TAS in the Field:
Studying governance challenges in high-risk autonomous systems

Mixed Reality Laboratory, School of Computer Science, University of Nottingham

Glenn McGarry,
Andy Crabtree

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

mixed reality lab

Scope of the Studies

Our ongoing studies investigate and elaborate governance challenges that confront current practices surrounding autonomous systems.

We position our findings so far against future governance (e.g. EU Artificial Intelligence Act) and aim to inform discourse in this area.

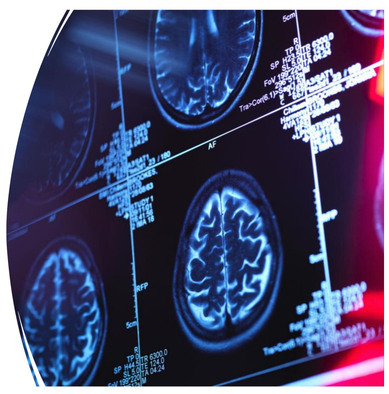
Method

We take a qualitative social science approach based on the 'ethnographic' tradition of observational studies in HCI.

Current studies limited to informal interview with practitioners working in the field of high-risk autonomous systems.

1 High-risk Medical Devices

This study explores the work of high-risk medical device developers. Designing autonomous systems for this tightly regulated market – with its well-defined stakeholder relationships and risk-based processes - is highly challenging.



Key Observations

Risk is accounted for at every step

"Risk assessment and mitigation has to be built into your processes. It goes into the project planning and needs to be recorded into a system that is trusted by the regulator so that they can check that all of these steps have been done."

AI's framing is problematic for medical device regulation (MDR)

"The machine learning community has come in from the side and said "hey, I think our system can actually replace the doctor"; which I can tell you was not the right way to enter the market."

Context is an overlooked variable

"In a big clinical trial they put automation in two different places in the workflow and got completely different health economics, sensitivities and specificities out of it. The system now behaves differently, and it's really complicated to understand what is the risk? What is the impact of the error of your system?

Software as a medical device (SaMD) is a risky concept

There are small companies that are building AI for medical devices who say, "I'm going to build this machine learning and it (will) just run everywhere". But if the hardware is made slightly faster, slightly higher resolution, what happens to the AI then, is it still performing the same way? How do you know?

Good data science will be crucial for prospective medical AI

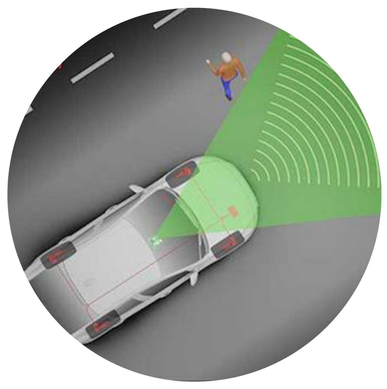
"Understanding what goes in the quality of the input data is important, because normally you make a medical device and then you test it on people, (with) machine learning, you take the data from a medical device, learn, and then stick it back in. So perhaps in the future there may be a company that has the expertise to verify if somebody has made machine learning properly."

Engineering and Physical Sciences Research Council

UKRI Trustworthy Autonomous Systems Hub

2 'Trust' in Automated Braking systems

This study explores the work of a consultant to the automotive industry involving the cyber resilience of automated emergency braking systems (AEB).



Key Observations

Every sensor is augmented by ML (whether you know it or not)

"(There are) 18-23 sensors in a braking system. We poisoned the machine learning in the sensor, so it would not report accurately. The safety engineer responsible said "There is no machine learning on my vehicle." but actually there is."

Road vehicles are evolving into a system of systems

"The complexity of connections way exceeds anything that anybody is analysing. Control systems are often in conflict: I've got privacy obligations in relation to my telemetry box, but if you encrypt the data that I need to brake then all I have to do is get rid of the (security) key and you can't brake."

You can't rely on algorithms to manage problems

"If I stimulated the machine learning so that the control system said, 'We're going to have to stop the car' that's a sensible thing to do. However, if over an hour that happened to 40 million vehicles, you're not moving food, we haven't got any medicine; restarting a system of that scale isn't a very easy prospect."

You don't even need to get into the IT to hack a vehicle's sensors

"All I have to do is shine bright lights at the sensors in order to overwhelm them and it actually makes a mistake about what's happening … I don't even have to get into the IT … for every vehicle that's got that system, I could turn the brakes on and all I've got to do is flash the headlights."

Digital models cannot represent the real world

"It is impossible for a digital system to be a complete and accurate representation of your analogue real world system … all I have to do as an attacker is find a gap. You've got to understand your system will change, it spends most of its time in operation not in design … understanding how that all fits together … and being able to respond is fundamental to doing anything that's trustworthy."

Governance Implications (EU AI Act)

Study 1

- Redundancy in relation to most areas of MDR
- The sector is not prepared for scrutiny of data driven systems (Annex VII)
- Implies systematic roles in verifying data practices

Study 2

- The key finding conceives of "trust" as safe and legally defensible with clear requirements to:
  - Understand complex systems-of-systems
  - Do thorough adversarial testing
  - Identify risk of harm
- Implied in the proposed EU AI act are systematic roles:
  - In risk management system (Article 9)
  - In post-market monitoring (Article 61)

Next Steps

- Further mapping of the medical device domain: findings to follow from interviews with regulators.
- Completed findings will feed into a paper with more detailed governance implications.

BAE SYSTEMS · ACFR Australian Centre for Field Robotics · SECURE · RESCUE GLOBAL Crisis and Disaster Response · University of Southampton