



Lloyd's Register
Foundation



UNIVERSITY
of York

Safety of Autonomy: Challenges & Strategies

PROFESSOR JOHN McDERMID, OBE FRENG

DIRECTOR: ASSURING AUTONOMY INTERNATIONAL
PROGRAMME

21ST MAY 2018



Lloyd's Register
Foundation



UNIVERSITY
of York

Outline

- Autonomy and Robotics
- A selective history
- Some challenges to safety orthodoxy
- Strategies for addressing the challenges
- The Assuring Autonomy International Programme
- Conclusions



Lloyd's Register
Foundation



UNIVERSITY
of York

Autonomy

- The Oxford English Dictionary says that autonomy is
 - The ability of a person to make his or her own decisions (or self-government, independence, or ...)
- Autonomous *systems* make decisions, not the humans; considerable current interest, but the idea is not new
 - Electric kettles that switch themselves off
 - Adaptive gearboxes in cars
 - Vacuum cleaners ...



Lloyd's Register
Foundation



UNIVERSITY
of York

Robotics & Autonomous Systems

- Can undertake tasks cost-effectively and provide benefits to society, e.g.
 - Avoiding boring, repetitious, dangerous and slow work in factories and warehouses
 - Supporting social care & independent living
 - Shipping, removing seafarers from harms way
 - Driving, reducing accidents due to human error





Lloyd's Register
Foundation



UNIVERSITY
of York

Outline

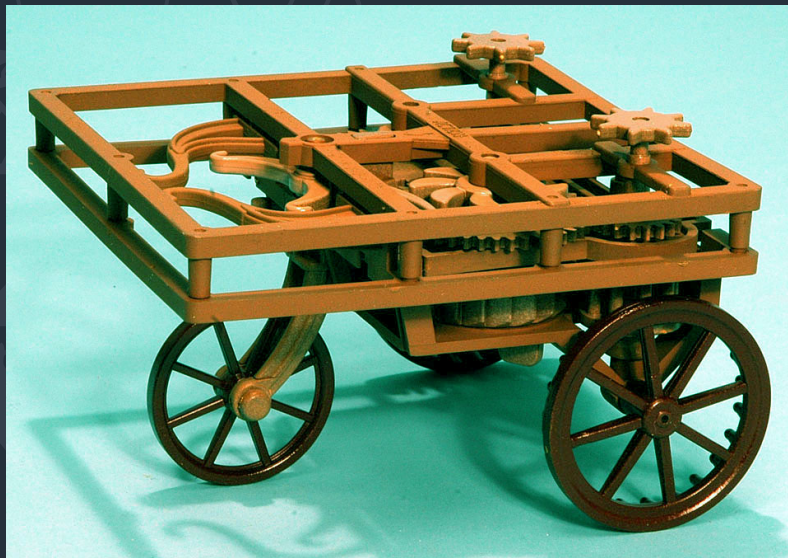
- Autonomy and Robotics
- A selective history
- Some challenges to safety orthodoxy
- Strategies for addressing the challenges
- The Assuring Autonomy International Programme
- Conclusions



Lloyd's Register
Foundation



UNIVERSITY
of York



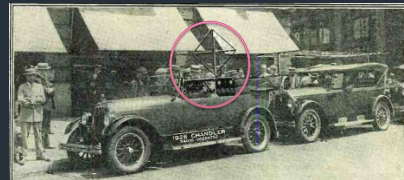


Lloyd's Register
Foundation



UNIVERSITY
of York

- Radio controlled passenger car
 - Used on the the streets in New York in 1925
 - And later in other cities, e.g. Fredericksburg in 1932
- On board technologies from around the 1940s
 - But we still use remote control in some applications



FREDERICKSBURG, VA., SATURDAY, JUNE 18, 1932.

"Phantom Auto" to Be Operated Here

Driver-less Car to be Demonstrated
About City Streets Next Saturday — Controlled Entirely by Radio.



Lloyd's Register
Foundation




UNIVERSITY
of York

- Industrial robotics exclude humans from work area
 - Simple interlocks, e.g. power is isolated if the gate is opened so the robot movement halts
- Principles well understood
 - Many encoded in standards or EU Directives
- With collaborative robots
 - Need to manage interaction in a more subtle fashion






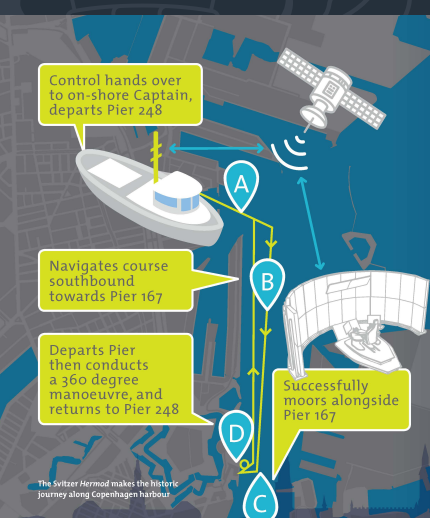




Lloyd's Register Foundation



UNIVERSITY of York




The diagram shows a map of Copenhagen harbour with a ship's path marked by points A, B, C, and D. Point A is at Pier 248, Point B is at Pier 167, Point C is at Pier 167, and Point D is at Pier 248. The ship's path is indicated by a yellow line with arrows. A satellite is shown in the sky, connected to the ship by a signal line. A callout box at the bottom left states: "The Svitzer Hermod makes the historic journey along Copenhagen harbour."

The world's first remote control commercial vessel

Key facts

- Rolls-Royce and Svitzer demonstrate the world's first remote controlled commercial vessel
- Test took place in Copenhagen harbour
- The 28 metre Svitzer Hermod was controlled by a Captain from shore
- It successfully demonstrated vessel navigation, situational awareness, remote control and communications systems
- Rolls-Royce Remote Operations Centre features state-of-the-art control
- Combination of Radar, Lidar and camera technology ensures Captain's awareness of surroundings

The tech	The test	The vessel
On board sensors to give Captain full awareness of surroundings Sensors covering Radar, Lidar, camera and audio State-of-the-art Remote Operations Centre on shore Rolls-Royce Dynamic Positioning systems control position of the vessel via satellite	400+ individual validations met 42 individual safety requirements met Passed 61 mandatory cyber security tests Completed 16 hours of remote control operation and overseen by Lloyd's Register	28 metre tug Svitzer Hermod Built in 2016 2 x MTU 16V4000 M63 diesel engines



Rolls-Royce



Lloyd's Register
Foundation



UNIVERSITY
of York

Outline

- Autonomy and Robotics
- A selective history
- Some challenges to safety orthodoxy
- Strategies for addressing the challenges
- The Assuring Autonomy International Programme
- Conclusions



Lloyd's Register
Foundation



UNIVERSITY
of York

Safety Challenges

- Autonomy
- Learning and adaptation
- Systems of Systems (SoS)
- Safety of the Intended Function (SOTIF)
- Human interaction
- ...





Lloyd's Register
Foundation



UNIVERSITY
of York

Autonomous Systems & Assurance

- If decisions are made by a computer
 - How do we know they are good, appropriate ...?
 - Are they “ethical”?
- Modern autonomous systems
 - Use machine learning (artificial intelligence)
 - How do we know what has been learnt is “right”?
- How do we give confidence to users, the public, regulators ... that the systems are safe to operate?

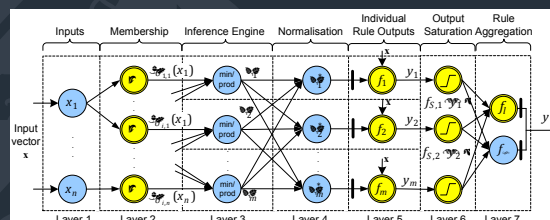


Lloyd's Register
Foundation

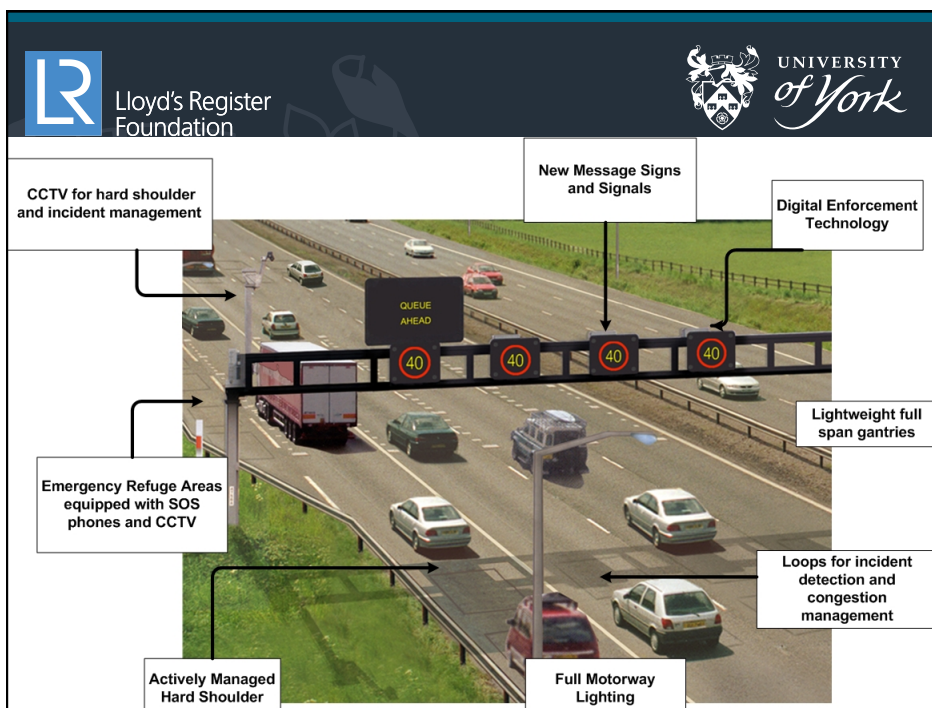
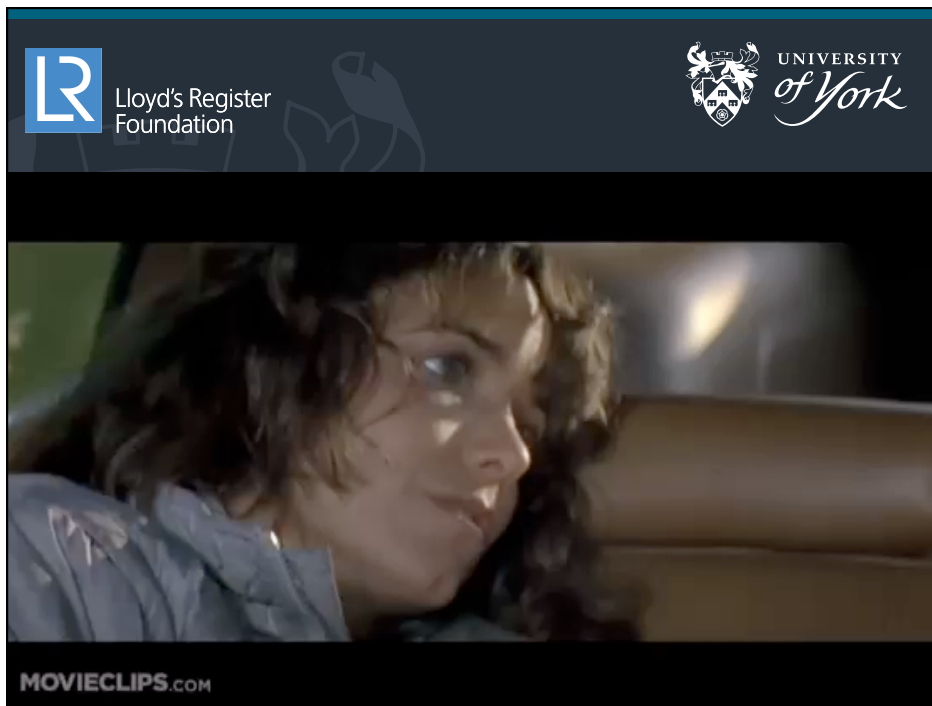


UNIVERSITY
of York

- If systems learn how can they be assured?
 - Common problem that what is learnt is invisible, e.g. weights in a neural network



- Tesla cars learn from others – a system of systems
 - Is what is learnt appropriate?





Lloyd's Register
Foundation



UNIVERSITY
of York

- Used classical techniques, e.g. FMEAs on the signs
- Novelty was analysing rule sets (and transitions)
 - Transitions are hazardous, and aim was to minimise risk
- Early years of operation accident rate halved



Lloyd's Register
Foundation



UNIVERSITY
of York

Is it safe in all situations?
Does it interact safely with other functions?
Does it interact safely with other systems?



Lloyd's Register
Foundation



UNIVERSITY
of York

- If “totally autonomous” human interaction is limited
 - Probably biggest concern is (public) acceptance
- In many situations, there is shared autonomy
 - System and human have distinct decision responsibilities
 - Role of humans and systems changes, e.g. for automated highway driving, when reverting to driver control
 - In emergencies
 - On leaving the highway
- Also issue of social cognition (NB collaborative robots)
 - How do humans perceive and then predict the behaviour of autonomous system (and vice versa)?



Lloyd's Register
Foundation



UNIVERSITY
of York





Lloyd's Register
Foundation



UNIVERSITY
of York

Why is Assurance Hard (1)?

- Assurance means
 - Confidence or certainty in a system's abilities
- Can we test it? Consider autonomous cars
 - Currently ~3.7 million miles between fatalities in the West (with drivers)
 - Systems such as Oxbotica's have done 10,000 miles
- Can we prove it?
 - Often what they have learnt is invisible
- How do we know that we have covered all scenarios?



Lloyd's Register
Foundation



UNIVERSITY
of York

Why is Assurance Hard (2)?

- Most cases are not fully autonomous
 - How do autonomous and “human operated” systems “understand each other”?
 - How do we manage relationship (e.g. a “handover”) between system and human
 - How can we be sure that the human has sufficient awareness of the situation?
- We also have to address how people judge and predict the behaviour of others (people, machines)



Lloyd's Register
Foundation



UNIVERSITY
of York

Outline

- Autonomy and Robotics
- A selective history
- Some challenges to safety orthodoxy
- Strategies for addressing the challenges
- The Assuring Autonomy International Programme
- Conclusions

Already shown one strategy:
the M42 System of Systems



Lloyd's Register
Foundation



UNIVERSITY
of York

Strategies

- Safety of learning
 - Expose the learning to enable analysis
 - More dynamic approach to risk management
- Safety of Intended Function (SOTIF)
- Risk management
 - Risk and benefit trade-offs
- Specific issues
 - Image analysis as an example

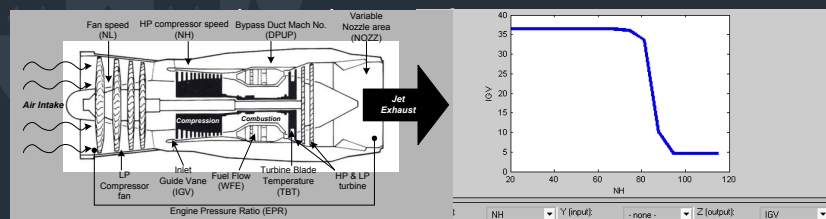


Lloyd's Register
Foundation



UNIVERSITY
of York

- York developed SCANN
 - A variant of neural networks that “exposed” learnt behaviour – weights for the network nodes
 - This made it possible to reason about safety
 - Applications, e.g. inlet guide vanes for aircraft engines
- General approach, can be used on other technologies

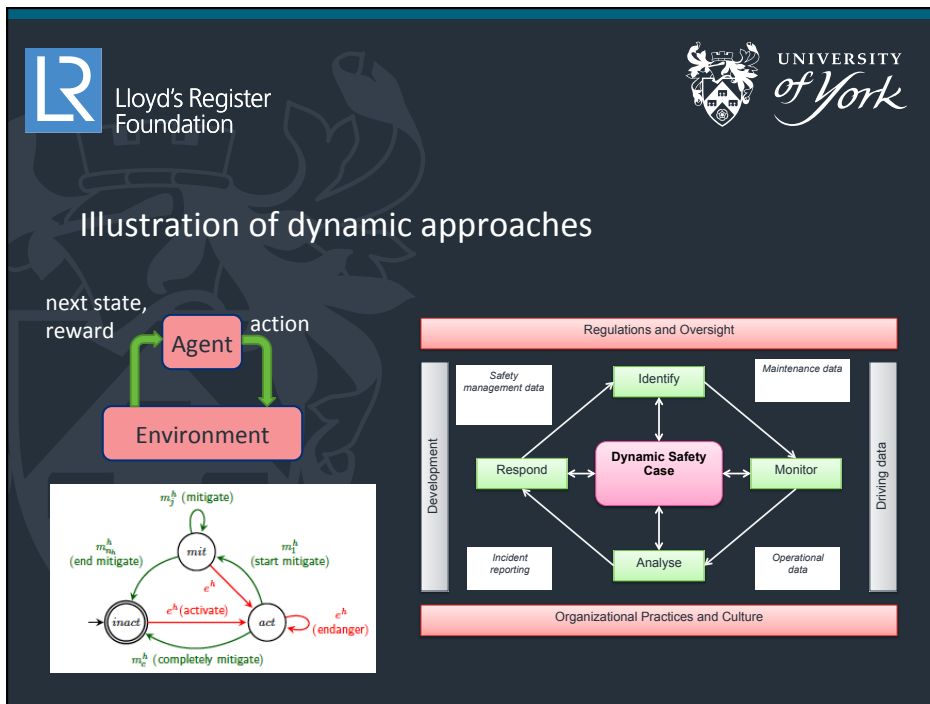


Lloyd's Register
Foundation



UNIVERSITY
of York

- Alternative approaches to safety of learning
 - Expose what is learnt, e.g. SCANN
 - Note requirement for explanations of AI under GDPR
 - Constrain learning
 - For example, objective functions for reinforcement learning that include safety goals
 - Update understanding of risk dynamically
 - Risk aware system design
 - Dynamic safety/assurance case



Lloyd's Register Foundation **UNIVERSITY of York**

- Emergency brake function (SOTIF)
 - Stops vehicle in the event of obstacle detection
 - Stopping may lead to a rear end collision (hazard)
- Should assess as part of hazard analysis
 - Also need to consider function interactions – with own system and others in SoS (adapt Functional Hazard Analysis (FHA))

Diagram illustrating a vehicle with emergency brake lights activated, showing a hazard warning triangle symbol on the rear.

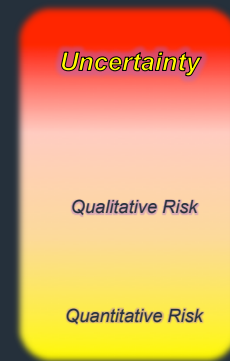


Lloyd's Register
Foundation



UNIVERSITY
of York

- For risk management we assume
 - Can quantify failure rates, or
 - Can assess risk qualitatively (rank via hazard risk index)
- Autonomy introduces uncertainty (means uncertainty can't be ignored)
 - Need to have a process that manages uncertainty, e.g. progressive usage
- Autonomy brings benefits
 - Need to balance risks and benefits



Lloyd's Register
Foundation



UNIVERSITY
of York

- DRABTO – Dynamic Risk and Benefit Trade-Off
 - Risk of malfunction balanced against benefits of normal functions including autonomy
 - Dynamic, as risk varies over time (minutes, hours ...)
 - Must influence design – risk as a run-time concept
 - Potentially draws together work on dynamic risk and dynamic safety cases ...
- More speculative
 - But might be the “paradigm shift” needed to get public acceptance as well as technical solutions?



Lloyd's Register
Foundation



UNIVERSITY
of York

- Several specific problems
 - Image understanding is critical
 - Sensing and analysis capability very difficult
 - Subject to false positives and false negatives
- May be benefits in adapting principles from reuse (3Cs)
 - Concept
 - Context
 - Content

Augmented aircraft landing system – only had to find the runway centre line ...



Lloyd's Register
Foundation



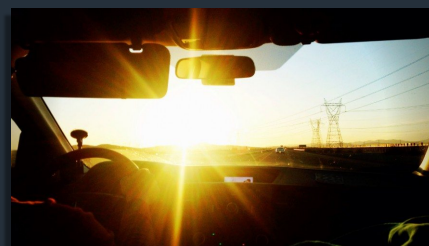
UNIVERSITY
of York



Rain and snow can affect **radar** performance



A heavy woolen coat can affect the performance of **ultrasonic sensors**



Rising sun in the front of the vehicle can affect the performance of a **video camera**



Lloyd's Register
Foundation



UNIVERSITY
of York

Outline

- Autonomy and Robotics
- A selective history
- Some challenges to safety orthodoxy
- Strategies for addressing the challenges
- The Assuring Autonomy International Programme
- Conclusions



Lloyd's Register
Foundation



UNIVERSITY
of York

Assuring Autonomy (1)

- The Programme is funded by the Lloyd's Register Foundation and the University of York motivated by the Foundation's 2016 Review of RAS
 - A gap in assurance and regulation of RAS
 - Assurance and regulation needs to “catch up with” and positively influence technology development
- Programme for five years from 1st January 2018
 - Programme led and directed from York, with worldwide collaborators



Lloyd's Register
Foundation



UNIVERSITY
of York

Assuring Autonomy (2)

- Four main strands of work
 - Work on assurance and regulation in real-world prototypes
 - Basic research, e.g. on dynamic risk and assurance of AI
 - Education and training for professionals in RAS & safety
 - Support to the international community
- International perspective is important as many problems are global & need harmonised regulation



Lloyd's Register
Foundation



UNIVERSITY
of York

Assuring Autonomy (3)

- Status
 - First demonstrators being set up, with projects in the UK, Italy and Sweden, Japan likely to follow soon
 - Healthcare, manufacturing, quarrying, automotive
 - Basic research starting, including work to develop a widely accessible, expert Body of Knowledge (BoK)
 - Wiki-like repository, on safety and assurance of RAS
 - Education and training starting soon, first understanding what's needed, then developing teaching materials
 - Linking with other industry and academic projects



Lloyd's Register
Foundation



UNIVERSITY
of York

Outline

- Autonomy and Robotics
- A selective history
- Some challenges to safety orthodoxy
- Strategies for addressing the challenges
- The Assuring Autonomy International Programme
- Conclusions



Lloyd's Register
Foundation



UNIVERSITY
of York

Conclusions

- Potential benefits from RAS in many domains
 - Taking people “out of harm’s way”, providing effective care, eliminating/reducing sources of (human) error ...
- Safety, assurance and regulation a key focus
 - Some companies are well aware of the issues, but a significant gap in technical capability and competence in some areas – including in regulation
- Assuring Autonomy International Programme gives an opportunity to address those problems



Lloyd's Register
Foundation



UNIVERSITY
of York



Professor John McDermid OBE FREng
Director: Lloyd's Register Foundation ,
Assuring Autonomy International Programme
Department of Computer Science
University of York
Deramore Lane,
York YO10 5GH

www.york.ac.uk/assuring-autonomy



@York_RAS / @LR_Foundation

E-mail: assuring-autonomy@york.ac.uk