Trustworthy Autonomous Systems Town Hall meeting

Summary of discussions

Background

UKRI is planning to invest up to £33 million in a multi-disciplinary research programme focussed on the Trustworthy Autonomous Systems (TAS). This funding is part of the Strategic Priorities Fund (SPF). It is anticipated that the TAS programme will be composed of a central Hub that will provide coordination and leadership, and several research focused nodes, each led by a multidisciplinary team and with a responsibility to undertake in-depth fundamental research into a key challenge within trustworthy autonomous systems.

Ahead of this significant investment, UKRI held a Town Hall meeting in London on 18 October 2019, as detailed here: https://epsrc.ukri.org/newsevents/events/ukri-trustworthy-autonomous-systems-programme-town-hall-meeting/

The Town Hall meeting aimed to:

- Share the context, importance and vision of the TAS programme
- Bring the research communities and industrial, government and third sector stakeholders with an interest in TAS, to foster new collaborations and relationships
- Explain how the TAS programme will be delivered, delivery timelines and the specific requirements of the SPF funding.

Please note that the Trustworthy Autonomous Systems Town Hall meeting was NOT a scoping workshop, and the objectives of the programme, its remit and the approach to delivery was determined prior to the event. Also, **please note** that it is not an expectation that the following points will be considered in the applications for the Hub or the nodes; applicants may or may not take these into account when preparing their proposals.

This document is a summary of the discussions held at the Town Hall meeting. Attendees were split in 8 groups to network/discuss on the different elements of the programme: the Hub, and the seven research nodes (trust, responsibility, resilience, security, functionality, verifiability and legality). The questions raised at the event have been collated into the FAQ document published alongside the Hub Outline call document https://epsrc.ukri.org/funding/calls/trustworthy-autonomous-systems-hub-outline-call/

Hub

The Hub will be central to the TAS programme, lead on both the **coordination** and **advocacy & engagement** objectives, building an interconnected research community, and proactively engaging across stakeholders. The vision and key aspects of the hub were discussed during this session:

- The Hub should be a recognised academic centre that engages with policy makers, coordinates conversations with the public and users and engages with national and international partners, supporting the inter-node collaboration and helping to promote the commercialisation of the node's outcomes.
- There is a need for the Hub to focus and be involved in its own fundamental research. Attendees discussed that some of the complementary research topics could include system integration and standardisation.

Research Nodes:

The research nodes will undertake fundamental, creative and multidisciplinary research in a number of areas of current un-met need. The nodes will be separate entities connected to the hub and they are expected to address one of the following research challenges in depth: trust, responsibility, resilience, security, functionality, verifiability and legality.

Trust

Discussions were focused on the following:

- The need for a trust framework, including how the paths of trust can be identified, the context that would enable identification of trust and distrust scenarios.
- Extensive human and autonomous systems interaction studies may be required to understand how the trust is developed. This approach would span across disciplines and user groups.
- The potential need to establish methods that would help measure trust and assess trustworthiness of the given system. This could potentially be achieved by enabling multi-level dynamic trust, where the autonomous system can adapt various trust expectations, as well as learn to become trustworthy based on experience.

Responsibility

This area requires an understanding of computational, legal and ethical boundaries for autonomous decision making. The discussion raised the following key points:

- There is a clear need to establish what it means for the machine to behave responsible and what its responsible use would look like.
- To ensure responsibility (especially if human is outside the loop), it may be necessary to identify and select appropriate values (i.e. what they are,

where they are from, is bias being imprinted) that are being applied onto the system. The key here is also how to ensure and check that the system is following the imprinted values.
- There may be a need for the meaningful and feasible levels of interpretability of autonomous systems to be established.

Resilience

To be resilient to undoubted failures, technically, socially and politically, systems must be designed to recognise and assess such changes and failures both internally and in their environment. The discussions held around resilience considered the following:

- A highly multidisciplinary team would be most suited for this task. A high level of interaction with other nodes is anticipated.
- Wide variety of approaches and tools, as well as establishing failsafe mechanisms would help promote resilience and ensure that functionality is maintained even after system failure.
- Ensuring resilience will involve, amongst others, monitoring the data quality and identify false data, identifying potential points of failure, as well as risk of unintended errors or failures.

Security

An autonomous system must be resilient to attacks at both software and hardware levels. The points discussed within the scope of this node included the following:

- A multidisciplinary team of experts is required to tackle challenges such as: cybersecurity, security by design (i.e. software and hardware levels), legislation, insurance and data protection (inputs and outputs) as well as social aspects of human involvement on both sides of the fence. It may be important to explore what security means to other disciplines.
- A culture change may be required when it comes to security approaches (e.g. from reactive to proactive). There is a need for a clear and approved security metric, that would help to establish if the system is secure, as well as ways of monitoring its behaviour during the use and reassure the users of continued system security.
- Unintended use of the system is a key threat to the overall security and may need to be explored and monitored.

Functionality

To address functionality, there is a need to understand first what the system is designed to do (i.e. what its autonomous functions are), what it does and what it might do, its performance as a function of time. Points discussed included:

- How and to what extent these functions adapt to the environment and how the level of automation can be dynamically adjusted for each function?
- A framework may be required to make decisions on the functions and their distributions (e.g. who does what, where, why), as well as necessary

mechanisms to be able to convey the risk, uncertainty in the functions and address them.

- Lifecycle of the technology should be taken into the account and planned for.

Verifiability

In order to deploy and utilise autonomous systems on a large scale, a clear understanding of suitable verification levels and strategy is required across all stages of its development (e.g. procurement, construction, implementation) and deployment (user).

- Development of common methodologies (e.g. certification, guidelines, benchmarking, toolkits) and their integration may be required.
- Cross-application verification may need to be addressed (e.g. shared autonomy).
- There is, potentially, a need to establish how to deal with changing environment and uncertainty.

Legality

In order to deploy and ensure adoption of autonomous systems they need to be legal and insurable, as well as regulated, with issues around liability and potential failure being resolved.

- A multidisciplinary team (including lawmakers) is crucial in order to be able to address the challenges of this node.
- There may be a need for creating a new accident investigation entity for vehicles, as well as new legal concepts that would help explore the implications of how systems evolve and function, including principles that need to change when human decision maker are removed, followed on by the ability of autonomous systems to learn to behave within the law.
- Certification processes may need to be established for the component parts and systems as a whole.