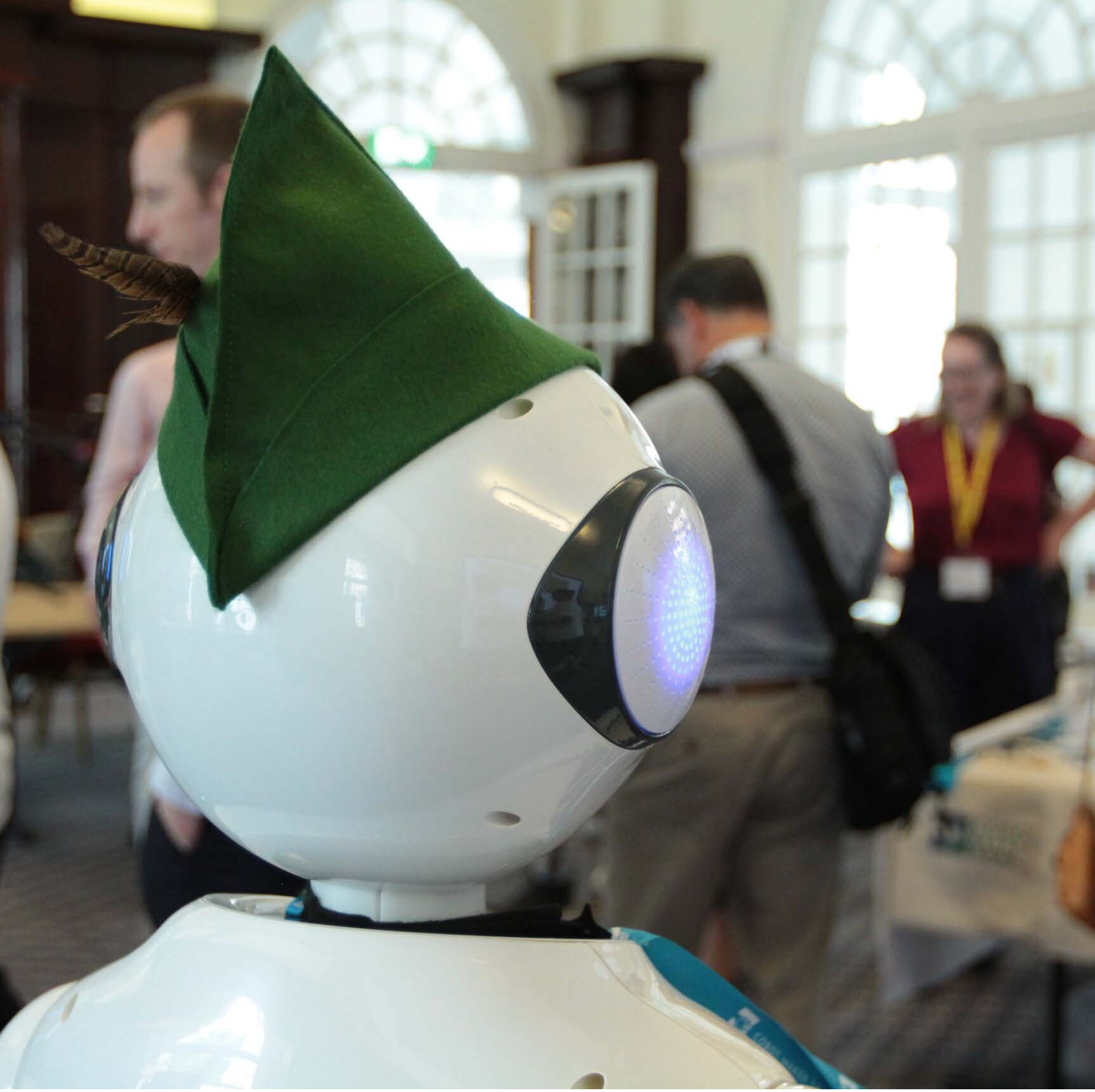


UKRI Trustworthy Autonomous Systems Hub

GOVERNANCE AND REGULATION

REGULATING THE ROBOTS – THE CHALLENGES OF GOVERNING EVER-SMARTER TECHNOLOGY

As Autonomous Systems (AS) and Artificial Intelligence (AI) take significant steps forward in capabilities, a number of challenging questions emerge. How do we ensure that our intelligent machines continually meet certain standards? How can the law keep pace with machine learning and its diverse applications? Where does human input end – and who or what is responsible if something goes wrong?

These complex issues are among those being addressed by the UKRI Trustworthy Autonomous Systems (TAS) Programme – a £33m multi-disciplinary research programme funded as part of the Strategic Priorities Fund. Governance and Regulation makes up one of the six TAS Nodes – focused research projects examining individual aspects of trust in autonomous systems. It was also the first topic to be examined in a series of multi-disciplinary workshops, exploring the complexities and looking at what the future may hold.

The world of AS and AI is fast-moving and difficult to govern. How, then, do we begin assessing what governance and regulation is needed? And once that is in place, how do we enforce and monitor it?

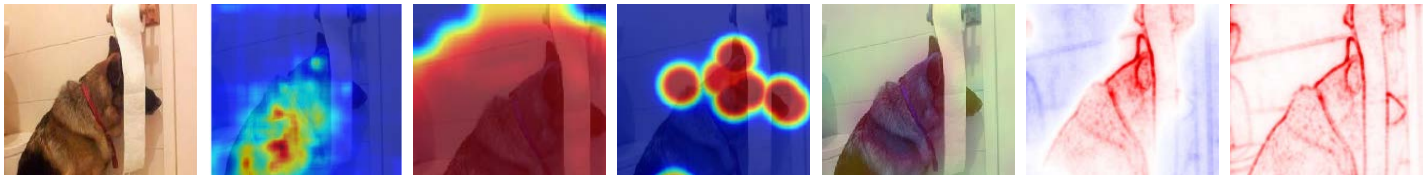
A range of methodologies and early results are being examined by the TAS programme, broadly within the parameters of four main areas of research: legal and social implications; machine learning, data and ethics; formal testing methods and implications in the design process. The aim is to create a Responsibility Framework which can inform and assist the regulators, backed up by real-world examples.

How to govern an AI world

Designing and deploying autonomous systems requires a multi-pronged approach. The use of standards in the design phase requires agreement on what those standards are. Once the systems are in use, new issues arise: human input is unpredictable, as is human interpretation of information. Where does our input end and machine learning begin? We need the real-world data from the autonomous systems in action, but how do we interpret this, and how can we use that data to help shape future governance?

TAS researchers have been working with key partners on a range of studies in different fields to examine this. Medical devices and AI analysis are one of the most fast-growing and important applications of TAS and Governance and Regulation research.

One example is tumour detection. Machine learning can be used to review MRI images of the brain, looking for patterns that might indicate tumour growth. Evaluation of this information, combined with other methods such as heatmaps, has the capacity to greatly advance technological capabilities and is an exciting glimpse into the future of medical testing.

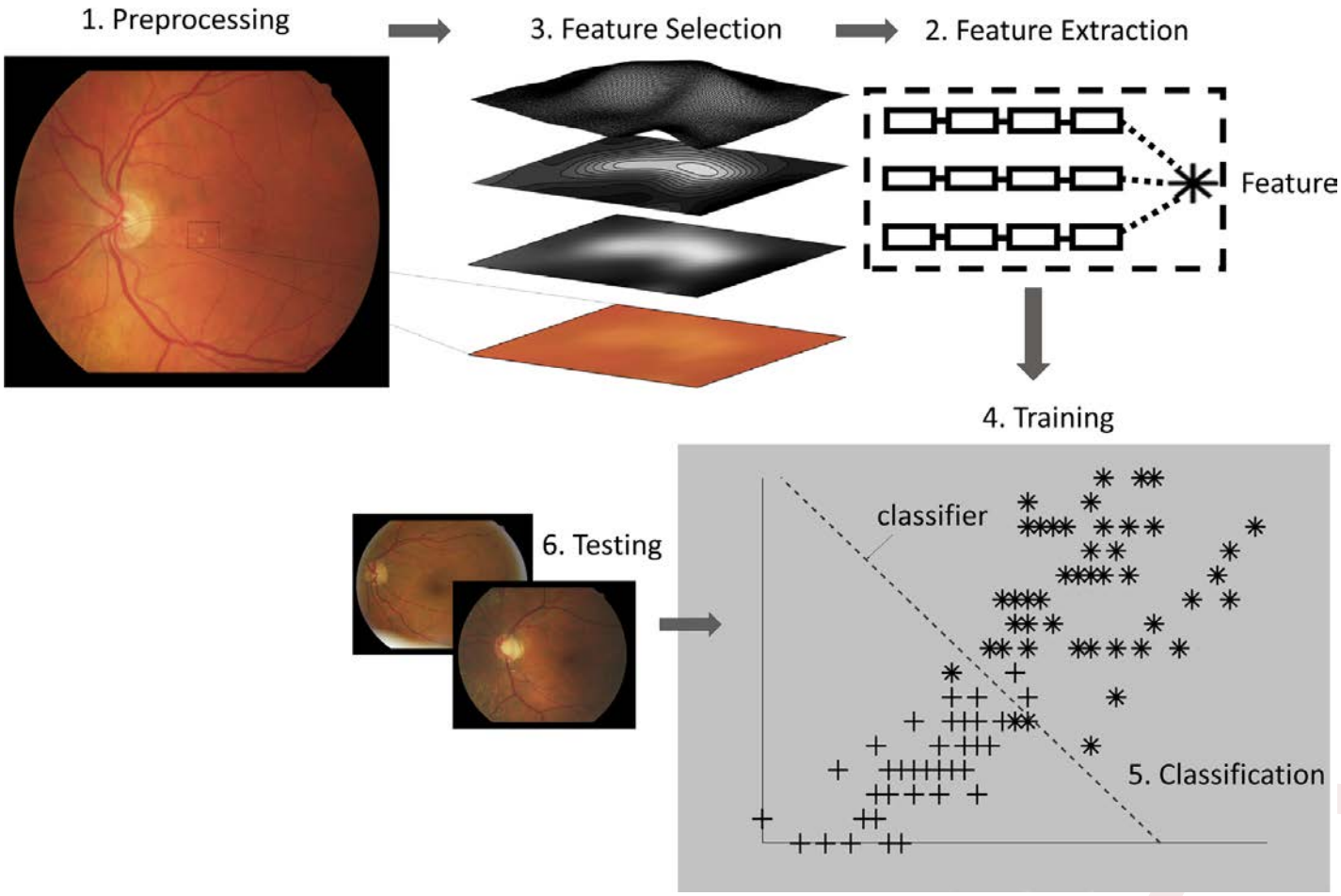


(a) Input (b) DC-causal (c) DC-SBFL (d) Extremal (e) RISE (f) RAP (g) LRP

Explaining the classification of a partially occluded image of a German Shepard dog (a), DC-Causal (b) performs best of the classification explanation tools (from Chockler et al., 2021)

Elsewhere, researchers are in contact with NHS Scotland in relation to the screening process for Human Papilloma Virus (HPV), a key indicator for development of cervical cancer. Autonomous systems and devices are also being used to help research colorectal cancer and explain the gut microbiome.

In the field of optometry, researchers have been looking into how AI can help mitigate human errors and improve diagnosis of eye disease, including early age-related macular degeneration. The use of machine learning works particularly well in helping understand decisions and errors in human-agent teams. As Dr Hana Chockler of King's College London says, this is exciting work in progress: "We're not yet treating patients, but the autonomous system and the machine learning components help in research, and their prediction seems to be a very large part of causality."



Schematic of a supervised machine learning pipeline for age-related macular degeneration in Optometry (from Pead et al., 2019)

The world of transport poses new questions regarding responsibility and regulation. How can AI account for and adapt to discrepancies in human behaviour and unpredictable real-world scenarios? For autonomous vehicles (AVs) to function correctly and compliantly, they must make crucial decisions on fast-changing data that comply with existing traffic laws. Such situations, however, are not always clear-cut. Likewise, laws themselves contain many discrepancies, real-world complexities and contradictions and are widely open to human interpretation.

Prof Burkhard Schafer from the University of Edinburgh explains: “How can we or should we at all represent these things, such as a near miss? We let humans get away with it. We might not want to let machines get away with it because machines are a different type of accuracy for norms than a human driver as we make excuses for human drivers that we do not necessarily want to make for a machine.”

Another aspect of autonomous vehicles is that of licensing. If cars were fully autonomous, would you need a driving licence? Not needing a licence seems highly unlikely. Even in systems where automation has been achieved to a very high extent, such as trains, there is still typically a human operator ready to take over if necessary. It may be that we may need to move to a very different type of a driving licence – perhaps aligned with what currently exists with automatic and manual licences - and one that requires updating as new automated features are introduced.

1	Ensure that CAVs reduce physical harm to person.
2	Prevent unsafe use by inherently safe design.
3	Define clear standards for responsible open road testing.
4	Consider revision of traffic rules to promote safety of CAVs and investigate exceptions to non-compliance with existing rules by CAVs.
5	Redress inequalities in vulnerability among roads users.
6	Mange dilemmas by principles of risk distribution and shared ethical principles.
7	Safeguard informational privacy and informed consent.
8	Enable user choice, seek informed consent options and develop related best practice industry standards.
9	Develop measures to foster protection of individuals at group level.
10	Develop transparency strategies to inform users and pedestrians about data collection and associated rights.
11	Prevent discriminatory differential service provision.
12	Audit CAV algorithms
13	Identify and protect CAV relevant high-values datasets as public and open infrastructural resource.
14	Reduce opacity in algorithmic decisions.
15	Reduce opacity in algorithmic decisions.
16	Identify the obligations of different agents involved in CAVs.
17	Promote a culture of responsibility with respect to the obligations associated with CAVs.
18	Ensure accountability for the behavior of CAVs (duty to explain).
19	Promote a fair system for the attribution of moral and legal culpability for the behavior of CAVs.
20	Create fair and effective mechanism for granting compensation to victims of crashes or other accidents involving CAVs.

Twenty principles for the ethics of connected autonomous vehicles to support researchers, policymakers, manufacturers and deployers (from European Commission, 2020)

MITIGATING RISK AND FAILURE

Another area of complexity arises from when things go wrong. Autonomous systems force us to examine how we might restore trust in a system after there's been a problem or a failure. Where does the responsibility lie? Where does the apology come from if something goes wrong? Insuring against risk is complex and can have unintended consequences. If tough rules and stringent insurance regulations are imposed universally, for example, would this deter smaller companies from participating in developing the technology? And would imposing high insurance premiums price some innovators out the market, thereby limiting growth?

According to Prof Ramamoorthy from the University of Edinburgh: “We often think about insurance just in terms of safety; however, insurance plays an important role in the development of the sector. If you're a small company, trying to deploy a new product with these kinds of failure modes, there's simply no way unless there's some way of ensuring the company itself.”

Insurance companies are important in shaping regulation from the legal perspective. But to quantify risk, companies need data. Part of the insurance business model is to insure human skills; exactly how much risk can it take?

Prof Burkhard Schafer highlights the issue: “The moment you have an event that affects all [autonomous] cars because they're all using the same software and it fails at the same time. If that is your risk model, then our insurance companies are massively undercapitalized to cover that sort of thing.”

THE COMPLEXITIES OF INTERNATIONAL REGULATION

Cross-border legal questions surrounding AI are far from clear cut. Different countries have different approaches to governance and jurisdiction, so a single set of regulations is an impossibility. The EU have published a proposal for regulation on AI, which will affect most autonomous systems under development. It is being examined in depth by TAS who conclude that the idea of a multi-national regulatory approach is ambitious, but a good one. However, overall, it has significant shortcomings - in particular with potential dangers to the rule of law and democracy. Who can enforce the regulations? Does it give too much power to technology-setting bodies with little judicial control?

In a post-Brexit UK, we would also encounter some issues reminiscent of GDPR changes. The EU AI Act would have some overseas reach, and with a similar US initiative being announced, the UK risks falling between two large regulatory blocks.

These questions open the door for TAS researchers to work closer with standard-setting bodies to ensure that ethical and legal considerations are properly reflected in these standards.

Trust and the law

There are also challenges around how we create laws that are flexible enough to temporarily respond to an emergency without undermining trust in the technology and the legal system itself.

During the COVID pandemic, for example, there was a need for a rapid technological solution to a huge social problem – in essence, the rollout of the COVID-19 Track and Trace system. The laws on data permissions were temporarily overwritten to ensure the system could be implemented more quickly.

Prof Burkhard Schafer said the research drew some interesting conclusions: “We looked into responses to emergency and crisis and one of the findings is that some safeguards were unnecessary, especially in the UK context, where Parliament can't bind [devolved] Parliament and there is no constitutional court to necessarily set things back after the crisis is over.”

There are legislative techniques that can mitigate this, such as sunset clauses – provision that sets an expiry date when part of a law will cease to have effect. However, this poses technical challenges for the design of autonomous systems. For example, they may have to be designed with ‘contestability’ and ‘temporality’ as features, just in case legislators want to change the applicable law after they are deployed.

THE FUTURE OF AUTONOMOUS SYSTEMS

What next, then, for the regulation and governance of autonomous systems and AI? Some of the key findings of the Governance and Regulation workshop prove interesting reading.

One of the most notable conclusions relates to the difference between ideology and what is actually achievable in practice. According to Prof Ram Ramamoorthy: “There is a big gap between what we wish we could say about AS and what we can actually say now, and we may never achieve the security level of robustness alluded to.”

There is still much work needed in the areas of ethics, responsibility and machine learning. Do we need separate regulations for autonomous systems in isolation, as opposed to when humans and AS are both instrumental in their operation?

Dr Hana Chockler highlights the issue: “Physicians are getting acquainted with machine learning systems, showing them where the tumour is or what are the markers, but what if the physician is even more hands off and this is the norm. For example, directing the autonomous system to perform the surgery- a completely different strategy will be needed.”

In the end, it's all about mitigating risk, and the reality is that governance and regulation has to be carefully tailored and adapted as technology advances – and one size most definitely does not fit all.

REFERENCES

Chockler, H., Kroening, D., Sun, Y., 2021. Explanations for Occluded Images. in Proceedings of International Conference on Computer Vision (ICCV).

https://arxiv.org/pdf/2103.03622.pdf

Pead, E., Megaw, R. Cameron, J., Fleming, A., Dhillon, B., Trucco, E., MacGillivray, T., 2019. Automated detection of age-related macular degeneration in color fundus photography: a systematic review. Survey of Ophthalmology. Vol. 64, p. 498-551.

https://doi.org/10.1016/j.survophthal.2019.02.003

European Commission, Directorate-General for Research and Innovation, 2020. Ethics of connected and automated vehicles, Publications Office,

https://doi.org/10.1016/j.survophthal.2019.02.003

About the Trustworthy Autonomous Systems (TAS) Hub

The TAS Hub sits at the centre of the £33M Trustworthy Autonomous Systems Programme, funded by the UKRI Strategic Priorities Fund. Its role is to coordinate and work with six research nodes to establish a collaborative platform for the UK to enable the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice by individuals, society and government. For more information, please visit the website: https://www.tas.ac.uk/.