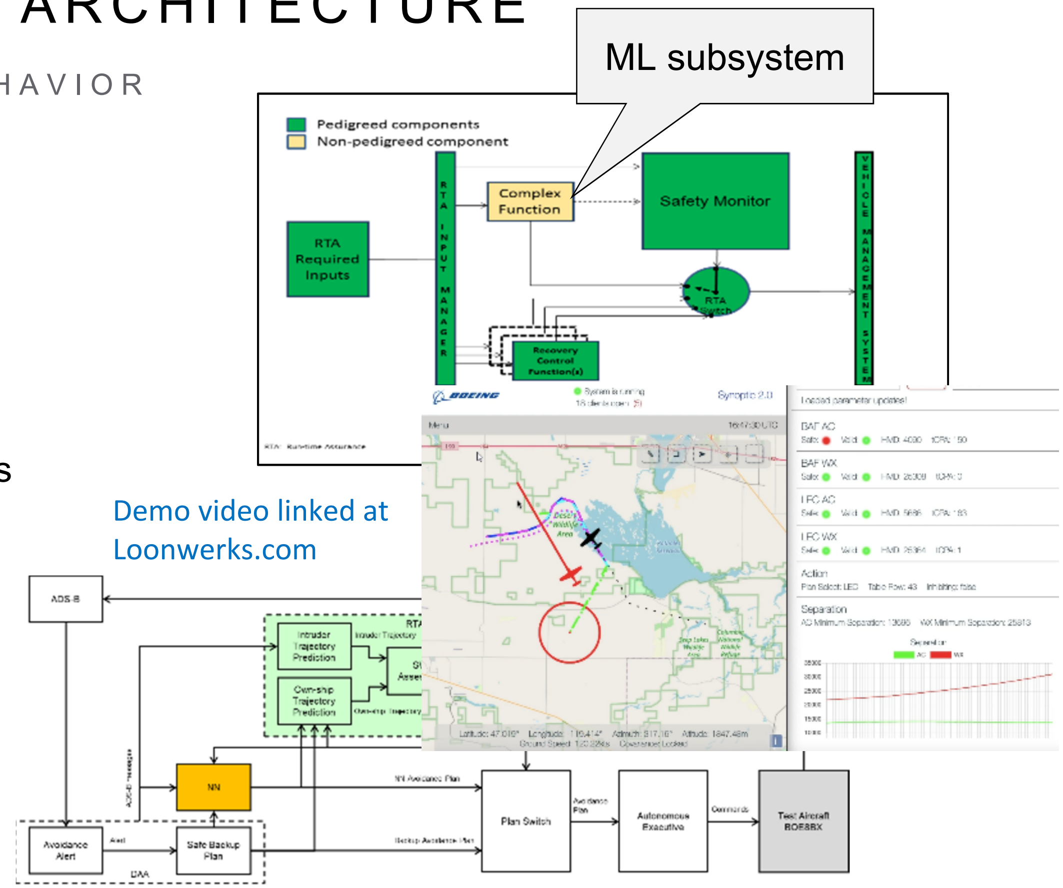
This document does not include any export controlled technical data.

RUN-TIME ASSURANCE ARCHITECTURE

BOUND RESIDUAL UNINTENDED BEHAVIOR

- Run-time monitors detect unsafe/unexpected behaviors
- May monitor inputs, outputs, system state
- Switch to alternative safe behavior or recovery function, possibly with limited performance
- ML subsystem may still contain **unintended behavior**, but architecture ensures that there is no impact on system safety

- ASTM F3269
  - *Standard Practice for Methods to Safely Bound Behavior of Aircraft Systems Containing Complex Functions Using Run-Time Assurance*



Demo video linked at Loonwerks.com

This document does not include any export controlled technical data.

NO *ONE SIZE FITS ALL* SOLUTION

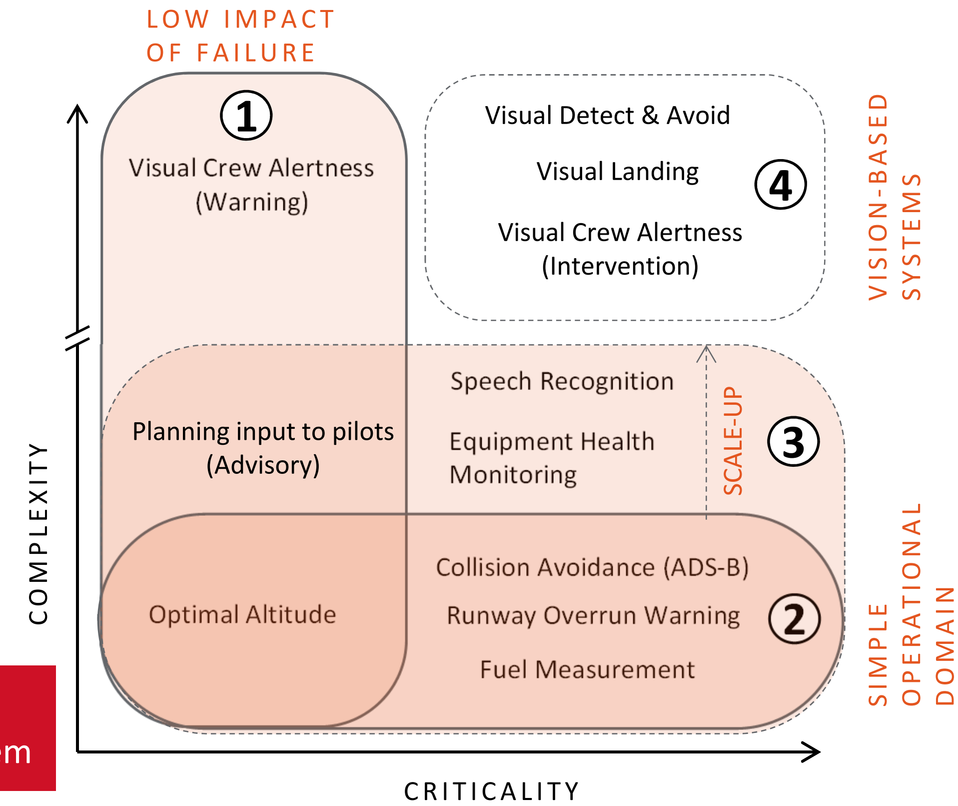
1. **Low-criticality systems** : *low impact of failure*
   Machine learning components treated as “black box” similar to traditional system certification, demonstrate compliance with requirements through testing

2. **Low-complexity systems** : *smaller models, simple operational domain, potentially with higher criticality*
   New analysis tools and advanced testing methods developed on DARPA Assured Autonomy program

3. **Increasing complexity** : *larger models and operational domains*
   Scalability improvements in analysis tools for larger systems

4. **Complex high-criticality systems** : *vision-based systems in safety-critical applications*
   Additional research on scenario-based completeness metrics and increased scalability

New methods of compliance to address gaps will vary depending on complexity and criticality of target system



This document does not include any export controlled technical data.

SUMMARY

- AI/ML will be used to meet demand for increasingly autonomous aircraft
- AI/ML faces both technical challenges and certification barriers (unknown unknowns)
- New certification guidance is being developed for AI/ML
- New assurance technologies are being developed for AI/ML
- No one size fits all solution

Unintended Behavior in Learning-Enabled Systems: Detecting the Unknown Unknowns
D. Cofer. Digital Avionics Systems Conference (DASC), October 2021

Flight Test of a Collision Avoidance Neural Network with Run-Time Assurance
D. Cofer, R. Sattigeri, I. Amundson, J. Babar, S. Hasan, E. Smith, K. Nukala, D. Osipychev, M. Moser, J. Paunicka, D. Margineantu, L. Timmerman, J. Stringfield.
Digital Avionics Systems Conference (DASC), September 2022.

Formal Methods use for Learning Assurance (ForMuLA)
M. Baleani, A. Clavière, D. Cofer, E. DeWind, L. Di Guglielmo, O. Ferrante, G. Gentile, D. Kirov, D. Larsen, L. Mangeruca, S. Fulvio Rollini, G. Cima, R. Schneider H. Semde, G. Soudain
Collins Aerospace / EASA IPC Report

Loonwerks.com

This document does not include any export controlled technical data.