
Proceedings Paper:
Preston, Kate Elizabeth, Sujan, Mark Alexander and Habli, Ibrahim (2025) Assurance of AI and autonomous technology in complex environments: a human-centered perspective. In: Proceedings, Safety Critical Systems Symposium SSS'25. Safety Critical Systems Symposium, 04-06 Feb 2025, The Milner York. Safety Critical Systems Club, GBR.

Assurance of AI and autonomous technology in complex environments: a human-centered perspective

Kate Preston, Mark Sujan and Ibrahim Habli

University of York

York, UK

Abstract The development of AI and autonomous technologies is increasing rapidly across multiple sectors. The safety assurance of such systems often takes a narrow and technology-driven perspective, focusing on technical criteria such as overall performance metrics. However, these systems operate within broader human, technical, and organisational contexts. For example, an autonomous ship exists within a maritime system involving other functions and services including, among others, port and docking services, maintenance services, navigational services etc. Without consideration of these broader contexts, hazardous scenarios can arise, which may result from dysfunctional interactions between the autonomous functions and other elements of the maritime system. Effective safety assurance requires, therefore, consideration of the interactions between different elements of the work system, where the AI or autonomous system is just one such element. In this paper we put forward human-centered reflections based on an analysis of 22 demonstrator projects from diverse application domains.

1 Introduction

The development of AI and autonomous technology is increasing, with research highlighting the potential use and benefit across sectors such as healthcare, maritime and transportation (Sujan et al., 2022; Munim et al. 2020; Abduljabbar et al. 2019). The Centre for Assuring Autonomy (CfAA), previously the Assuring Autonomy International Programme (AAIP), was created to take a multi-disciplinary perspective to the safety assurance of AI, autonomous technology and robotics across different sectors. The AAIP supported 22 demonstrator projects across sectors, including work within manufacturing around how collaborative robots (Cobots) can support the sector and the importance of assessing the safety of human-robot collaboration (COBOTS demonstrator, 2022). Demonstrators have also been completed in

healthcare, one of which focused on the safety assurance of an existing AI platform piloted in Copenhagen and applied to the Welsh Ambulance Service (ASSIST demonstrator, 2023).

The CfAA has also developed several guidance documents to support the assurance of AI and autonomous technology. This includes the Assurance of Machine Learning in Autonomous Systems (AMLAS), a structured methodology for the creation of compelling and detailed safety cases for the machine learning component of any automated technology (Hawkins et al. 2021). Additionally, the Safety Assurance of Autonomous Systems in Complex Environments (SACE) framework was developed to work alongside AMLAS, providing a defined and detailed process for creating a safety case for autonomous technology (Hawkins et al. 2022). The methodology takes a holistic view of the autonomous system and its environment and provides a safety assurance process that leads to the creation of a safety argument and evidence. Research has also been completed on developing an argument pattern for AI and autonomous technology's ethical assurance. This ongoing work has developed an assurance case framework for communicating sufficient confidence in the overall ethical acceptability of AI and autonomous technology in their intended context (Porter et al. 2024). Understanding and researching the safety assurance of AI and autonomous technology has also been seen within the Safety Critical Systems Club (SCSC). For example, the SCSC Safety of Autonomous Systems Working Group has developed a guidance document on autonomous technology that should be managed throughout its lifecycle within safety contexts (SCSC publications. 2024). Further, the SCSC has organised seminars and symposiums and published articles focusing on developing safe AI systems (SCSC publications. 2024).

Despite the work done within the CfAA and the SCSC, there are still challenges to assuring AI and autonomous technology. These challenges include AI and autonomous technology experiencing a reduction in performance once integrated into practice; for example, a study focusing on the accuracy of machine learning versus clinicians for classifying skin lesions found that the AI tool accuracy was greater than a clinician when using data with similar qualities to the data the technology was trained on (Tschandl et al. 2019). However, performance decreased significantly once the AI was tested using images outside the training dataset. A further challenge relates to who will take responsibility for the AI or automated technology when integrated. The lack of clear responsibility can result in adverse events due to poor oversight of the technology's status, causing hazards to be missed (Porter et al., 2023). In addition, the introduction of autonomous technology may affect how other elements of the work system interact (Sujan et al, 2022). For example, the use of autonomous healthcare technology in intensive care could potentially affect how nurses and patients interact and remove nurses from the bedside, and the introduction of autonomous cargo vessels could lead to disrupting existing relationships between ship owners and cargo owners. These potential challenges may result from a limited understanding of the sociotechnical work system where the new AI or autonomous technology will be integrated and how the components within that work system will interact to create processes and outcomes (Salwei et al. 2022). For

a conceptual visualisation of how future AI or autonomous technology may interact with the other components within the work system, see Fig.1.

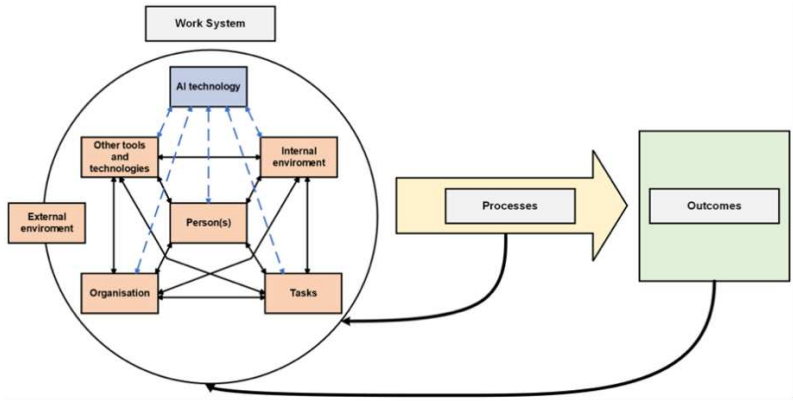


Fig. 1. Extended System Engineering Initiative for Patient Safety (Salwei et al. 2022)

To understand the sociotechnical work system, a human-centred approach can be taken throughout the development lifecycle and support AI and autonomous technology developers in considering the work system and the corresponding interactions. A human-centred approach is concerned with placing the humans at the centre and understanding the interactions between different elements of the work system rather than just focusing on any one element, such as AI, in isolation (see Fig. 1). For example, from this perspective it is important to consider how people would interact with the technology, for what purposes, under what contextual and environmental conditions, and how technology use is supported or hindered by organisational factors such as workload, staffing levels, competency and skill mix.

Accordingly, a human-centred assurance framework for AI and autonomous technology in complex environments will need to consider what kinds of processes and evidence are required to demonstrate that interactions among elements of the work system are such that overall performance is acceptably safe (or whatever the overall assurance goal is). This research aims to start the development of a human-centred assurance framework based on previous literature and demonstrator projects completed within the AAIP. The current paper discusses reflections from examining 22 AAIP demonstrator projects.

2 Methods

A demonstrator review was completed, assessing 22 projects completed within the AAIP that focused on developing AI-related, robotic, and autonomous technologies. Table 1 provides an overview of the demonstrator projects. The demonstrators were analysed to understand where human-centred approaches had been taken.

Additionally, a post-hoc assessment was completed to understand where a human-centred approach could have added value.

Table 1. Overview of AAIP demonstrator projects

Sector	Demonstrator title	Description of project (Verbatim in past tense)
Health and social care (n=7)	Safe robots for assisted living (ALMI Demonstrator)	The project demonstrated how novel robotic technology, environment monitoring capabilities, verification techniques, and adaptation methods could be integrated and applied to address concerns for autonomous robots used in people's homes.
	Machine learning in healthcare (SAFR Demonstrator)	This project helped to establish a safety assurance framework to support healthcare manufacturers and deploying organizations in assuring their ML-based healthcare technology and meeting their regulatory requirements.
	AI in ambulance response (ASSIST Demonstrator)	The project team adapted an existing Corti AI platform, which had been piloted in Copenhagen, for use within the Welsh Ambulance Service (WAST). The assurance activities contributed to the development of a real-world Body of Knowledge for assurance cases of AI in critical sectors.
	Safety of the AI clinician (Safety of the AI clinician Demonstrator)	This demonstrator project investigated how to assure the safety of an AI-based DSS for sepsis treatment in intensive care. Through this, it helped to establish general regulatory requirements for AI-based DSS.
	Assistive robots in healthcare (UWE Demonstrator)	The work focused on a series of experiments designed to validate a range of practical use cases derived from potential end-users, including occupational and physiotherapists, paid carers, regulators, and potential commercial manufacturing partners.
	Medication management (SAM Demonstrator)	The study focused on the clinical system rather than the technology itself, addressing safety assurance challenges at the intersection of engineering and human factors.
	Social credibility (Social credibility Demonstrator)	This small feasibility project was divided into two sections: introductory and experimental work. The introductory work (Menon, 2019) found that the social effects of assistive robots are often overlooked in hazard analysis. The experimental work was designed to validate the proposed link between social credibility and safety.
Automotive (n=4)	Shared control in autonomous driving (SafeSCAD Demonstrator)	The team developed a novel DNN-based framework that predicted driver takeover behavior (e.g., takeover reaction time) to ensure that a driver was able to safely take over control when engaged in non-driving tasks. They investigated formal analysis techniques for neural networks whose results could feed directly into the system-level design of autonomous systems and applied them to the DNN developed by the project to quantify its aleatory uncertainty.

Sector	Demonstrator title	Description of project (Verbatim in past tense)
	Automatic rating system for autonomous systems (ATM Demonstrator)	The team's mechanism exploited the observation that if trajectories that commonly caused catastrophic accidents were similar to trajectories commonly taken by humans when interacting with the RAS, then a slight error in the RAS would likely have caused the accident.
	Explaining autonomous decisions (SAX Demonstrator)	In on-road and off-road driving scenarios, the project team studied the requirements of explanations for key stakeholders (users, system developers, regulators). These requirements informed the development of algorithms that generate causal explanations.
	Adapting current engineering processes (TIGARS Demonstrator)	The TIGARS project started with a joint a UK-Japan workshop where the team: identified assurance gaps in an experimental vehicle, set up experimental facilities and developed experimental and theoretic approaches to static analysis and dynamic assurance. The team undertook experimental research with a donkey car platform, and research to address some of the assurance challenges with autonomous road vehicles that they identified.
Maritime (n=3)	Boundaries of autonomy (BOAUT Demonstrator)	This project explored bounding behaviour (i.e. the identification of, and adherence to, limits that allow autonomous systems to stay safe) of maritime autonomous surface ships (MASS) as they deviate from their planned paths.
	Safe unmanned marine systems (ALADDIN Demonstrator)	The ALADDIN project developed a smart anomaly detection and fault diagnosis for marine autonomous systems (MAS) by introducing and implementing new methods for the detection and identification of adverse behaviour for MAS.
	Regulation and liability in autonomous shipping (Swansea University Demonstrator)	The work focused on two key areas: 1) to elaborate the scope of any new legal framework that may need to be put in place to ensure the safe operation of such vessels through SCCs; and 2) to define the legal position of seafarers on board such ships, and those in remote control centres. The primary objective of this study was to highlight the regulatory and legal challenges that need to be addressed so that maritime autonomous surface ships (MASSs) can operate in UK territorial waters without complication.
Manufacturing (n=2)	Assuring the safety of cobots (COBOT demonstrator)	The project used two case studies, placing specific emphasis on digital twins for safety analysis, machine learning for vision-based proximity detection, synthesis of safety controllers, testing approaches for analysis of hazards, and security policy, user authentication, and intrusion detection.
	Flexible manufacturing (RECOLL Demonstrator)	This project studied the safety-related human-robot behaviours (e.g. movements, layout occupation, voluntary/accidental contacts, near misses, etc) when the operations in a prototype machining production setup need frequent reallocation of human/robot tasks, uneven distribution of human location, and subtask-dependent potential physical interaction with machines.

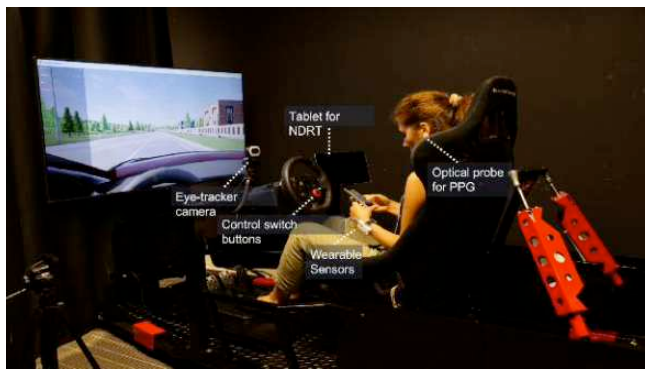
Sector	Demonstrator title	Description of project (Verbatim in past tense)
Aviation (n=2)	Remote inspection using drones (SAFEMUV Demonstrator)	This project developed a process for a systematic robustness assessment of UAV teams. This was underpinned by methods for the specification, generation, and testing of collaborative inspection scenarios, enabling the progressive transition from simulation to lab-based operations and to real-world operations.
	Wizard of Oz prototyping for automated decision-making tools in air traffic control (WIZARD Demonstrator)	The project built on the outcome of the recently completed A2URE project by testing the initial stages of a methodology for designing, developing, and testing automated decision-making tools. It developed and evaluated an innovative approach to allow for future ATC automated decision-making capabilities to be prototyped and validated in a more affordable and less time- and resource-intensive way.
Mining (n=1)	Assuring the safety of UAVs for mine inspection (ASUMI Demonstrator)	Using lab facilities in the Institute for Safe Autonomy at the University of York, the project team performed testing and simulations of multiple UAVs conducting mine inspections. Following this, the team conducted real-world analyses in Boulby Mine. The project team developed, defined, and validated safety requirements and a safe operating concept for multiple UAVs performing mine inspections to ensure safe operation and guarantee early intervention where required.
Space (n=1)	Assuring autonomy in space (ACTIONS Demonstrator)	Using autonomous in-orbit fire detection to support wildfire emergency response as the driving application, this project considered the safety assurance of ML algorithms onboard small satellites.
Agriculture (n=1)	Robots to support farming (MeSAPro Demonstrator)	This project complemented the team's existing work on the functionality and technical capabilities of the robots by defining safety requirements for the sense-understand-decide-act components of a soft-fruit production RAS, developing methods to detect deviations from safe behavior and ways to mitigate the effect of such deviations, and formally verifying the sensing, understanding, and deciding components of a soft-fruit production RAS.
Quarrying (n=1)	Assuring system-of-systems (SUCCESS Demonstrator)	The SUCCESS demonstrator project explored numerous aspects in the safety assurance of cooperating SoS, with a special focus on the construction machinery domain.

3 The use of the term ‘Human-centred’

When reviewing demonstrator projects, various uses of the term ‘human-centred’ were found. However, many studies often assume that AI or autonomous technology is human-centred if applied in a setting that directly impacts humans, suggesting different uses and understandings of the term. Box 1 illustrates this use of human-centred terminology within the demonstrators.

Shared control in autonomous driving (SafeSCAD Demonstrator)

Purpose of demonstrator: This demonstrator project focuses on developing a deep neural network framework to predict driver takeover behaviour, to ensure the driver can take over control when engaging in non-driving tasks. An overview of the system can be seen below:



Reflection: In the implication’s sections of the final project report, it is stated that they had developed a ‘human-centred’ framework. However, while they applied an approach that focused on the human-AI interaction, including involving users in the evaluation, overall, the project is not underpinned by the human-centred perspective as described in the current paper. This highlights that what is considered ‘human-centred’ may be different depending on perception.

Box 1. Example reflecting on the use of the term ‘human-centred’

The concept of ‘human-centred’ AI and autonomous technology has increased in recent years and has been the basis of several guidelines and principles, including the EU AI Act and the OECD principles (European Commission. 2024; OECD. 2019). When adopting a human-centred approach to developing AI and autonomous technology, the underlying principle is ensuring that the technology is designed with humans at the centre and that the technology benefits them in their everyday work (Ryan. 2024). However, this may be misinterpreted and lead to developers believing that only ensuring the AI or autonomous technology algorithm supports a perceived human need or positively impacts their everyday work will make it human-centred. While it is positive that humans are involved in this process and that their needs are considered when algorithms are developed, this alone cannot be considered a

human-centred approach. This is because the research focuses solely on the technological development of the algorithm and not on how the new technology will integrate and interact with the wider sociotechnical system. Therefore, there is a need to shift the understanding of what constitutes a human-centred approach to ensure accuracy. Additionally, this further highlights the need for a human-centred assurance framework as this can ensure the correct approach has been taken during the development of the AI or automated technology.

4 Understanding the wider work system

Another observation from the literature was around the need for a more holistic understanding of the work system in which AI and autonomous technology will be integrated. As mentioned in the introduction, AI will only be one part of the wider complex sociotechnical environment and will, therefore, interact with the other components in that work system. Without understanding these interactions, issues may arise when AI or automated technology is integrated into that work system. Despite the abovementioned importance, only the ASSIST demonstrator applied a method to understand the wider sociotechnical work environment. An overview of the ASSIST demonstrator can be seen in Box 2.

AI in ambulance response (ASSIST Demonstrator)

Purpose of demonstrator: The ASSIST demonstrator (2023) aimed to understand the operational environment from a systems perspective by applying the Systems Engineering Initiative for Patient Safety (SEIPS) model (Holden et al. 2013, Carayon et al. 2020) identify any assurance requirements and constraints for using a specific AI technology for the ambulance service.

Reflection: The demonstrator through observations and interview used the SEIPS model to understand the context where the AI system would be utilised. The SEIPS model is made up of six elements: persons, tools and technology, external environment, physical environment, organisation and tasks, which were used to understand the overall work system and how the different elements interact to create an outcome. The ASSIST demonstrator was able to use this understanding of the work system support the development of recommendation for the use of the AI technology from a systems perspective.

Box 2. Example reflecting on understanding the wider work system

The need to understand the work system is considered one of the key outputs from the post-hoc analysis, as in many of the demonstrators, this could have benefitted the research. For example, a demonstrator focusing on agriculture aimed to develop robots to support the fruit-picking process (MeSAPro demonstrator. 2021). In the project, the researchers created process scenarios, which were then used as a basis for analysis. These scenarios could have greatly benefited from understanding the work system and how the components already within the setting may interact with the fruit-picking robot. This understanding of the work system could then be used to verify that the system is working as planned and highlight any potential

hazards. Understanding the work system would also have been useful in an automotive demonstrator, which focused on assuring trust and trustworthiness of automotive vehicles (TIGARS demonstrator. 2020). In the demonstrator, the researchers state that an assurance case should, at a minimum, address what the system is, the environment it will work within, how much trust is necessary given the environment, whether it is trustworthy enough to deploy, and whether it will continue to be trustworthy when changes to the environment occur. By completing an analysis of the work system, a comprehensive overview of that environment and how this could change with the introduction of new technology may have supported the development of the assurance case. Overall, understanding the work system would allow for a holistic view of the environment where the new AI or autonomous technologies will be integrated, which could support a number of research activities.

5 Taking a human-centred lens to safety

During the analysis of the demonstrators, there was a clear focus on ensuring appropriate safety assurance when deploying AI and autonomous technology. An example where safety assurance was looked at from a human-centred lens is the ASSIT demonstrator. In this demonstrator, the researchers aimed to understand and critique assurance techniques from a sociotechnical perspective to ensure that the whole clinical system was considered during this process. A further example of where a human-centred lens was taken for safety was a demonstrator within the automotive sector applied the Systems Theoretic Process Analysis (STPA) and Functional Resonance Analysis Method techniques to develop resilience requirements for the technology design (TIGARS demonstrator. 2020). Additionally, the STPA method was applied in the quarrying sector to identify potential accidents and their causes and in the manufacturing sector to complete a hazard analysis of a Cobot system case study (SUCCESS demonstrator. 2020; COBOTS demonstrator, 2022).

While several studies applied a human-centred approach to focusing on the safety of AI and automated technology, these were still in the minority. Based on the post-hoc analysis, several demonstrators could have benefitted from a human-centred approach to ensure the full system was considered. Box 3 provides an example of one of these demonstrators.

Robots to support farming (MeSAPro Demonstrator)

Purpose of demonstrator: This demonstrator aimed to understand the use of autonomous robotic systems to support human fruit pickers and reduce workplace accidents.

Reflection: The demonstrator highlighted several hazards with a corresponding severity score, but it was unclear if they had considered the full system, meaning that some hazards could have been missed. Therefore, this demonstrator could have applied a human-centred approach, similar to the demonstrators described previously and applied methods such as STPA to ensure full consideration for the system where the AI or autonomous technology will be implemented. Additionally, this demonstrator could have applied a human-centred approach to their development of safety requirements.

Box 3. Example reflecting on taking a human-centred lens to safety

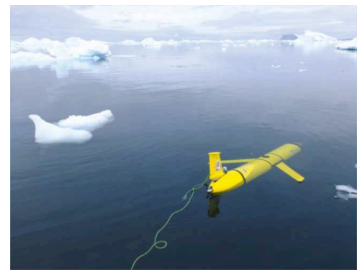
Within the healthcare sector, a demonstrator focusing on an AI system providing sepsis management guidance could have taken a human-centred approach to understanding safety (Safety of AI Clinicians demonstrator. 2022). The demonstrator applied AMLAS, where a sociotechnical perspective could have been taken for several stages to ensure an understanding of the work system where the AI will be used. Overall, by taking a human-centred approach, researchers would better understand the interactions that may occur with the introduction of AI and autonomous technology, which may influence overall safety.

6 Engaging users and stakeholders

One of the most common human-centred approaches found in the demonstrators was the engagement of stakeholders. Engagement with stakeholders is a key aspect of human-centred development as it allows AI and autonomous technology to be developed in line with the system where it will be integrated. A number of different stakeholders were engaged within the demonstrators; for example, a healthcare demonstrator (SAFR demonstrator. 2023) completed stakeholder engagement with manufacturers to develop guidance for developers and regulators, using a framework of questions based on the stages of AMLAS to gain their input. Further, the ALADDIN demonstrator engaged stakeholders, as seen in Box 4.

Safe unmanned marine systems (ALADDIN Demonstrator)

Purpose of demonstrator: The project aimed to develop a smart anomaly detection and fault diagnosis technology for marine autonomous technology. An example of the autonomous technology can be seen below:



Reflection: The demonstrator engaged stakeholders by bringing together a panel to discuss several questions relating to the adoption of AI and automated technology within the maritime sector. The discussion included topics on human-AI teaming, transparency and explainability and data needs to ensure any future maritime autonomous system is developed in line with stakeholder needs.

Box 4: Example reflecting on engaging users and stakeholders

However, alongside the engagement of stakeholders, the needs of future users of AI and automated technology should be considered. Some studies did consider the users and their requirements; for example, within healthcare, a study looking at the development of a robotic assistant for assisted living engaged medical personnel, caregivers and potential end users through surveys to understand what their needs would be for using the robot in day-to-day routines (ALMI demonstrator. 2023). Overall, it is encouraging that some demonstrators engaged stakeholders and users. However, both users and stakeholders must be engaged in the future, as this will allow for consideration of the wider environment through stakeholders such as the developers and regulators and the needs of those directly using the technology. It would, therefore, be potentially beneficial to develop a 'people map' similar to the one completed in the ASSIST demonstrator. This 'people map' would then allow for an understanding of all stakeholders and users involved and ensure all their views were considered during the development of AI and autonomous technology (ASSIST demonstrator. 2023; Svedung & Rasmussen. 2000).

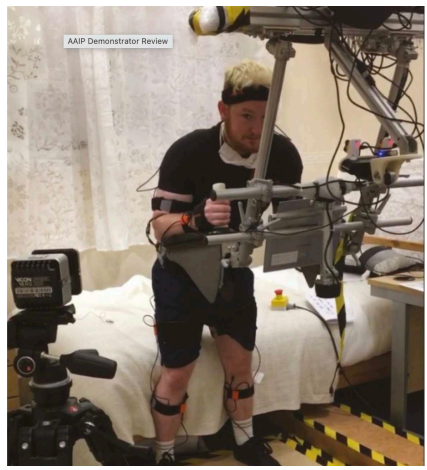
7 Human-AI interactions

The demonstrator projects also highlighted the need to understand how humans interact with AI or autonomous technology, as challenges can arise without consideration. One challenge concerns the handover between AI or autonomous technology and human operators. Within the automotive sector, handover could involve the driver taking over from the automated vehicle, which can lead to difficulties if

the driver is not actively engaged and has limited situational awareness of the task (Sujan et al. 2020). A further challenge is a potential change in the role of the humans involved, where they are no longer active participants but act in a supervisory role. An example of this can be seen in the maritime sector, where human operators are now remote and supervising several automated ships. This leads to a potential loss of skills, poor situational awareness, cognitive overload and the operators feeling displaced. (Sujan et al. 2020) A human-centred approach to the development and assurance of AI and autonomous technology can be beneficial to ensure these challenges are overcome. This includes considering the skills and training needed for those working alongside the new AI or automated technology. The Assistive Robots in Healthcare demonstrators highlighted the need for increased training, as seen in Box 5.

Assistive robots in healthcare (UWE Demonstrator)

Purpose of demonstrator: This demonstrator aimed to investigate the safety and regulatory requirements for the use of physically assistive robots. The overarching aim of the demonstrator was to identify future knowledge and training needs for the use of these physically assistive robots in healthcare. An image of a participant in a Xsens suit used for the demonstrator is below:



Reflection: Alongside technical research the demonstrator using a review of the literature and a survey wished to understand the training needs of healthcare professionals who would operate the robotic assistant. The results would then be used to develop future training material, and support in defining industry standards for operator training.

Box 5: Example reflecting on human-AI interactions

Alongside training, the design may consider the human-AI interaction, with methods such as Wizard of Oz. This Wizard of Oz method was used by an Aviation demonstrator (WIZARD demonstrator. 2023) and is a prototyping method that can be used for interaction design to understand how the human user will work alongside the machine.

8 Conclusions

The demonstrator review highlighted how previous projects applied or could have applied a human-centred approach to developing AI and autonomous technology across sectors. Analysing the literature provided reflections that may need further research or thought, including more education on what constitutes a human-centred approach and how they can support research within AI and autonomous technology. These reflections also highlight that a human-centred assurance framework would be useful to ensure that approaches are conducted correctly and that the benefit of applying them to AI and autonomous technology is known. Future research will use these observations and further results from a scoping review to create the initial requirements for the human-centred assurance framework. Once the initial requirements have been established, a number of activities may be undertaken to create the human-centred assurance framework. These activities could include consensus work and focus groups to ensure that the important approaches are captured and that the framework will integrate well with previously developed frameworks, such as AMLAS and SACE. Overall, using human-centred approaches to develop AI and automated technology is beneficial and can ensure a good understanding of the sociotechnical work system and the interactions within that work system where the technology will be used. However, a human-centred assurance framework would be useful to ensure that these approaches are known and used during AI and autonomous technology development.

Acknowledgments This work was funded by the Engineering and Physical Sciences Research Council (Assuring Responsibility for Trustworthy Autonomous Systems, EP/W011239/1) and supported by the Centre of Assuring Autonomy, a partnership between Lloyd's Register Foundation and the University of York

References

Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. Sustainability, 11(1), 189.

ACTIONS Demonstrator (2022) Assuring autonomy in space. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/assuring-autonomy-in-space/. Accessed 14 October 2024

ALADDIN Demonstrator (2022) Safe unmanned marine systems. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/safe-unmanned-marine-systems/. Accessed 14 October 2024

ALMI Demonstrator (2023) Safe robots for assisted living. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/safe-robots-assisted-living/. Accessed 14 October 2024

ASSIST demonstrator (2023) AI in ambulance response. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/ai-ambulance-response/. Accessed 14 October 2024

ASUMI Demonstrator (2023) Assuring the safety of UAVs for mine inspection. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/uav-boulby-mine/. Accessed 14 October 2024

ATM Demonstrator (2021) Automatic rating system for autonomous systems. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/automatic-rating-system/. Accessed 14 October 2024

BOAUT Demonstrator (2022) Boundaries of autonomy. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/boundaries-of-autonomy/. Accessed 14 October 2024

Canella, S., Flis, V., Robert Roj, I., & Malkar, I. (2024). Perception and acceptability of social robots in healthcare: ethnographic research based on a qualitative case study. Salute e società: XXIII, 2, 2024, 88-102.

Carayon, P., Wooldridge, A., Hoonakker, P., Hundt, A. S., & Kelly, M. M. (2020). SEIPS 3.0: Human-centered design of the patient journey for patient safety. Applied ergonomics, 84, 103033

COBOT demonstrator (2022) Assuring the safety of cobots. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/assuring-the-safety-of-cobots/. Accessed 14 October 2024

Demirel, H. O., Goldstein, M. H., Li, X., & Sha, Z. (2024). Human-centered generative design framework: an early design framework to support concept creation and evaluation. International Journal of Human–Computer Interaction, 40(4), 933-944.

European Commission (2024b) Coordinated plan on artificial intelligence. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/plan-ai. Accessed 14 October 2024

Hawkins, R., Osborne, M., Parsons, M., Nicholson, M., McDermid, J., & Habli, I. (2022). Guidance on the safety assurance of autonomous systems in complex environments (SACE). arXiv preprint arXiv:2208.00853.

Hawkins, R., Paterson, C., Picardi, C., Jia, Y., Calinescu, R., & Habli, I. (2021). Guidance on the assurance of machine learning in autonomous systems (AMLAS). arXiv preprint arXiv:2102.01564.

Holden, R. J., Carayon, P., Gurses, A. P., Hoonakker, P., Hundt, A. S., Ozok, A. A., & Rivera-Rodriguez, A. J. (2013). SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics, 56(11), 1669-1686.

Lutzhoft, M., Hynnekleiv, A., Earthy, J. V., & Petersen, E. S. (2019, October). Human-centred maritime autonomy-An ethnography of the future. In Journal of Physics: Conference Series (Vol. 1357, No. 1, p. 012032). IOP Publishing.

MeSAPro Demonstrator (2021) Robots to support farming. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/robots-to-support-farming/. Accessed 14 October 2024

Munim, Z. H., Dushenko, M., Jimenez, V. J., Shakil, M. H., & Imset, M. (2020). Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. Maritime Policy & Management, 47(5), 577-597.

OECD (2019) AI-principles overview. https://oecd.ai/en/principles. Accessed 14 October 2024

Porter, Z., Habli, I., McDermid, J., & Kaas, M. (2024). A principles-based ethics assurance argument pattern for AI and autonomous systems. AI and Ethics, 4(2), 593-616.

Porter, Z., Ryan, P., Morgan, P., Al-Qaddoumi, J., Twomey, B., McDermid, J., & Habli, I. (2023). Unravelling Responsibility for AI. arXiv preprint arXiv:2308.02608.

Rasmussen, J., & Suedung, I. (2000). Proactive risk management in a dynamic society. Swedish Rescue Services Agency.

RECOLL Demonstrator (2020) Flexible manufacturing. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/flexible-manufacturing/. Accessed 14 October 2024

Ryan, M. We're only human after all: a critique of human-centred AI. *AI & Soc* (2024). https://doi.org/10.1007/s00146-024-01976-2

SAFEMUV Demonstrator (2022) Remote inspection using drones. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/remote-inspection-using-drones/. Accessed 14 October 2024

SafeSCAD Demonstrator (2022) Shared control in autonomous driving. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/autonomous-driving/. Accessed 14 October 2024

SCSC publications (2024) All publications. Safety Critical Systems Club. https://scsc.uk/scsc-198. Accessed 30 October 2024

Safety of the AI clinician Demonstrator (2022) Safety of the AI clinician. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/safety-ai-clinician/. Accessed 14 October 2024

SAFR Demonstrator (2023) Machine learning in healthcare. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/machine-learning-healthcare/. Accessed 14 October 2024

Salwei, M. E., & Carayon, P. (2022). A sociotechnical systems framework for the application of artificial intelligence in health care delivery. Journal of cognitive engineering and decision making, 16(4), 194-206.

SAM Demonstrator (2022) Medication management. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/medication-management/. Accessed 14 October 2024

SAX Demonstrator (2022) Explaining autonomous decisions. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/explaining-autonomous-decisions/. Accessed 14 October 2024

Social credibility Demonstrator (2019) Social credibility. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/social-credibility/. Accessed 14 October 2024

SUCCESS Demonstrator (2020) Assuring system-of-systems. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/assuring-system-of-systems/. Accessed 14 October 2024

Sujan, M., Furniss, D., Hawkins, R. D., & Habli, I. (2020). Human factors of using artificial intelligence in healthcare: challenges that stretch across industries. In Safety-Critical Systems Symposium. York.

Sujan, M. A., White, S., Habli, I., & Reynolds, N. (2022). Stakeholder perceptions of the safety and assurance of artificial intelligence in healthcare. Safety science, 155, 105870.

Swansea University Demonstrator (2022) Regulation and liability in autonomous shipping. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/regulation-maritime-remote-control-centres/. Accessed 14 October 2024

TIGARS Demonstrator (2020) Adapting current engineering processes. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/adapting-current-engineering-processes/. Accessed 14 October 2024

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., *Colquhoun*, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Annals of internal medicine, 169(7), 467-473.

Tschandl, P., Codella, N., Akay, B. N., Argenziano, G., Braun, R. P., Cabo, H., ... & Kittler, H. (2019). Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. The lancet oncology, 20(7), 938-947

UWE Demonstrator (2021) Assistive robots in healthcare. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/assistive-robots-healthcare/. Accessed 14 October 2024

WIZARD Demonstrator (2023) Wizard of Oz prototyping for automated decision-making tools in air traffic control. Assuring Autonomy International Programme. https://www.york.ac.uk/assuring-autonomy/about/aaip/demonstrators/automated-decision-making-air-traffic-control/. Accessed 14 October 2024