THALES

Operationalising the adoption of AI to counter cyber attacks

Peter Davies
Director Security Concepts

8TH JUN 2022

Some Questions …

- How s ML / AI different from your regular complex system?
- Where are you using ML / AI?
- What is your failure mode for your ML / AI system?
- What is the standard and type of evidence required of your ML system?

Some Questions …

FLiNT

Calls for global 'data extradition treaties' after Brighton road tragedy

The United Road Transport Union is calling for new international data extradition treaties to be signed to help authorities get to the ground truth in road traffic incident investigations. The call comes after a young family was killed on the M23 just outside Brighton last month, in the latest road tragedy to be blamed on a cyber-security breakdown. Their hatchback was involved in a fatal collision when the Automated Lane Keeping Assist of a Dutch-owned HGV allegedly failed and the Ukrainian driver lost control of the vehicle. The HGV driver, who escaped the collision unhurt, was initially suspected of falling asleep at the wheel. But in a statement released through his union yesterday he insists that the Lane Assist feature "froze-up" in the minutes before the accident.

He claims that the feature was remotely "hacked" and that he had struggled desperately for a minute or more to recover control of the steering system, trying to take corrective action to avoid the fatal collision. With the support of his union, he is now calling for the vehicle software manufacturer XG to share user data that he believes would help prove the automated Lane Assist was compromised. He hopes these data will also provide evidence to show that he took prompt and appropriate action in his attempts to regain control of the vehicle.

XG have refused to allow access to such data, citing European privacy and security regulations. They deny any liability in the cluster of recent cases involving their automated driver assist products.

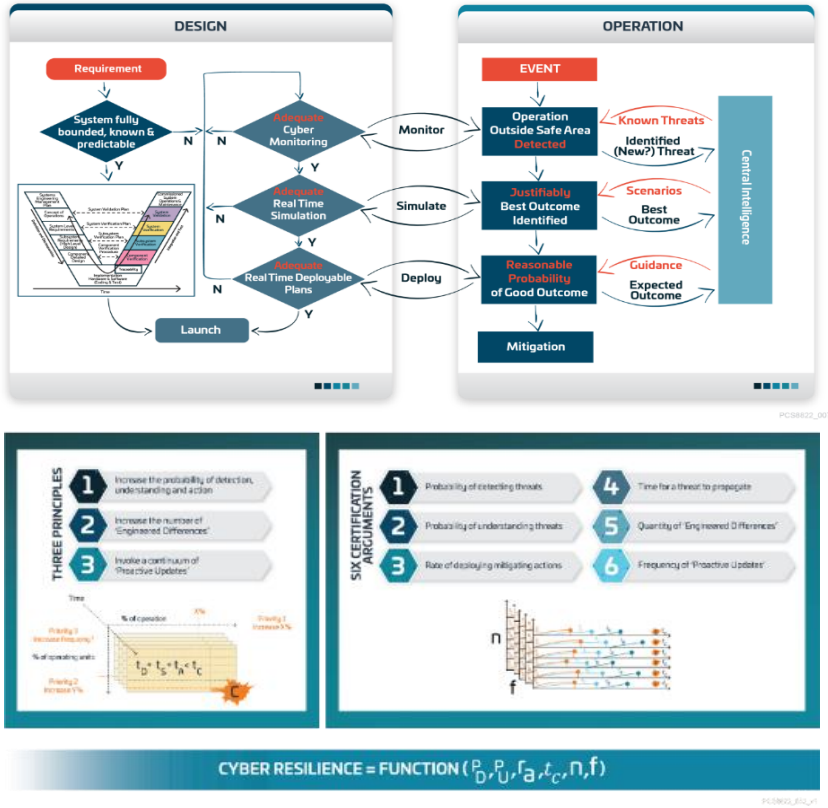
URTU spokesman William Johnson read a statement earlier today by the memorial flowers laid out at the scene of the accident

"This tragedy highlights the importance of sharing data across jurisdictions" he said. "If this accident had been caused by faulty brakes or tyres, then police could check the physical evidence. They need access to the systems data to investigate properly. This isn't about cyber security, it's about cyber safety." The URTU will ballot their members on strike action to support this case on Monday.

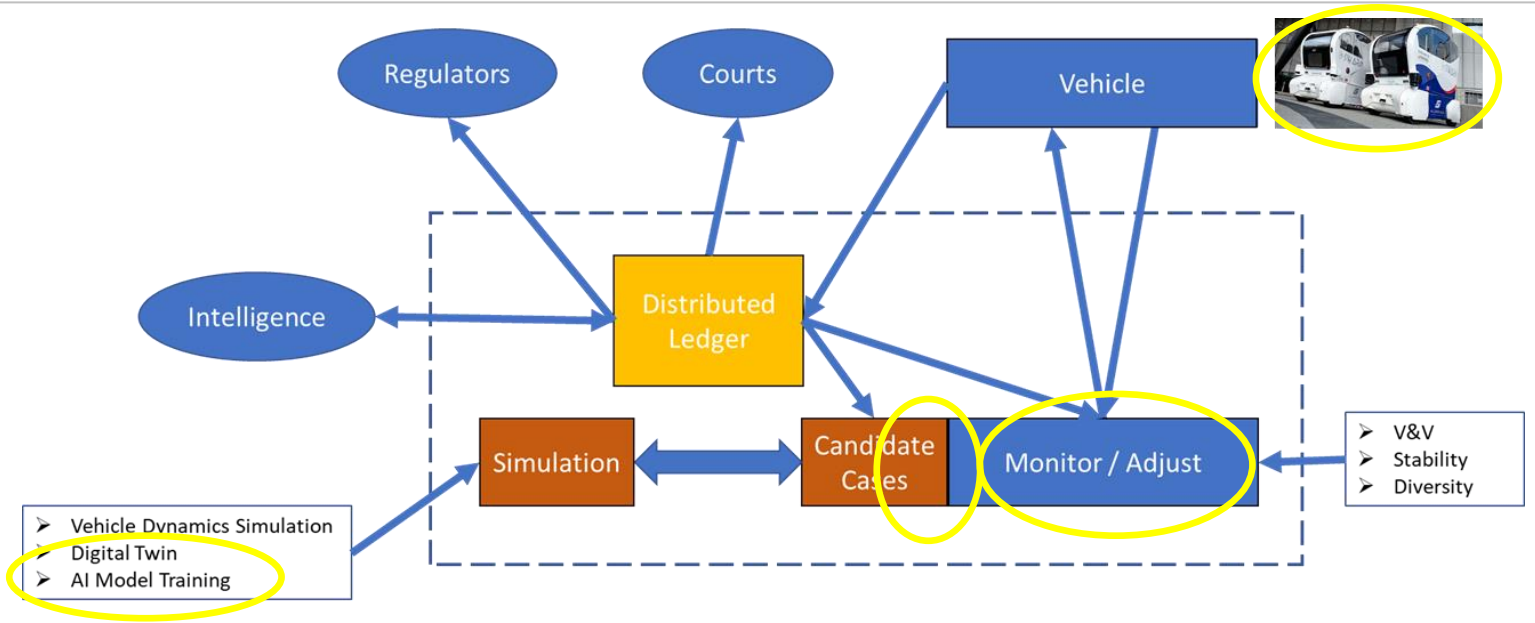
What are you trying to acheive …

An operational methodology, suitable for standardisation, for which:

- **The methodology itself** is capable of being tested in court or by publicly appointed regulators.
- **Operators** understand what evidence should be produced by it and are able to measure the quality of that evidence.
- **The evidence produced** is capable of being tested in court or by publicly appointed regulators.



… and where are we finding AI / ML …



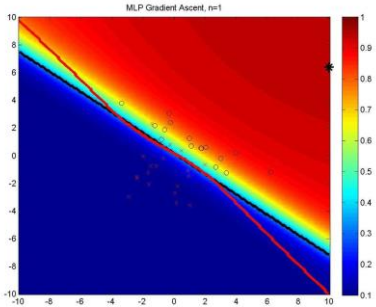
There are two main elements of admissibility: the physical element (the artefact) and the process (technical) by which the artefact has been handled.

'Digital forensics is meant to be based on science, not supposition'

One of the key tenets of any operational Cyber Resilient methodology must be that it should generate evidence in a style and of form that can be taken to court.

Cyber Resilience [The Goal] …

“A System is Resilient if, and only if, there is justifiable and enduring confidence that it will function as expected, when expected”



- **It is Secure if it displays this property in the face of an Adversary;**
- **It is Cyber Secure if it displays this property in the face of an Adversary that need not be co-located.**

Principles and observations that also apply to ML / AI …

- A fundamental requirement of digital system design should be its ability to be analysed
- All possible behaviours must be bound to quantify both safety and security

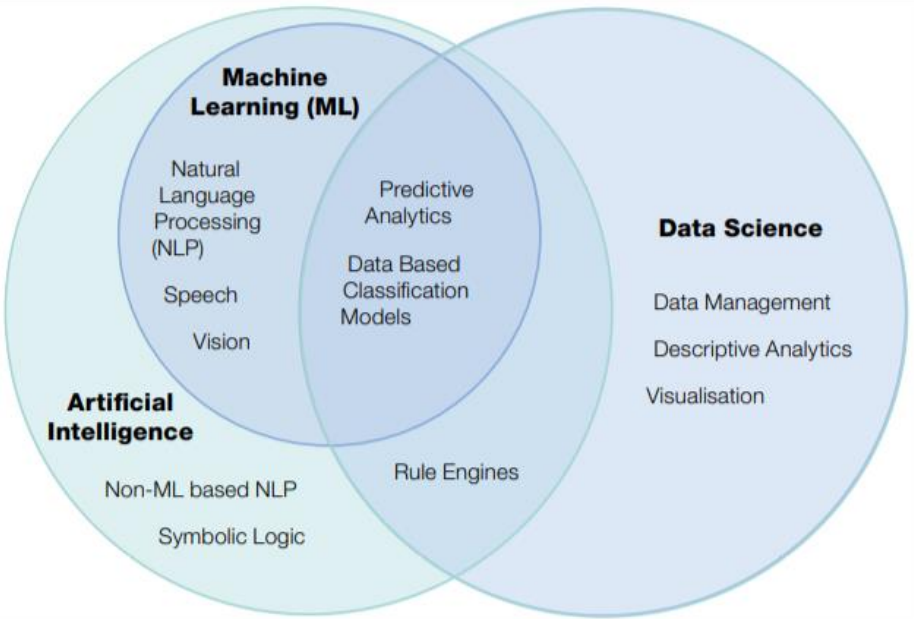
- Paradox of digital systems that they can be both unpredictable and deterministic
- Rice's theorem: no algorithm exists to predict a priori the behaviour of a generic information processing system. If the system was finite, but had an exponentially large number of states, then it is effectively undecidable – simulation or testing cannot determine all possible behaviours.
- Digital systems can be unpredictable. A 1-bit variation in a program, or the input to a program, can produce extremely different results. Digital systems possess a Lyapunov exponent of > 0.
- Adequacy of modelling – different levels of scientific enquiry can be applied to the security of digital systems but verification and validation is only as good as the model itself and analysis is bounded by the efficacy of the model.
- Hardware and software is created at huge cost recouped replication it many times over. Whilst useful from an interoperability perspective, this determinism is advantageous to a threat.

What is AI?

No universally accepted definition

In Thales, AI is defined as:

- *"AI: To make a machine do what humans do usi intelligent / cognitive capacities*
  - *Perceive – rich, complex and subtle information*
  - *Learn – within an environment*
  - *Abstract – to create new meanings*
  - *Reason – to plan and decide*
  - *Act – to achieve rational goals."*



Dstl definitions:

- AI – theories and techniques developed to allow comp systems to perform tasks normally requiring human or b intelligence.
- Machine learning (ML) – A field that aims to provide computer systems with the ability to learn and improve automatically without having to be explicitly programmed.

Dstl AI Biscuit Book –
The_Dstl_Biscuit_Book_WEB.pdf (publishing.service.gov.uk)

Successful and safe application of AI to real world problems

"AI systems will only fulfil their promise for society if they can be relied upon. This means that the role and task of the system must be properly formulated; that the system must be bug free, be based on properly representative data, and can cope with anomalies and data quality issues; and that its output is sufficiently accurate for the task."

Hand, D.J. and Khan, S., 2020. Validating and verifying AI systems. *Patterns*, *1*(3), p.100037.

- **Real world problems and applications are badly behaved in many ways.**
- **Humans aren't good at expressing their requirement**
- **Compiling code and knowing the theory is just the first step**
- **Standard safety techniques aren't sufficient due to the nature of AI solutions**

Simulation and Real Time V&V

- Taking existing results from simulating based on accident database and moving to ‘real time’
- Gaming and Model Based
- If we find a problem then how do we replicate and prove it fixed?
- Granularity / fidelity of the model
- Skills and access to resources

Simulation and Real Time V&V

Interim Conclusions:

- **Automatic selection of test cases**
- **Gaming and Model Based**
- **Deterministic Outcome From Game Theoretic**
- **Combinatorial Explosion from Physics**
  - Use of Supercomputers vs Cloud
  - Use of Sensors / Actuators as part of the simulation
- **Can we achieve a conclusion in most cases without needing to go to expensive resources**
- **Defining the Interface for Service**
- **Cost of maintaining required specialist technology services**

Interim Conclusions:

Skills

- Medical Robotics
- Adversarial Simulation

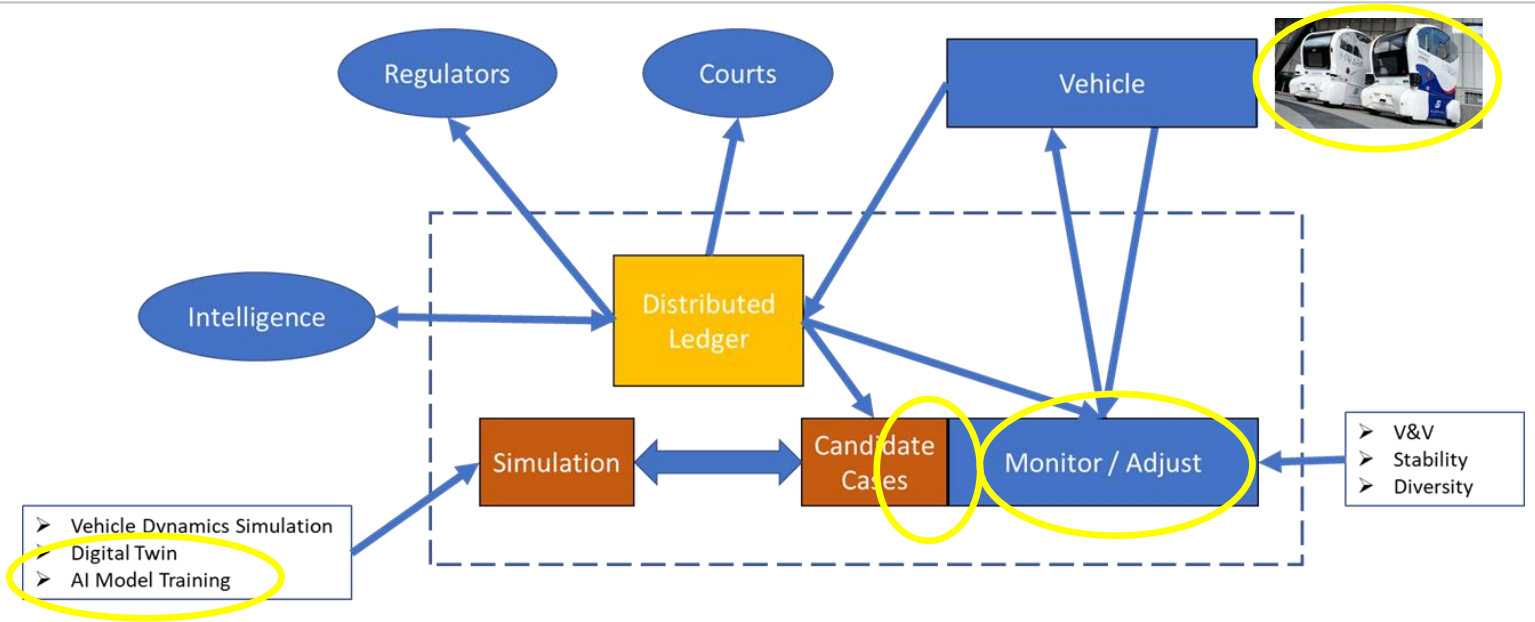
Tools

- The simulation frameworks e.g. unreal are not yet sufficiently flexible or granular (effort required)

Access to Capital

- Potential requirement for specialist compute might be difficult to capitalise.

What might a Cyber Resilient Solution Look Like …



There are two main elements of admissibility: the physical element (the artefact) and the process (technical) by which the artefact has been handled.

'Digital forensics is meant to be based on science, not supposition'

One of the key tenets of any operational Cyber Resilient methodology must be that it should generate evidence in a style and of form that can be taken to court.

What Should You Take Away …

- **Stop pretending.** We do not currently, or foreseeably, have engineering or scientific practices to guarantee to any legally satisfactory level the behaviour of large scale cyber-physical systems. Not now, not in the long term or in unforeseen circusmstances and certainly not in the face of a global cyber attack.
- Cyber Techniques that cannot scale increases liabilities
- CyRes is designed to take advantage of what we know using tools and techniques where they can achieve outcomes.

- AI correctly applied is a critical factor in achieving the kind of real time V&V necessary to maintain purposeful operation of complex systems.

FLiNT

Calls for global 'data extradition treaties' after Brighton road tragedy

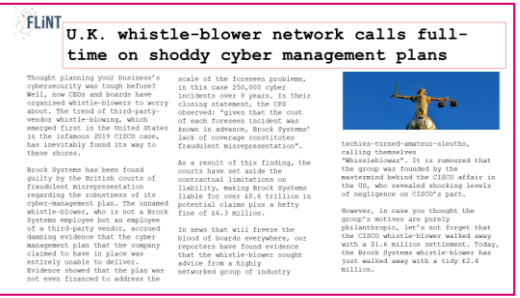
The United Road Transport Union is calling for new international data extradition treaties to be signed to help authorities get to the ground truth in road traffic incident investigations. The call comes after a young family was killed on the M23 just outside Brighton last month, in the latest road tragedy to be blamed on a cyber-security breakdown. Their hatchback was involved in a fatal collision when the Automated Lane Keeping Assist of a Dutch-owned HGV allegedly failed and the Ukrainian driver lost control of the vehicle. The HGV driver, who escaped the collision unhurt, was initially suspected of falling asleep at the wheel. But in a statement released through his union yesterday he insists that the Lane Assist feature "froze-up" in the minutes before the accident.

He claims that the feature was remotely "hacked" and that he had struggled desperately for a minute or more to recover control of the steering system, trying to take corrective action to avoid the fatal collision. With the support of his union, he is now calling for the vehicle software manufacturer XG to share user data that he believes would help prove the automated Lane Assist was compromised. He hopes these data will also provide evidence to show that he took prompt and appropriate action in his attempts to regain control of the vehicle.

XG have refused to allow access to such data, citing European privacy and security regulations. They deny any liability in the cluster of recent cases involving their automated driver assist products.

URTU spokesman William Johnson read a statement earlier today by the memorial flowers laid out at the scene of the accident

"This tragedy highlights the importance of sharing data across jurisdictions" he said. "If this accident had been caused by faulty brakes or tyres, then police could check the physical evidence. They need access to the systems data to investigate properly. This isn't about cyber security, it's about cyber safety." The URTU will ballot their members on strike action to support this case on Monday.



FLiNT

U.K. whistle-blower network calls full-time on shoddy cyber management plans

Thought planning your business's cybersecurity was tough before? Well, now CEOs and boards have organised whistle-blowers to worry about. The trend of third-party-vendor whistle-blowing, which emerged first in the United States in the infamous 2019 CISCO case, has inevitably found its way to these shores.

Brock Systems has been found guilty by the British courts of fraudulent misrepresentation regarding the robustness of its cyber-management plan. The unnamed whistle-blower, who is not a Brock Systems employee but an employee of a third-party vendor, accrued damning evidence that the cyber management plan that the company claimed to have in place was entirely unable to deliver. Evidence showed that the plan was not even financed to address the scale of the foreseen problems, in this case 250,000 cyber incidents over 8 years. In their closing statement, the CPS observed: "given that the cost of each foreseen incident was known in advance, Brock Systems' lack of coverage constitutes fraudulent misrepresentation".

As a result of this finding, the courts have set aside the contractual limitations on liability, making Brock Systems liable for over £6.6 trillion in potential claims plus a hefty fine of £4.3 million.

In news that will freeze the blood of boards everywhere, our reporters have found evidence that the whistle-blower sought advice from a highly networked group of industry

techies-turned-amateur-sleuths, calling themselves "Whistleblowers". It is rumoured that the group was founded by the mastermind behind the CISCO affair in the US, who revealed shocking levels of negligence on CISCO's part.

However, in case you thought the group's motives are purely philanthropic, let's not forget that the CISCO whistle-blower walked away with a $1.6 million settlement. Today, the Brock Systems whistle-blower has just walked away with a tidy £2.4 million.

Thank you

Peter Davies
Director Security Concepts

peter.davies@uk.thalesgroup.com