DELIVERING A SAFE AND INCLUSIVE AI FUTURE FOR THE UK

ASSURING AUTONOMY
INTERNATIONAL PROGRAMME

LR Foundation

SPOTLIGHT ON SAFETY

We are at a critical juncture in the safety discussion around society's adoption of AI. The Government has set out five key aims for the AI Safety Summit at the beginning of November. Additionally, the agenda for the summit calls directly on the scientific community to invest in research and bring solutions to safety challenges to the table.

At the Assuring Autonomy International Programme, we have been working with policy makers, industry and other research facilities to evaluate systems and develop methodologies which answer the big question: 'is it safe?'.

This document sets out three key approaches which will help the UK work towards answering that question and delivering the goal of safe and inclusive AI for the UK.

The Assuring Autonomy International Programme

The Assuring Autonomy International Programme at the University of York includes the UK's <u>foremost experts in safety-critical systems</u> and the development and deployment of safe AI.[1] We have over 35 years of experience researching safety critical systems, including safe machine learning, and safe autonomous systems. It's our position that maximising the potential benefits of AI in society, and doing so safely, centres around three main themes; evidence, community, and skills.

Here's what this should look like for policymakers:

1 Evidence - use AI safety cases to build public trust

Globally, three out of five people have issues trusting AI.[2] By stipulating the need for and benefit of AI safety cases[3] policymakers can foster greater public acceptance and trustworthiness of AI, whilst also providing organisations with a clear path to the safe deployment of AI-enabled systems. The case for implementing this recommendation is strengthened by the fact that <u>techniques already exist</u> and are being <u>actively applied by institutions</u> like NHS England, and developed into standards through The British Standards Institution. As leaders in the safety community, we've worked with many organisations to create safety and assurance cases which evidence and make the argument for the safe use of AI in sectors such as healthcare, road transport and maritime sectors.

Through this multidisciplinary work with industry we have identified that creating and maintaining safety cases is the most practical and realistic way to provide the required levels of confidence, transparency and trust to maximise the benefits of AI to society.

2 Community - harness the expertise of existing safety experts

The UK has a strong history and track record in safety. This is important because it means the UK is in the position to build in safety from the start by taking a whole systems approach[4] which can shape and inform safety culture around AI. And it's these timeless principles of safety which can catapult the UK forward to be a global leader in safe AI. Additionally, adopting this approach helps safeguard the UK from potential public concerns in the event of an AI-related incident on the global stage. We already have an entire community of safety professionals ready to work with teams developing and deploying AI systems.

By engaging with and securing the support of this existing safety community, policymakers gain ground in addressing some of the unique safety challenges around AI. For example, the inequitable distribution of risks and benefits across society, and the allocation of responsibility.
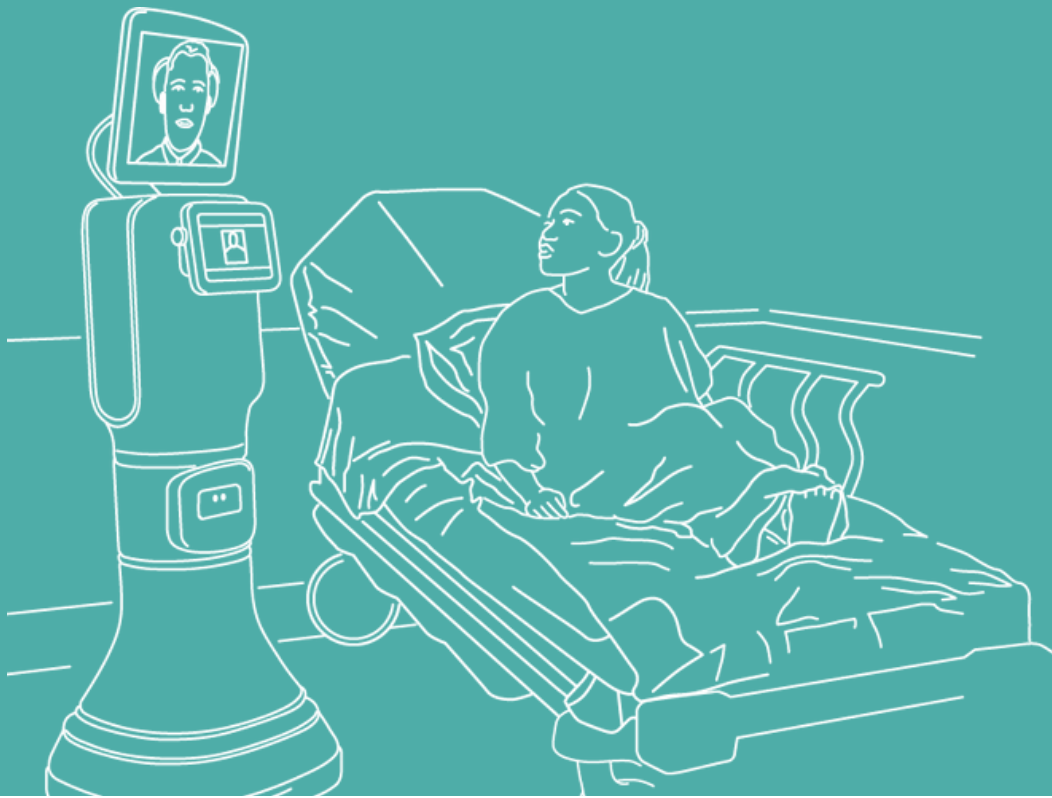
3 Skills - preparing people to work alongside AI

Ensuring people have the right skills to regulate, create, manage and use safe AI is vital for the UK's long-term global leadership of AI. Robust training, CPD and courses already exist to make this happen. The provision of training for industry, regulators and policy makers encompassing different sectors, knowledge levels, and organisational levels is central to the UK's decision making around the implementation of AI. It will enable us to ask the right questions at the right time from a position of innovation and credibility.

Innovation through safety

Whilst we recognise the urgency for progress and the need to establish the UK's position on the adoption of AI in the global market, we believe this cannot come at the risk of ignoring the potential for accidental or malicious harm from AI systems. By working proactively with established safety experts UK policymakers can champion safe innovation and effectively prepare the UK for the safe adoption of AI-enabled systems for the benefit of all society.

Assuring Autonomy International Programme
Institute for Safe Autonomy
University of York
York
YO10 5DD

Contact us:
Visit our website:
Follow us on LinkedIn
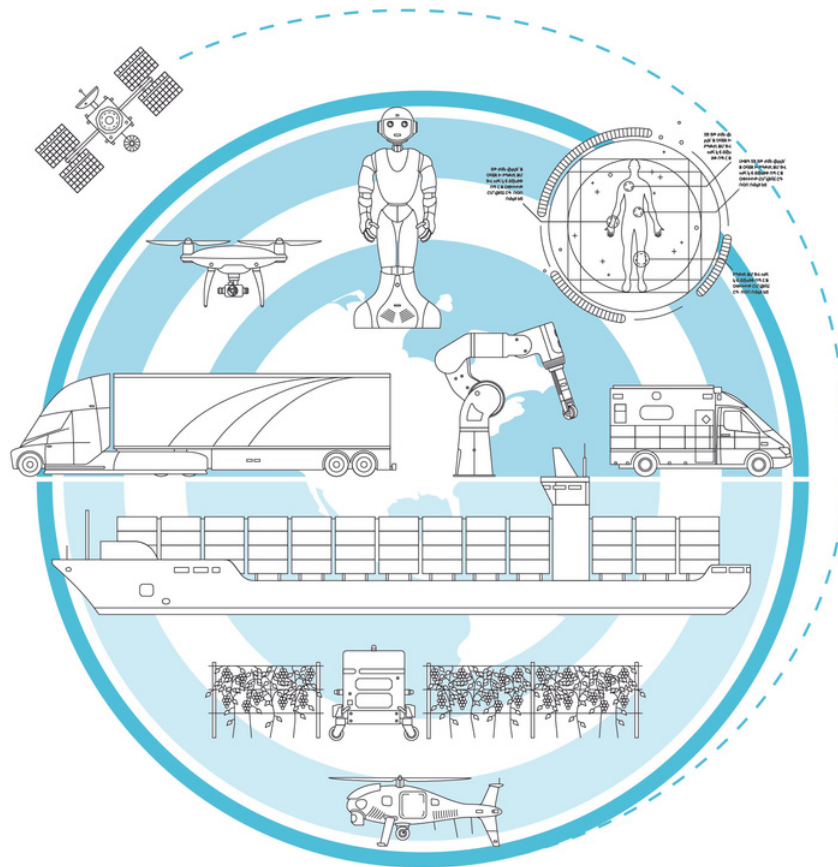
Notes:
1. The Assuring Autonomy International Programme (AAIP) is a £12m initiative funded by Lloyd's Register Foundation and the University of York established to address the global challenges around the safe development and deployment of AI, robotics and autonomous systems in society. We work with an international community of developers, regulators, researchers and others to ensure the public can benefit from the safe, assured and regulated introduction and adoption of RAS.
2. Statistic from Trust in Artificial Intelligence: Global Insights 2023
3. A safety case is a structured way of gaining confidence in a system's safety, and is something already mandated in many industries. It is based on combining evidence of risk and mitigation.
4. A whole systems approach is one that considers the legal, ethical, social and technical aspects in an integrated manner.