Assurance of Remote Inspection Robots: Some Perspectives

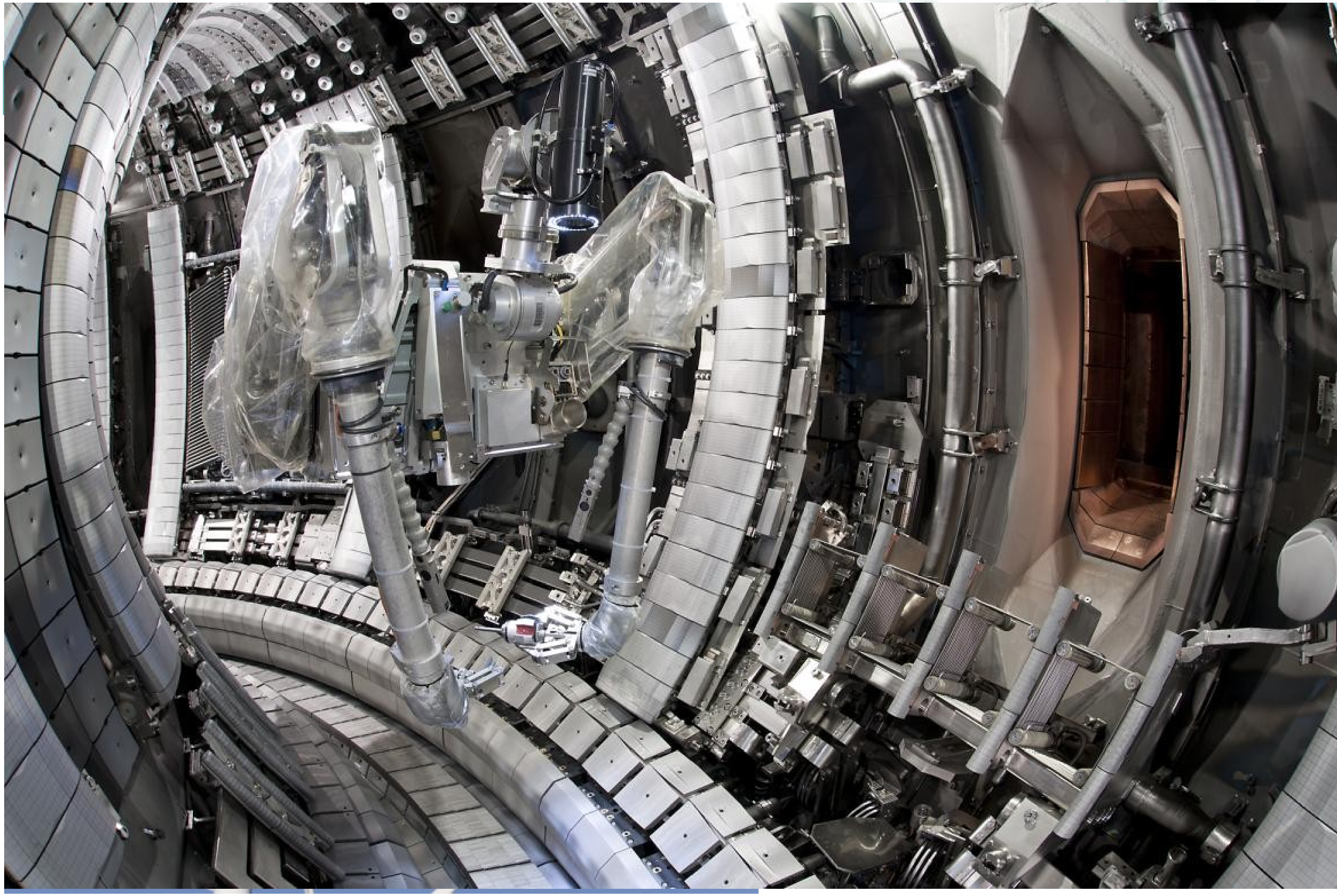
Prof John McDermid OBE FREng

Overview

Agenda

- Challenges for remote inspection robots
- Assuring Autonomy International Programme
- An assurance perspective
  - System models
  - Assurance of ML
  - Safety processes
- A legal perspective
- Insights
- Conclusions

Remote Inspection



Technical Challenges

Remoteness and Other Issues

- Autonomy
  - Able to make own decisions (but also shared control)
- Communication
  - Limited/no bandwidth and/or intermittent
  - Long round-trip delays and poor situational awareness
- Adaptive
  - Respond to changing environment and own state (repair)
- Long-lived
  - Missions of months or more

Challenges of Shared Control

Expectations on the Operator

- What is it realistic to assume of drivers
  - How long can they retain situational awareness?
  - How will they react in an incident? For example some data from Volvo relating to emergency braking
    - 1/3rd took control promptly
    - 1/3rd took control late, waiting for the autonomy
    - 1/3rd took no action, wanting to avoid “interfering” with the autonomy

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

Assurance Challenges

Safety and Other Properties

- Generic assurance and regulatory challenge
  - A safe system cannot be deployed or is frequently unavailable (losing benefit)
  - An unsafe system is deployed (as it is approved due to lack of contrary evidence)
  - Similar issues for availability, mission effectiveness …
- Addressing the technical challenges
  - Especially verification and validation for critical technologies including machine learning (ML)

Fundamental Challenges

AI/ML vs Human Decision-Making

- Autonomous systems
  - Transfer decision-making from human to machine (AI/ML)
  - ML learns future behaviour generalising from training data
- Humans have a semantic model, e.g. know what a valve is and its likely behaviour
  - Machines do not have these models
- Humans have contextual models, e.g. know what a pipeline is
  - And the effects of pressure, corrosion, silting up …
  - Machines do not have these models

Fundamental Challenges

AI/ML Safety

- Safety processes assume
  - Know system boundary and it is fixed
  - Know (can specify precisely) system behaviour
  - Know system environment and can assess hazards
  - Life-cycle progressively adds detail so can analyse easily
- With AI/ML
  - Behaviour not known precisely (learnt not specified)
  - Environment extremely complex (unpredictable)
  - Life-cycle highly iterative
  - Boundary and functions can also change

Overview

Agenda

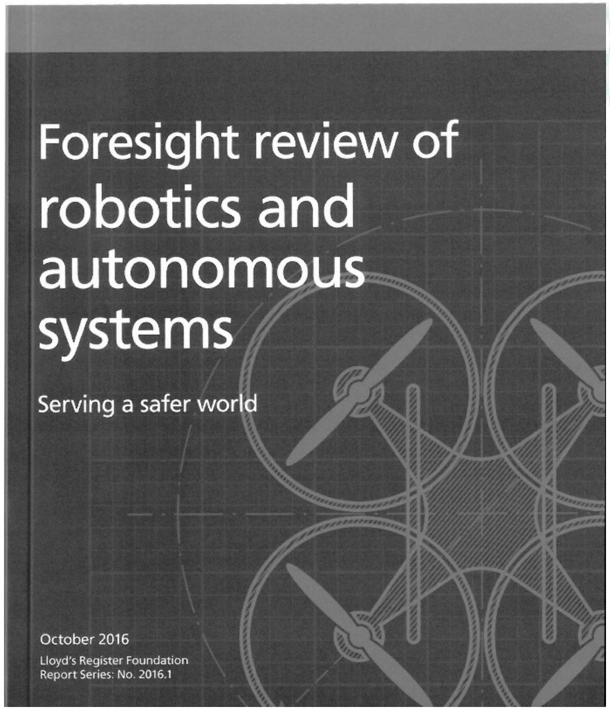
- Challenges for remote inspection robots
- Assuring Autonomy International Programme
- An assurance perspective
  - System models
  - Assurance of ML
  - Safety processes
- A legal perspective
- Insights
- Conclusions

Assuring Autonomy

Response to Foresight Review

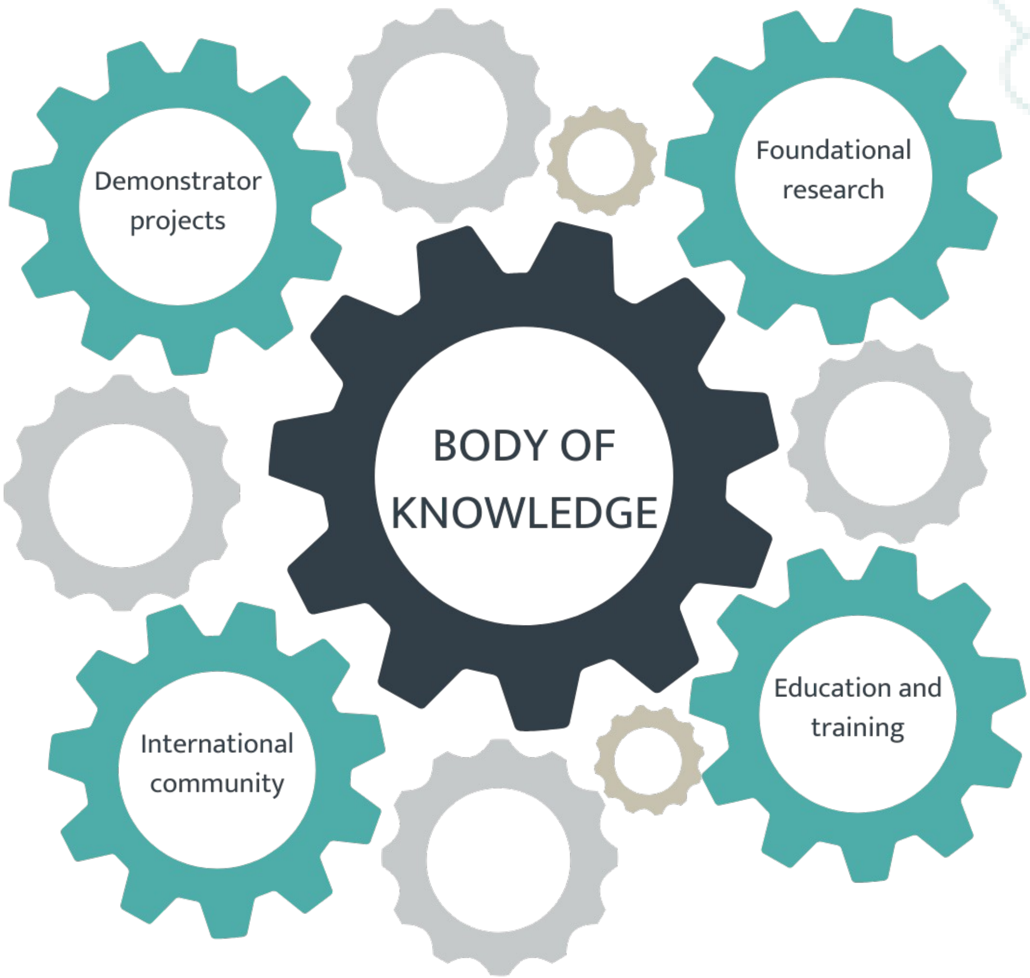
- Review published in October 2016
  - Identified "white spaces" in assurance and regulation of RAS
- York-led programme
  - January 2018 to December 2022(3)
  - A strong focus on 'demonstrators' and working 'bottom up'
  - Developing international links, and seeking to influence policies and regulations

Lloyd's Register Foundation



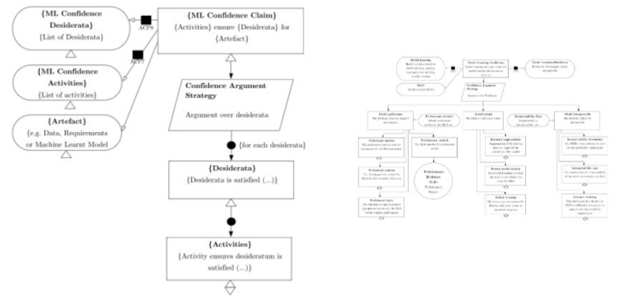
The Programme

Scope of Activities



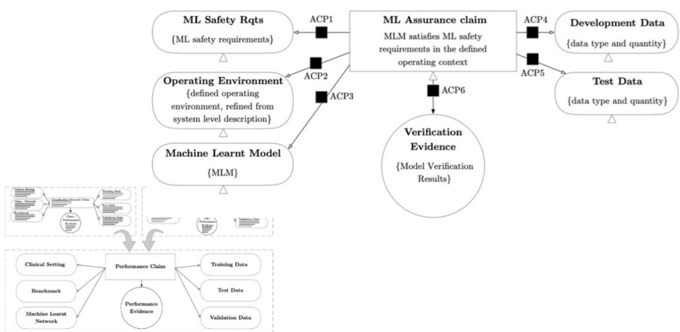
Research Summary

AI/Autonomy Safety Cases

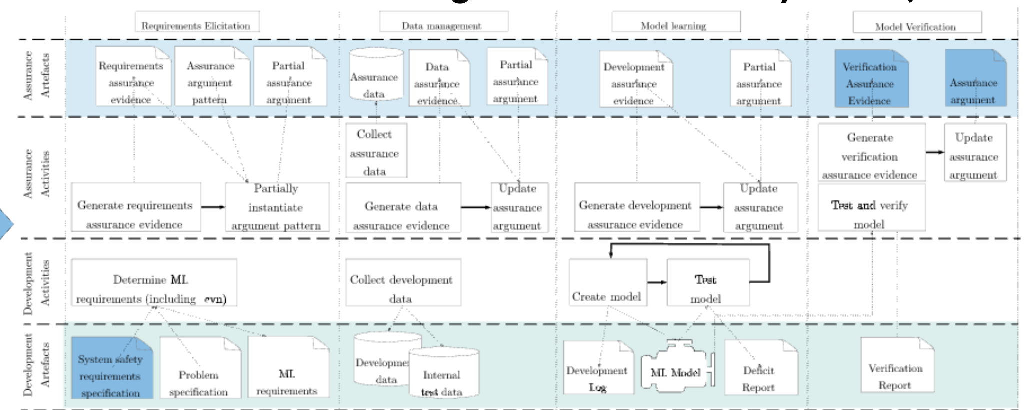


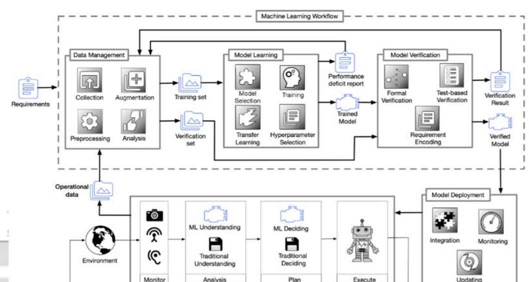
AI Ethics & Governance

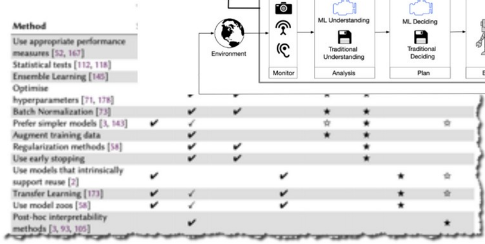
ML Assurance Cases



Assurance of Machine Learning for Autonomous Systems (AMLAS)

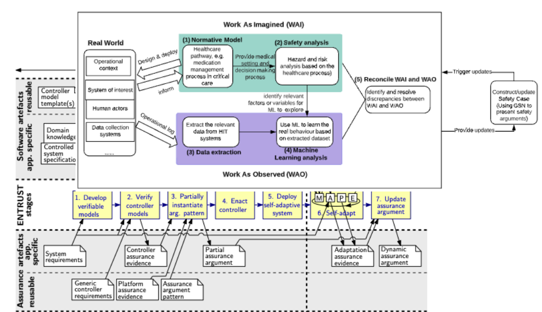






ML Assurance Survey

ML Verification



Dynamic Safety cases

Demonstrator Projects

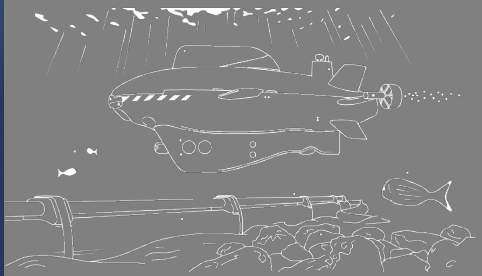
Relevant to Remote Inspection

Safe Airframe Inspection using Multiple UAVs (SAFEMUV)

Improving the safety of autonomous unmanned aerial vehicle teams through the creation of a systematic robustness assessment process

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 own decisions.



Assuring Long-term Autonomy through Detection and Diagnosis of Irregularities in Normal operation (ALADDIN)

Increasing the safety of unmanned marine systems by helping the vehicles identify the cause of their adverse behaviour.

Overview

Agenda

- Challenges for remote inspection robots
- Assuring Autonomy International Programme
- An assurance perspective
  - System models
  - Assurance of ML
  - Safety processes
- A legal perspective
- Insights
- Conclusions

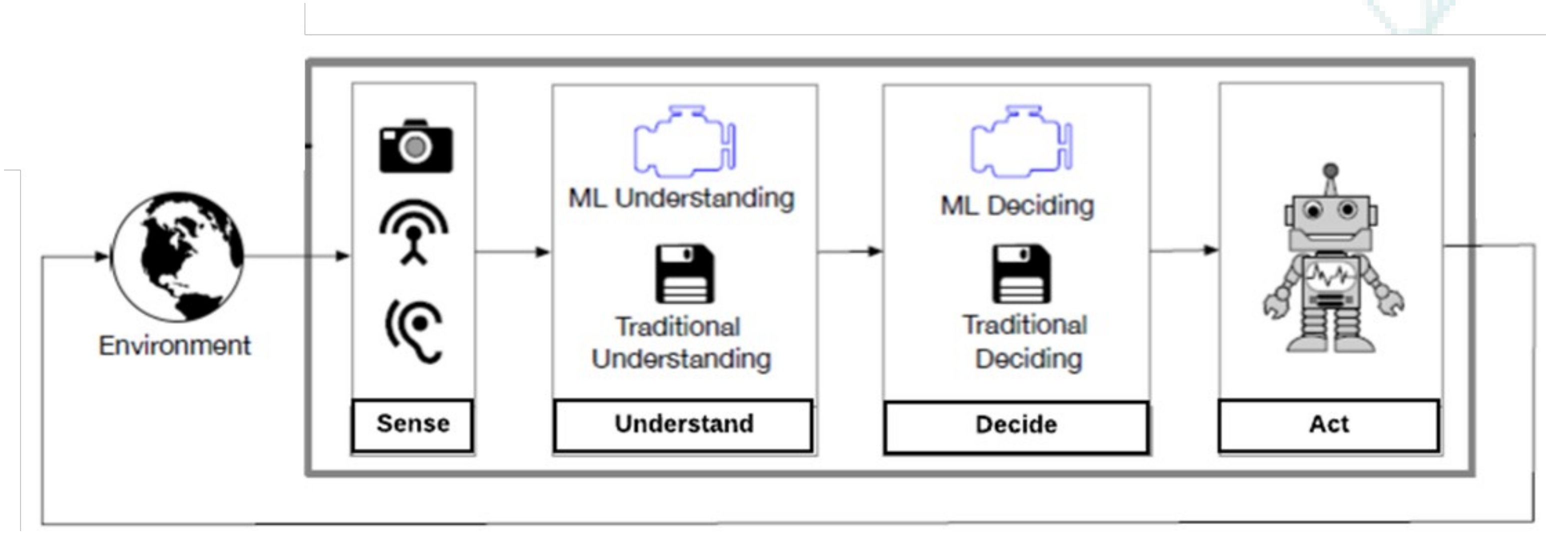
Assurance

Assurance and V&V

- Assurance can be thought of as:
  - Confidence that the system behaviour is as intended in the environment of use (as intended includes safe)
- For autonomy, three key elements to assurance
  - Defined intent – know what it should do and avoid doing (e.g. safety) [Validation]
  - Correct implementation – meets its intent [Verification]
  - Malfunction control – behaves appropriately when things go wrong, e.g. sensors are affected by weather, internal components, etc. [Verification & Validation]

A System Model

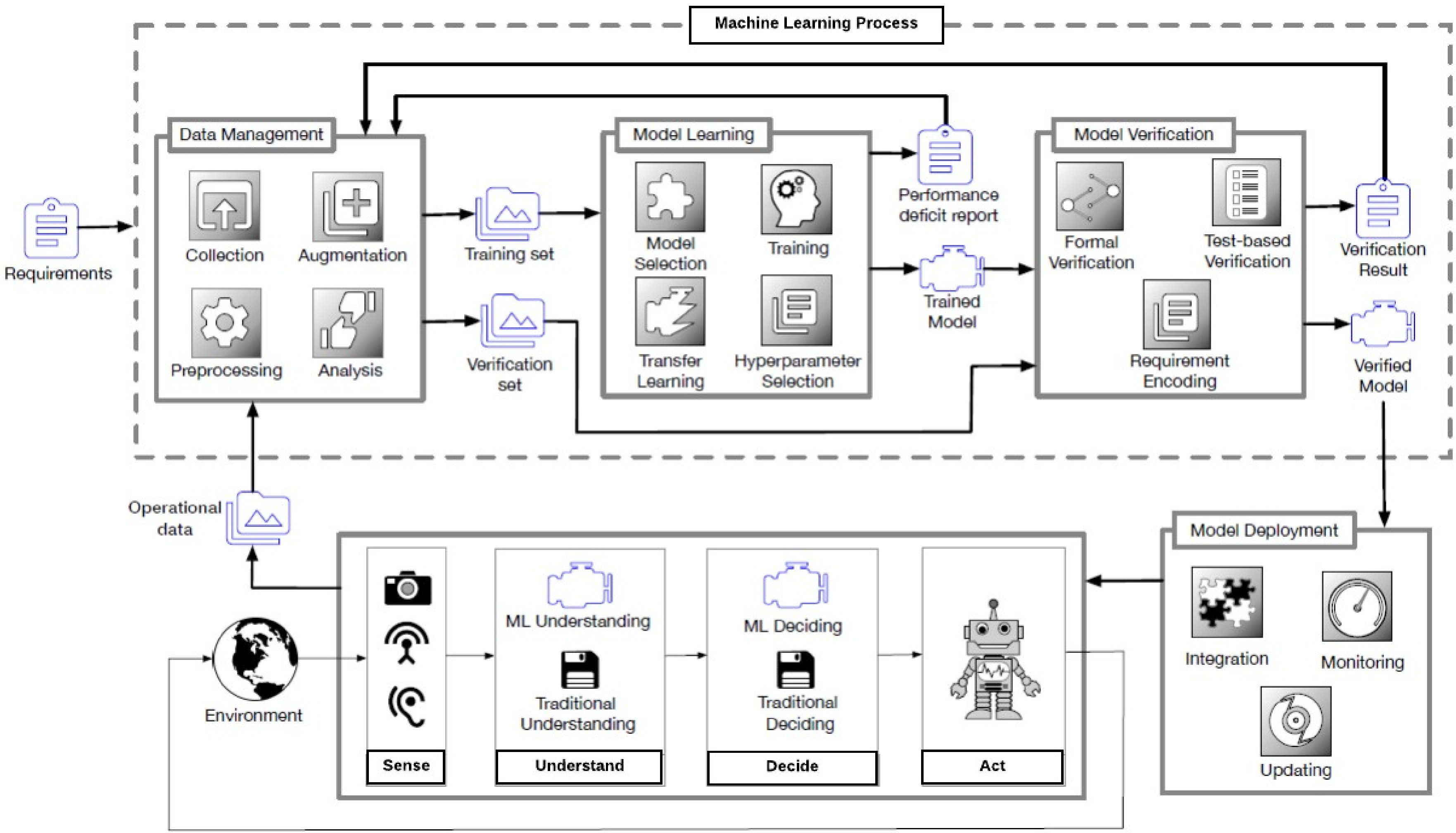
Sense-Understand-Decide-Act (SUDA)



- System operates cyclically
  - Understanding includes prediction, e.g. trajectory of drone
- AI/ML usually limited to Understand and Decide (SUDA)
- Variants of model, e.g. Sense and Understand merged

Assuring Machine Learning

ML Process and SUDA



Assuring Machine Learning

Table 4. Open challenges for the assurance concerns associated with the Model Learning (ML) stage

ID	Open Challenge	Desideratum (Section)
ML01	Selecting measures which represent operational context	Performant (Section 5.4.1)
ML02	Multi-objective performance evaluation at run-time	
ML03	Using operational context to inform hyperparameter-tuning strategies	
ML04	Understanding the impact of hyperparameters on model performance	
ML05	Decoupling the effects of perturbations in the input space	Robust (Section 5.4.2)
ML06	Inferring contextual robustness from evaluation metrics	
ML07	Identifying similarity in operational contexts	Reusable (Section 5.4.3)
ML08	Ensuring existing models are free from faults	
ML09	Global methods for interpretability in complex models	Interpretable (Section 5.4.4)
ML10	Inferring global model properties from local cases	

support reuse [2]					
Transfer Learning [173]	✔	✓	✔	★	☆

Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges

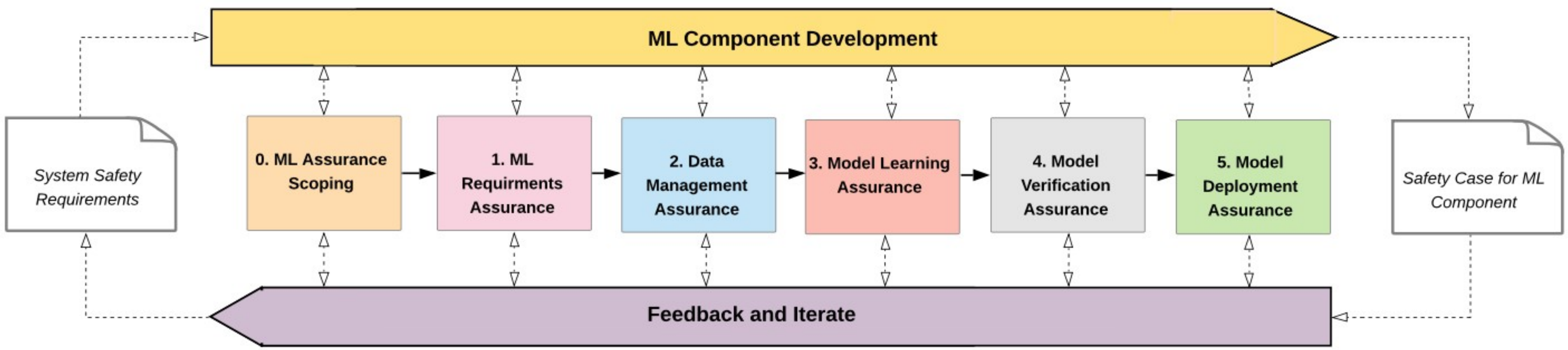
Rob Ashmore, Radu Calinescu, Colin Paterson

[†] ✔ = activity that the method is typically used in; ✓ = activity that may use the method

[‡] ★ = desideratum supported by the method; ☆ = desideratum partly supported by the method

Assuring Machine Learning

AMLAS - Assurance of Machine Learning for RAS

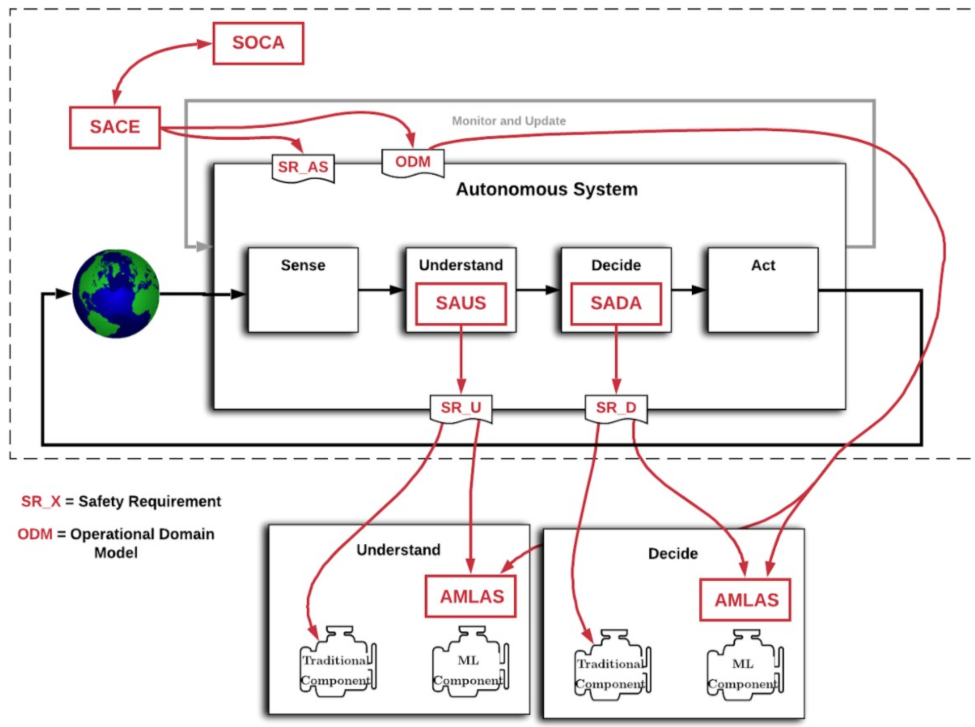


- Defined **assurance process for ML components**
- Results in a {compelling?} safety case for ML component(s) of the system
- Considers safety of ML in system context

An AI Safety Process

SUDA, AMLAS and More

- Safety processes
  - SOCA: acceptability
  - SACE: whole system
  - SAUS: understanding
  - SADA: decision-making
  - AMLAS: assurance of ML
- Shared control is addressed by SACE



SR – Safety Requirement

Overview

Agenda

- Challenges for remote inspection robots
- Assuring Autonomy International Programme
- An assurance perspective
  - System models
  - Assurance of ML
  - Safety processes
- A legal perspective
- Insights
- Conclusions

Legal Issues

Mind the Gaps

- In many legal frameworks need to “fix where responsibility lies” to have a case
  - Autonomy can introduce “liability gaps” – despite an accident can’t attribute responsibility (appropriately)
  - For example, the Tempe autonomous vehicle fatality – Uber found to have no case to answer under Arizona law
    - Likely to be a widespread issue
  - Also, ethical perspective on when it is appropriate to attribute responsibility to (legal individuals)

Burton et al. “Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective.” Artificial Intelligence, Volume 279, February 2020

RIMA Project

Robotics for Inspection & Maintenance

- EU Project – funded by European Union's Horizon2020 initiative
  - Major focus is on infrastructure
- Report written by University of York covering legal framework for operating RAS in different countries recently published
  - Takes a legal and regulatory perspective
  - Some of the legal issues and constraints likely to be of wider significance

D7.4 Review of legal frameworks, standards and best practice in verification and assurance for infrastructure robotics

Insights

From AAIP, RIMA, etc.

- Verification is hard
  - A lot missing, e.g. appropriate performance criteria, test coverage criteria informed by fault models for ML
- Validation is harder
  - Need to link to safety (availability, maintainability …)
- Adaptation goes beyond (most) current regulations
  - Will need to consider dynamic risk assessment
- Shared control is problematic (NB ALKS)
  - Need refined safety processes with input from human factors specialists

Regulatory Strategies

Regulation and Innovation

- **No response** – is “mute” about AI and RAS
- **Prevention-oriented** – proscribes use of aspects of AI and RAS, e.g. adaptation in operation
- **Control-oriented** – seek to control the technology
- **Toleration-oriented** – allow innovation, with a degree of scrutiny – i.e. largely responsive
- **Adaptation-oriented** – changes to respond to the technology – but how do we keep pace?

Conclusions

V&V for Inspection Robotics

- AAIP considering broad issues of RAS assurance
  - Focus on safety, but likely that approach to system models and ML assurance (AMLAS) of wider applicability
  - Some demonstrator projects of direct relevance
- Interested in collaborating on applications
  - Validate/refine AMLAS, encourage links for demonstrators
  - Address issues of “how much evidence is enough”
- Are open research challenges
  - For example, test coverage criteria, safe interaction of “swarms” of robots (and humans), and security-informed safety
  - Many will benefit from interdisciplinary approaches

UNIVERSITY of York

ASSURING AUTONOMY
INTERNATIONAL PROGRAMME

Funded by

Lloyd's Register Foundation

UNIVERSITY of York