UNIVERSITY *of York*

ASSURING AUTONOMY INTERNATIONAL PROGRAMME

Assuring Autonomy - developing a Body of Knowledge

Prof. John McDermid OBE FREng

Agenda

Key Topics

- Autonomy and Artificial Intelligence (AI)
- Assurance challenges and (some) new approaches
    - Current approaches to safety and SOTIF
    - Complex and open environment
    - Human-AV interaction
    - Complex system and AI lifecycles
    - Assurance case
- Regulation
- Related work
- Conclusions and discussion

Autonomy

Concept

- The Oxford English Dictionary says that autonomy is
  - The ability of a person to make his or her own decisions (or self-government, independence …)
- Autonomous *systems* make decisions, not humans
  - Electric kettles that switch themselves off
  - Adaptive gearboxes in cars
  - Vacuum cleaners …
- Autonomous not automatic
  - Open world, complexity of decisions, uncertainty, …

Autonomy and AI

AI, Machine Learning and Assurance

- Autonomous systems do not need to use AI
  - But many do, especially Machine Learning (ML)
- ML means
  - Getting computers to learn from data in the form of observations and real-world interactions in order to create a **model** of the real-world
- ML implements decisions moved from human to machine
  - Machine only does what its been trained to do (model)
  - No ‘general intelligence’ brought to bear (not in model)

Example HAV Function

Lane Keeping

Current Approaches

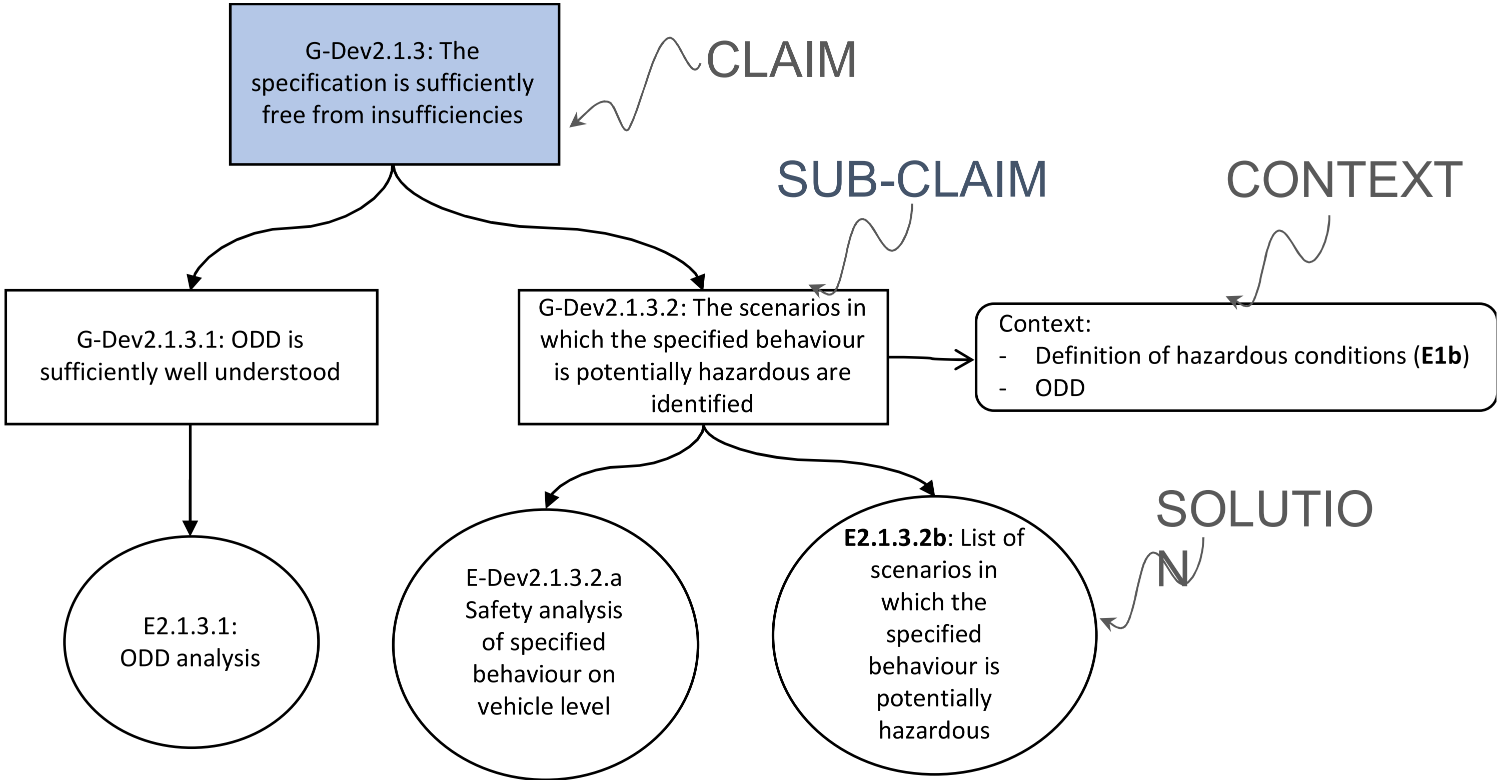
ISO 26262 and SOTIF

- ISO 26262 was developed by the industry
    - Derived from IEC 61508
    - Documenting ‘accepted good practice’ – drawing on other domains in some cases
- Some deliberate ‘omissions’ now being addressed
    - ISO 26262 only considers effect of ‘failures’
        - Doesn’t address HAVs or FAVs
    - Safety of Intended Function (SOTIF) in ISO PAS 21448
        - Considering is function, e.g. lane keeping, safe ‘in itself’?
        - Step towards dealing with autonomy (mainly HAVs)

Arguing Safety

ISO 26262 and SOTIF Assurance Case

- Argument supported by evidence



G-Dev2.1.3: The specification is sufficiently free from insufficiencies

CLAIM

SUB-CLAIM

CONTEXT

G-Dev2.1.3.1: ODD is sufficiently well understood

G-Dev2.1.3.2: The scenarios in which the specified behaviour is potentially hazardous are identified

Context:
- Definition of hazardous conditions (**E1b**)
- ODD

SOLUTIO
N

E2.1.3.1: ODD analysis

E-Dev2.1.3.2.a Safety analysis of specified behaviour on vehicle level

E2.1.3.2b: List of scenarios in which the specified behaviour is potentially hazardous

Why is Assurance Hard?

Complexity of the Environment

- Can we test to gain assurance?
    - Currently ~ 3.7 million miles between fatalities with drivers in developed nations (statistics vary)
    - Even the best current AVs (Waymo) “disconnect” about every 16,500 miles (worst around 1 mile)
    - Have we covered all credible operational scenarios?
    - On road-testing can never be enough
- Must using testing and simulation
    - Combined coverage of operational scenarios
    - Arguments needed to justify ‘sufficiency’

Why is Assurance Hard?

Open World

New Approach

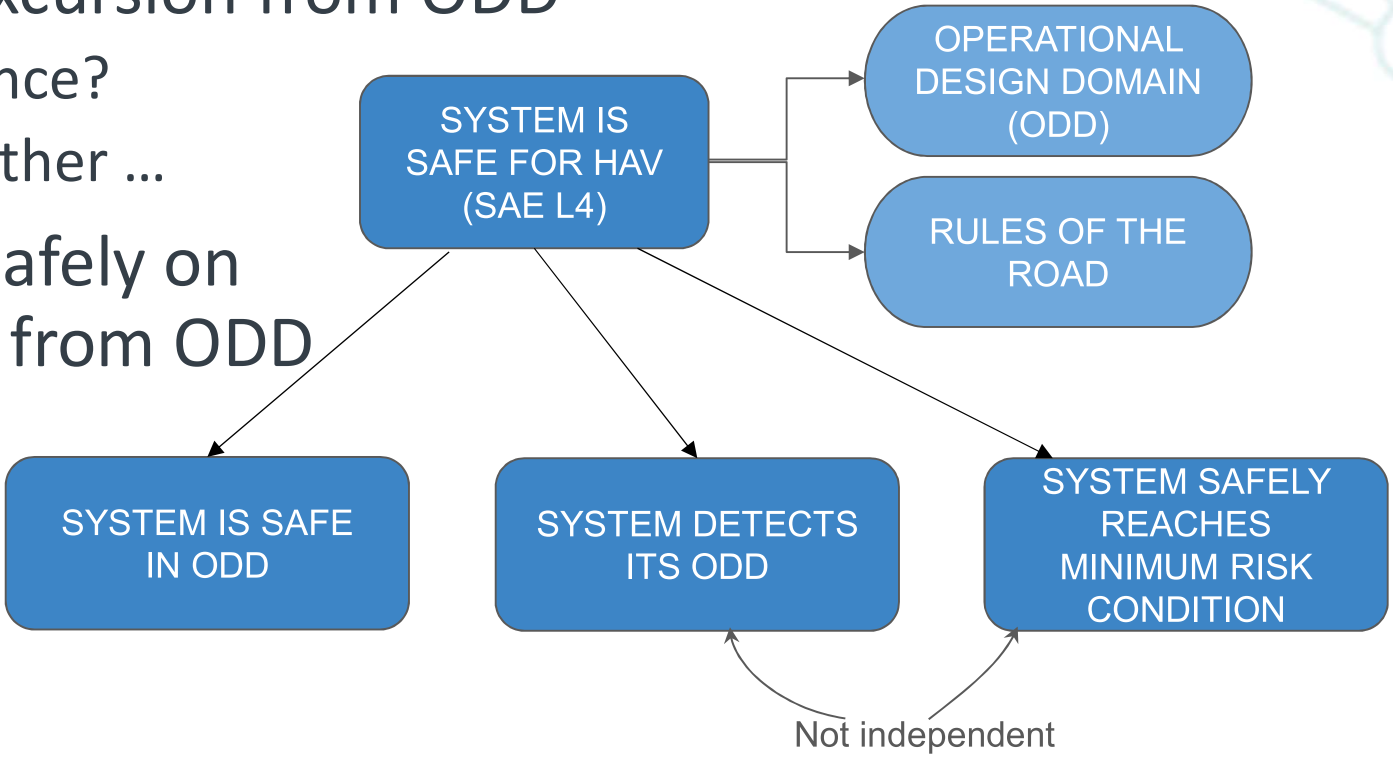
Constraining Behaviour

- Open world
  - Unconstrained, uncertain, hard to assure
- Closed world
  - Bounded, (fairly) certain, relatively easy to assure
- Defined world
  - Constrain open world to bound uncertainty – operational design domain (ODD)

New Approach

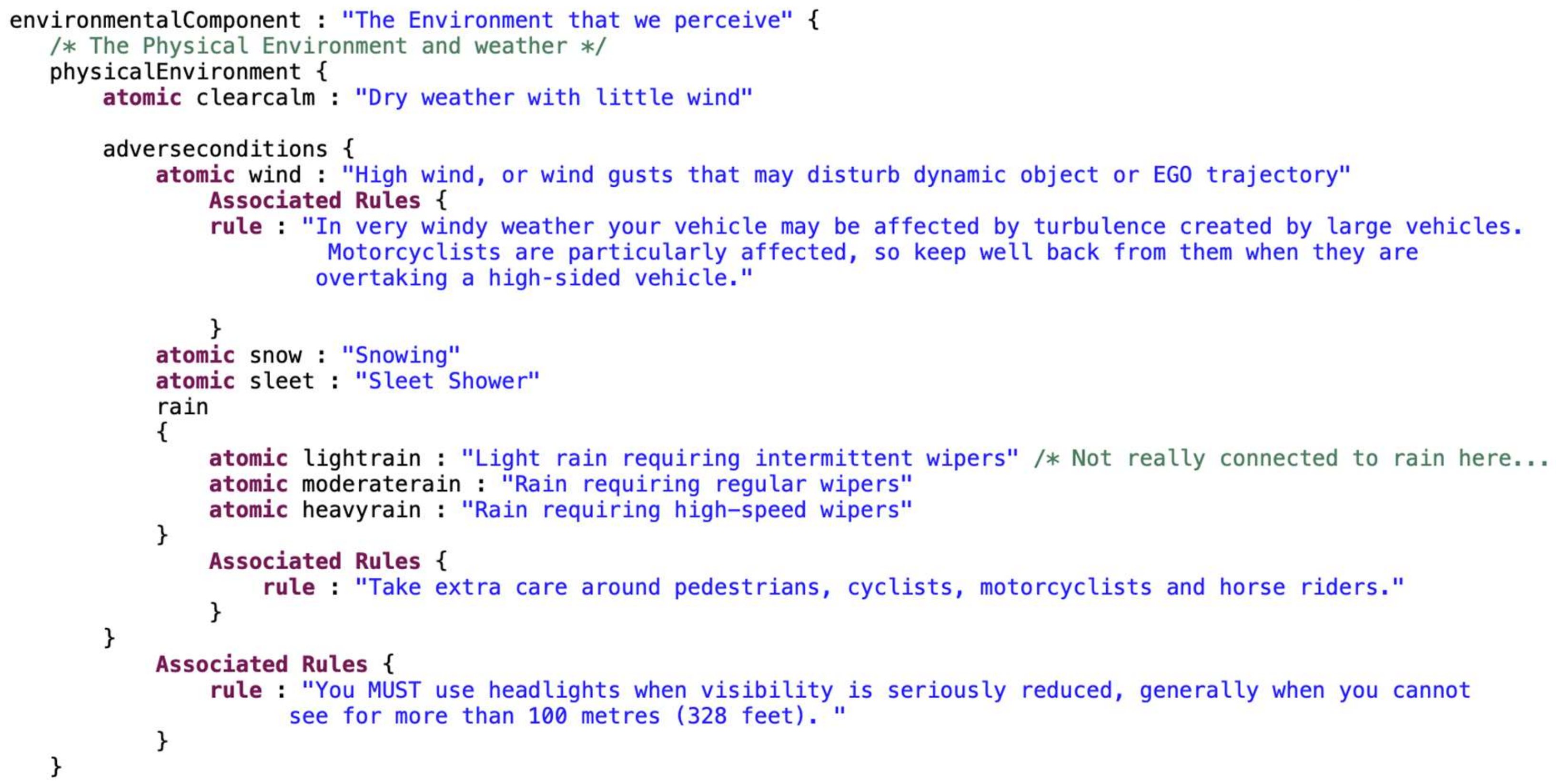
Exploiting the ODD

- Within ODD performs safely
- Detects excursion from ODD
  - In advance?
  - NB weather …
- Behaves safely on excursion from ODD


SYSTEM IS SAFE FOR HAV (SAE L4)
OPERATIONAL DESIGN DOMAIN (ODD)
RULES OF THE ROAD
SYSTEM IS SAFE IN ODD
SYSTEM DETECTS ITS ODD
SYSTEM SAFELY REACHES MINIMUM RISK CONDITION
Not independent

New Approach

Specify ODD and Associated Rules



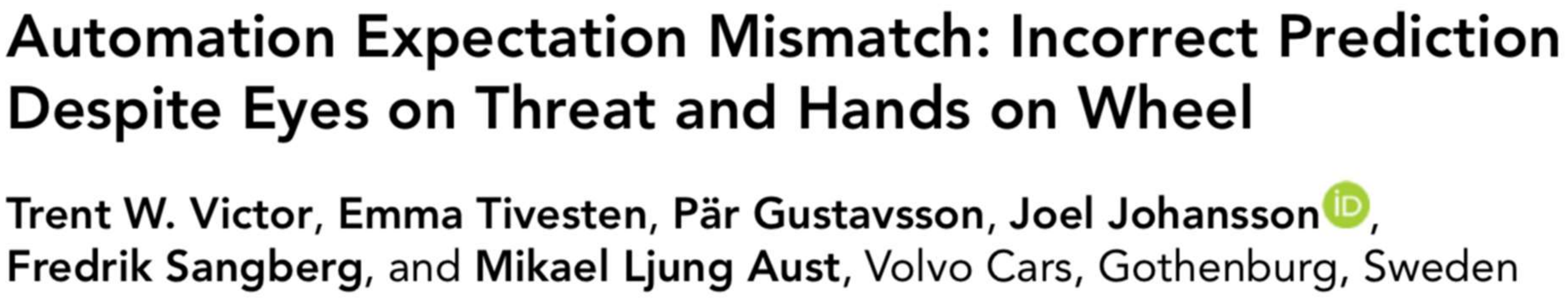
```
environmentalComponent : "The Environment that we perceive" {
   /* The Physical Environment and weather */
   physicalEnvironment {
       atomic clearcalm : "Dry weather with little wind"

       adverseconditions {
           atomic wind : "High wind, or wind gusts that may disturb dynamic object or EGO trajectory"
               Associated Rules {
               rule : "In very windy weather your vehicle may be affected by turbulence created by large vehicles.
                        Motorcyclists are particularly affected, so keep well back from them when they are
                       overtaking a high-sided vehicle."

               }
           atomic snow : "Snowing"
           atomic sleet : "Sleet Shower"
           rain
           {
               atomic lightrain : "Light rain requiring intermittent wipers" /* Not really connected to rain here...
               atomic moderaterain : "Rain requiring regular wipers"
               atomic heavyrain : "Rain requiring high-speed wipers"
           }
               Associated Rules {
                   rule : "Take extra care around pedestrians, cyclists, motorcyclists and horse riders."
               }
       }
           Associated Rules {
               rule : "You MUST use headlights when visibility is seriously reduced, generally when you cannot
                     see for more than 100 metres (328 feet). "
           }
   }
```

Approach being developed by FiveAI

Why is Assurance Hard?

Human System Interaction

- How do autonomous and human operated systems "understand" each other
  - Including intentions?
  - How do we manage "hand over"?
  - What is it realistic to expect of drivers?

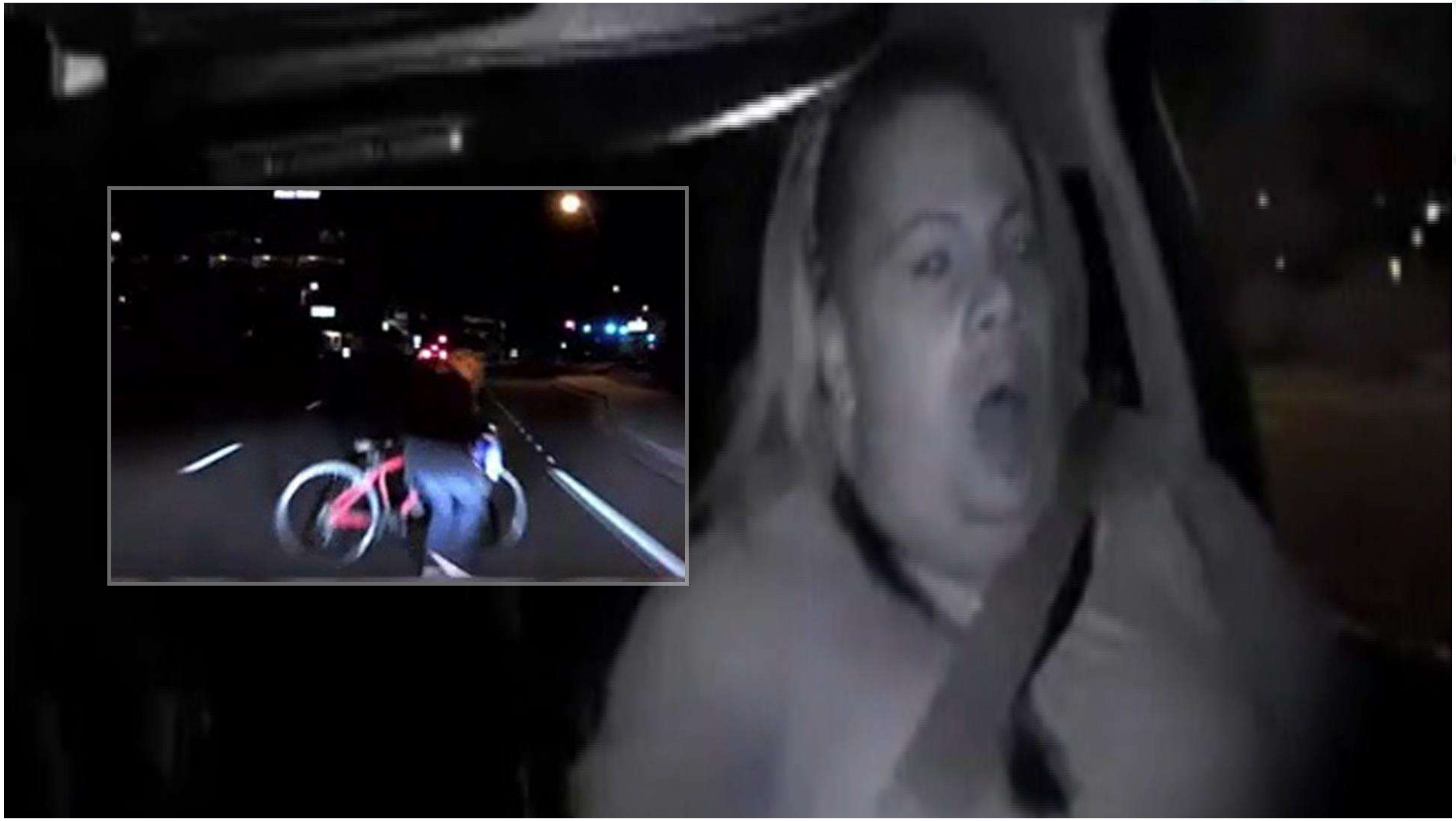


Automation Expectation Mismatch: Incorrect Prediction Despite Eyes on Threat and Hands on Wheel

Trent W. Victor, Emma Tivesten, Pär Gustavsson, Joel Johansson, **Fredrik Sangberg**, and **Mikael Ljung Aust**, Volvo Cars, Gothenburg, Sweden

Why is Assurance Hard?

Situational Awareness



Assurance and Humans

Human System Interaction

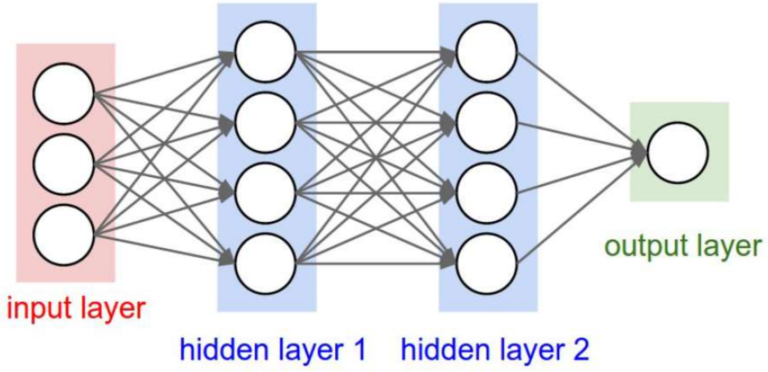
- Arguments and evidence need to address
  - Ability of driver to maintain concentration in monitoring mode
  - The length of time for the driver to regain situational awareness after disengaging from driving task
  - Ability of the system to alert the driver effectively
  - Social cognition (and disruption from autonomy)
- Many believe SAE L3 automation problematic
  - Easier to do SAE L4 HAV or FAV – but this shifts the problem of assurance to the ML components

System Complexity

The Appeal of Machine Learning

- Ability to learn and generalize beyond training data
  - By making sense of unstructured data, machine learning is particularly suited to open context systems
  - Deep learning enabled computers to learn tasks that seemed to be intractable for computer programs before

But learnt models very complex (100s of dimensions)



Source: www.cityscapes-dataset.com

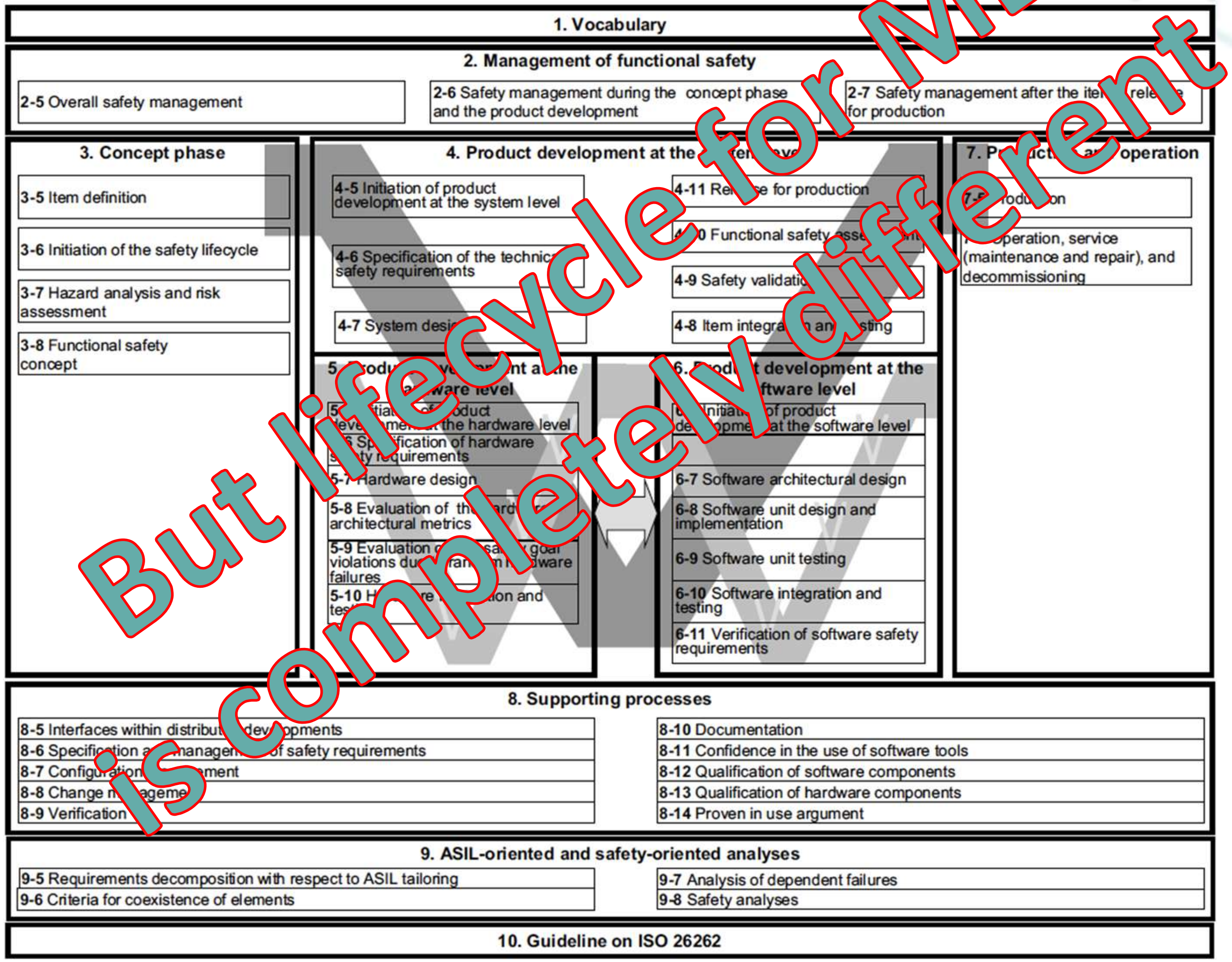
Why is Assurance Hard?

Complexity and Uncertainty

- Inherently uncertain (“80% sure, it is a vehicle”)
  - And uncertainty propagates
- Learnt models are high-dimensional, in general not understandable by humans
  - Models are ‘black box’ and what is learnt is hard to explain
- No structured verification methods for learnt functions and *currently no commonly agreed means to argue their safety*

Current Approaches

ISO 26262 Lifecycle Model(s)



1. Vocabulary

2. Management of functional safety

- 2-5 Overall safety management
- 2-6 Safety management during the concept phase and the product development
- 2-7 Safety management after the item's release for production

3. Concept phase

- 3-5 Item definition
- 3-6 Initiation of the safety lifecycle
- 3-7 Hazard analysis and risk assessment
- 3-8 Functional safety concept

4. Product development at the system level

- 4-5 Initiation of product development at the system level
- 4-6 Specification of the technical safety requirements
- 4-7 System design
- 4-11 Release for production
- 4-10 Functional safety assessment
- 4-9 Safety validation
- 4-8 Item integration and testing

5. Product development at the hardware level

- 5-5 Initiation of product development at the hardware level
- 5-6 Specification of hardware safety requirements
- 5-7 Hardware design
- 5-8 Evaluation of the hardware architectural metrics
- 5-9 Evaluation of safety goal violations due to random hardware failures
- 5-10 Hardware integration and testing

6. Product development at the software level

- 6-5 Initiation of product development at the software level
- 6-7 Software architectural design
- 6-8 Software unit design and implementation
- 6-9 Software unit testing
- 6-10 Software integration and testing
- 6-11 Verification of software safety requirements

7. Production and operation

- 7-5 Production
- 7-6 Operation, service (maintenance and repair), and decommissioning

8. Supporting processes

- 8-5 Interfaces within distributed developments
- 8-6 Specification and management of safety requirements
- 8-7 Configuration management
- 8-8 Change management
- 8-9 Verification
- 8-10 Documentation
- 8-11 Confidence in the use of software tools
- 8-12 Qualification of software components
- 8-13 Qualification of hardware components
- 8-14 Proven in use argument

9. ASIL-oriented and safety-oriented analyses

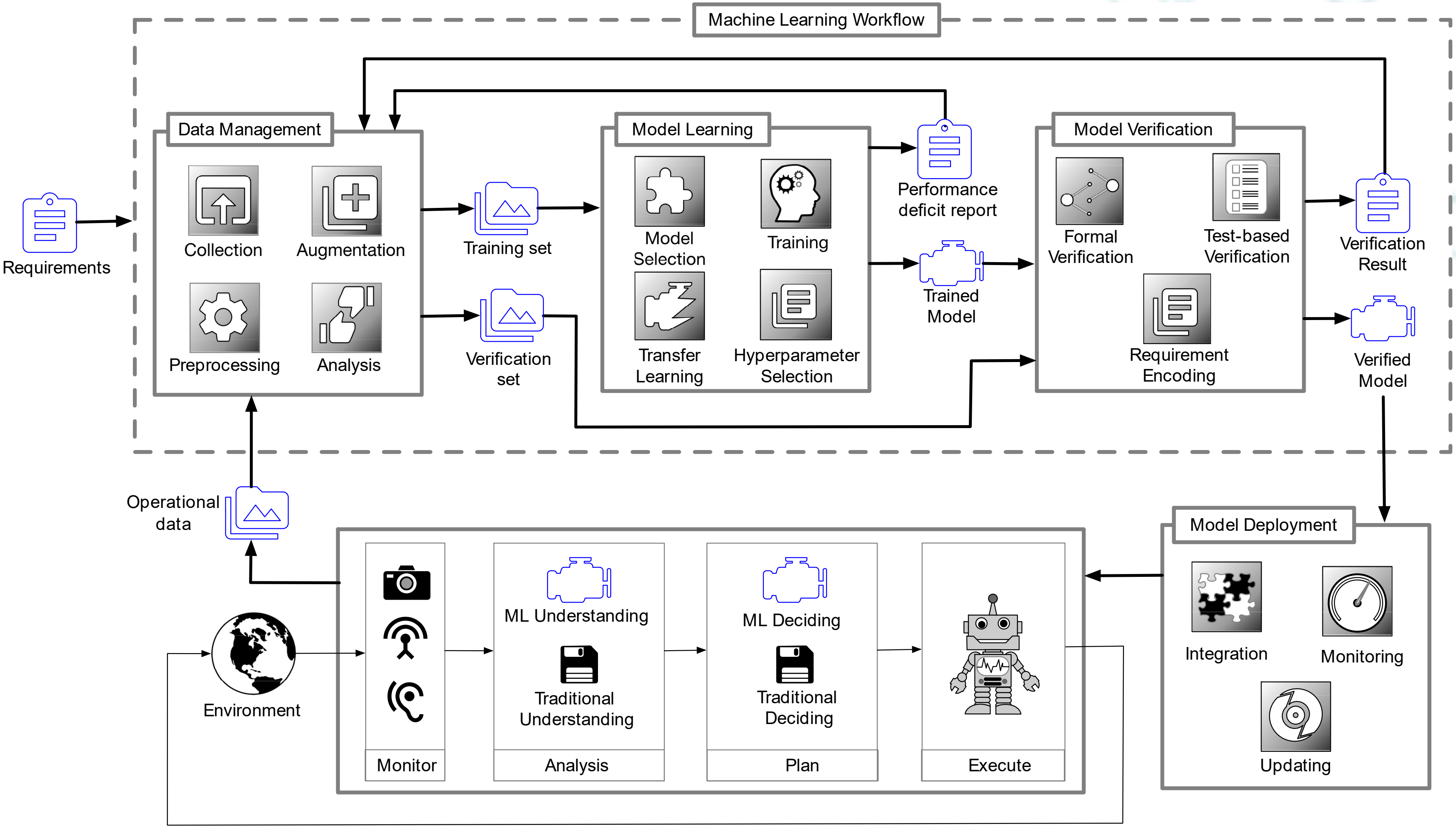
- 9-5 Requirements decomposition with respect to ASIL tailoring
- 9-6 Criteria for coexistence of elements
- 9-7 Analysis of dependent failures
- 9-8 Safety analyses

10. Guideline on ISO 26262

But lifecycle for ML is completely different

New Approach

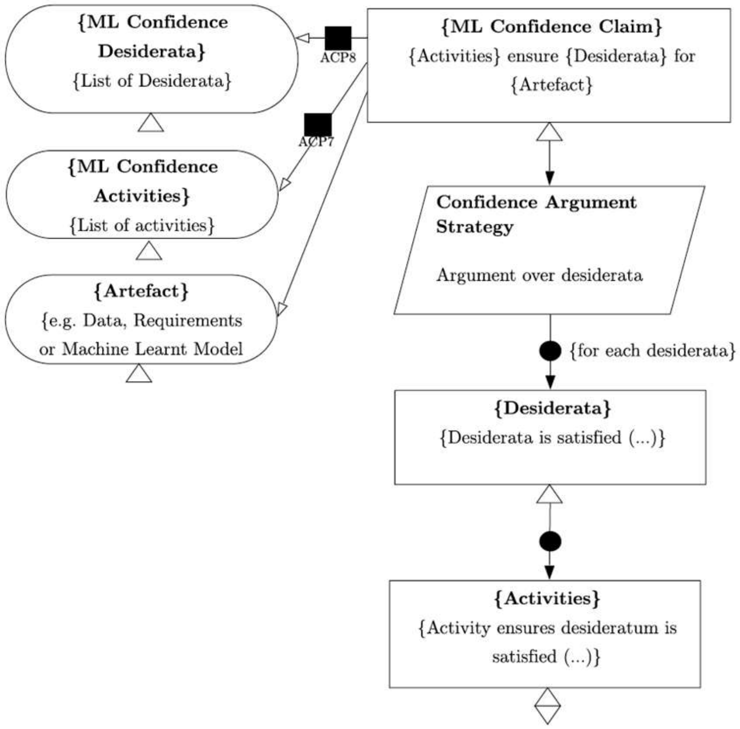
ML Lifecyle Model



New Approach

Assurance Argument and Evidence

- Core of assurance case remains the same as non-AV or non-ML case
  - Confidence argument addresses uncertainties
  - Example is the ML (model learnt) confidence claim and associated argument
- Currently experimenting with this approach in ISO 26262 context



{ML Confidence Desiderata}
{List of Desiderata}

{ML Confidence Activities}
{List of activities}

{Artefact}
{e.g. Data, Requirements or Machine Learnt Model

{ML Confidence Claim}
{Activities} ensure {Desiderata} for {Artefact}

Confidence Argument Strategy
Argument over desiderata

{for each desiderata}

{Desiderata}
{Desiderata is satisfied (...)}

{Activities}
{Activity ensures desideratum is satisfied (...)}

Regulation

UN ECE Objectives

- To define a performance requirement for (AI in) AVs such that

  1. AI never engages in careless, dangerous or reckless driving behaviour (*argue about SOTIF*)
  2. AI remains aware, willing and able to avoid collisions at all times (*safety process and confidence arguments*)
  3. AI meets, or exceeds, the performance of a competent & careful human driver (*requirements validation*)

- Approach to analysis of ML and assurance cases gives a basis for doing this (especially 1 and 2)
  - Can be enshrined in regulations

Related Work

A (Very) Selective Sample

- UL 4600 led by Phil Koopman
  - Assurance case approach, but very detailed
- BSI activities on AVs funded by the Centre for Connected Autonomous Vehicles (CCAV)
  - Various PAS and I chair the Advisory Board
- ISO 26262 work on SOTIF Assurance Arguments
  - Using York's Goal Structuring Notation (GSN)
- CCAV intend to develop a safety process for AVs
- Industry activities, e.g. FiveAI (I'm an advisor)

Demonstrator Projects

Sense-Assess-Explain (SAX)

Building autonomous vehicles, that can sense and fully understand their environment, assess their own capabilities, and provide causal explanations for their decisions.

Automatic Testing Mechanism (ATM)

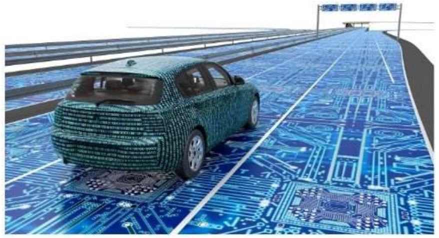
Designing an automatic testing mechanism for the autonomous software that controls a RAS. In the longer term it will form the core component of an automatic rating system to assess the safety of RAS (akin to the NCAP rating for cars).

Safe:SCAD - Safety of shared control in autonomous driving

Using an advanced semi-autonomous driving simulator to deliver methods for ensuring and assuring the safety of shared control in autonomous driving.

Towards identifying and closing gaps in assurance of autonomous road vehicles (TIGARS)

Investigating the assurance gaps and challenges for first generation autonomous systems and research into techniques and engineering processes for addressing them.



Programme Fellows

Dr Simon Burton

Organisation: Robert Bosch GmbH, Germany

Job title: Chief Expert – Vehicle Computer Software, Safety and Security

Fellowship focus: Safety assurance for machine learning in autonomous driving

Dr Jelena Frtunikj

Organisation: BMW AG, Germany

Job title: Machine Learning and Safety Expert for Automated Driving

Fellowship focus: the definition of safety related metrics/KPIs and techniques for evaluating the performance of machine learning algorithms

Lydia Gauerhof

Organisation: Robert Bosch GmbH, Germany

Job title: Research engineer

Fellowship focus: confidence arguments for machine learning algorithms

Dr Roger Rivett

Job title: Independent Consultant (former Functional Safety Technical Specialist, Jaguar Land Rover)

Fellowship focus: focusing on producing a generic automotive sector argument framework

Conclusions

Suggestions to the WG

- Need to be aware of other relevant activities and build on or complement them
  - I identified several, but there is much more work on standards internationally to consider
- Programme working across domains
  - But several activities focused on AVs
  - Believe ML lifecycle and approach to assurance unique
- Regulation is a major focus for the Programme
  - Happy to work with the WG and to assess the extent to which the Programme’s approach can support the WG

Discussion

Over to you

UNIVERSITY
of York

ASSURING
AUTONOMY
INTERNATIONAL PROGRAMME

Funded by

Lloyd's Register
Foundation

UNIVERSITY
of York