Towards human centred threat assessment

Security Node

Lancaster University & Cranfield University

Lancaster University: Corinne May-Chahal, Joe Deville, Catherine Easton, Luke Moffat

Cranfield University: Lisa Dorn, Anders Af Wåhlberg

"Those who are subject to manipulation, along with responsible governance actors, need technologies of humility the most to consider vulnerabilities, the framing and distribution of risks, and to learn together"

Monika Buscher et al (2022)

Engineering and Physical Sciences Research Council

UKRI Trustworthy Autonomous Systems Hub

Overview

Within the TAS-S Research Strand 3 (RS3) team, we have explored ways of bringing **insights from sociology, law, and psychology** to delivering Autonomous Systems (AS) that both deliver meaningful security and are ethically responsive.

Our research has aimed to highlight the value of bringing **an expanded understandings of both 'security' and 'safety'** to the design and deployment of AS, fully capable of taking account of the behavioural patterns and existential and social concerns that may be regarded as threats to AS within social systems.

A particular focus has been on understanding how questions of AS security relate to **Connected and Autonomous Vehicles** (CAVs), including collaborations with National Highways and research into the role of behavioural adaptation within AS security

- **How can threat assessment better integrate people's engagements with and expectations of AS?**
- **If we know that people adapt to AS over time, what does this mean for threat assessment?**
- **How can organisations themselves adapt securely to rapidly developing and changing AS?**
- **How can we move beyond ethics as 'tick box' in AS security design and deployment?**
- **Are standards and regulation socially responsive?**

In the context of the centrality of threat modelling to cybersecurity, we propose that **threat assessment needs to expand beyond modelling depending on quantifiable inputs** to incorporate a wider range of threats and other forces. Domains of relevance for AS threat assessment range from particular practices and types of expertise (law and ethics), to persistent patterns of human and organisational behaviour (experience, adaptation), to inputs coming from those that might be users of, or impacted by, AS (trust). The TAS Security node has found that:

Law and ethics

- Legal frameworks need to better respond to **how AS regulations and regulators are engaged with and understood socially**, pointing to an urgent need for **more agile and innovative forms of law-making**
- Industry alone is unlikely to have the critical capacities / tools to **set standards that are socially responsive**
- Participative contextual ethics requires **other ways of knowing AS design**

Trust

- Organisations shown to need **new tools for better assessing the trust of users and wider publics** in AS
- Survey of road users indicates **low confidence about safety of CAVs**, with a **lack of trust in businesses and government**
- This is particularly important given our research shows that **trust in AS heavily shaped by 'social trust'** in a makers' system and/or the regulators of AS technology

Experience

- **Behaviours in relation to AS shown to be highly dependent on experience**, however we have little or no inherited understanding of AS.
- More research required on how to construct **validated scales to measure experience with AS**
- But also need to make room for **diverse experiences with AS**, including recognising and even 'scaffolding' the right to refusal and to challenge the purported technological inevitability of AS

Adaptation

- **Siloing of knowledge within organisations** around AS risks inhibiting their ability to adapt to rapidly changing AS contexts
- An **urgent need for new professional roles, expertise and tools** to enable organisations to engage more critically with AS
- Evidence that **how users adapt to AS can compromise safety and security** demonstrates a pressing need for further research into the role of behavioural adaptation, including in relation to in-vehicle technology

Discussion: Implications for security, regulation and work of other TAS-S Research Strands (1 & 2)

Our work shows that **a wider set of factors must be taken into account when designing AS and assessing threats**. Approaches to threat assessment focusing solely on the interactions between a technology and a static user could miss the wider risks generated by the shifting behavioural and social environments of AS. This has **particular implications for the development of AS regulatory frameworks**.

These insights are also **being actively fed into the work of TAS-S Research Strands 1 and 2**, to further enhance both the security and response-ability of future AS design.

We are also working on **a practical toolkit for organisations**, to enable them to engage with AS design and deployment more critically and creatively, in the pursuit of a more socially response approach to AS security.

Cranfield University

Engineering and Physical Sciences Research Council

UKRI Trustworthy Autonomous Systems Hub

TAS-S

Lancaster University

This work is supported, in part, by the Engineering and Physical Sciences Research Council [grant number: EP/V026763/1]