Human Factors of Using Artificial Intelligence in Healthcare: Challenges That Stretch Across Industries

Mark Sujan[1], Dominic Furniss[1], Richard Hawkins[2], Ibrahim Habli[2]

1 Human Reliability Associates, UK

2 University of York, UK

Abstract *The use of artificial intelligence (AI) in healthcare is one of the fastest growing industries worldwide. AI is already used to deliver services as diverse as symptom checking, skin cancer screening, and recognition of sepsis. But is it safe to use AI in patient care? However, the evidence base is narrow and limited, frequently restricted to small studies considering the performance of AI applications at isolated tasks. In this paper we argue that greater consideration should be given to how the AI will be integrated into clinical processes and health services, because it is at this level that human factors challenges are likely to arise. We use the example of autonomous infusion pumps in intensive care to analyse the human factors challenges of using AI for patient care. We outline potential strategies to address these challenges, and we discuss how such strategies and approaches could be applied more broadly to AI technologies used in other domains.*

1 Introduction

Expectations for the use of artificial intelligence (AI) in healthcare are high. In the UK, as well as world-wide, politicians and policy makers are quick to highlight the potential health and economic benefits that the widespread adoption of AI can bring. This is underpinned by the establishment of new dedicated bodies, such as NHSX[1] in the UK and significant government funding to facilitate and speed up the development and adoption of AI in health services. AI is a major disrupter to health systems, and it will transform the way healthcare is delivered and accessed by patients (Coiera, 2018).

Examples of the use of AI in healthcare include machine learning algorithms that rely on pattern recognition, classification and prediction. For example, deep learning is particularly well suited to the interpretation of radiological images

[1] https://www.nhsx.nhs.uk/

because of the complexity and richness of the data (Saria et al., 2018). Deep neural networks (DNN) have been used to interpret head CT scans (Chilamkurthy et al., 2018), to identify skin cancer (Haenssle et al., 2018) and to recognise diabetes (Avram et al., 2019). AI-driven chatbots are another popular application domain, e.g. patient-facing symptom checkers (Semigran et al., 2015) or artificial agents delivering cognitive behavioural therapy to mental health patients (Fitzpatrick et al., 2017).

Evaluation studies of such AI algorithms have produced encouraging results. The evaluation of a bedside computer vision algorithm to identify and monitor behaviours of clinicians, such as hand washing, suggests that the algorithm can achieve 95% accuracy (Yeung et al., 2018). Skin cancer detection using algorithms might outperform dermatologists at this task (Esteva et al., 2017). Similarly, the developers of a DNN to detect diabetic retinopathy[2] found their algorithm achieved over 95% accuracy on two test sets (Gulshan et al., 2016). For the management of sepsis, the evaluation carried out by the developers of an algorithm trained by reinforcement learning found that on average patient mortality was lower when clinicians' management decisions matched those suggested by the AI (Komorowski et al., 2018).

However, looking across these studies, the focus of the evaluation is usually on the performance of the AI on a narrowly defined task. The evaluation is typically undertaken by the developers, and independent evaluation remains the exception. For example, the above evaluation of AI sepsis management has been criticised because the algorithm seemingly "learned" not to treat very ill patients – a strategy that fits with the training reward function, but is hardly suitable in a real clinical environment (Jeter et al., 2019). Sample sizes are often small, and prospective trials are infrequent. As a result, the evidence base to date about the actual performance of AI in real-world settings remains weak (Yu and Kohane, 2019).

There is relatively little evidence about the safety of using AI for patient care, and we argue that this is, in part, due to the focus on performance of the algorithms. The real challenges for the adoption of AI will arise when algorithms are integrated into clinical systems to deliver a service in collaboration with clinicians as well as other technology (Sujan et al., 2019d). It is at this clinical system level, where teams consisting of healthcare professionals and AI systems cooperate and collaborate to provide a service, that human factors challenges will come to the fore (Sujan et al., 2019b).

In this paper we analyse the human factors challenges of using AI for patient care as part of a clinical system, and we identify potential strategies for addressing these. The next section describes the scenario of autonomous infusion pumps in intensive care, which we use to illustrate the concepts. In section 3 we analyse

[2] Diabetic retinopathy is a condition of the eye that can affect people with diabetes. It is a leading cause of sight loss and blindness in the UK.

the scenario for human factors challenges and develop example strategies for dealing with them. In Section 4 we discuss how the identified strategies could be applied more broadly to AI systems used in other domains. We conclude the paper with a summary and outlook.

2 Scenario: Autonomous Infusion Pumps in Intensive Care

As a reference case we use a scenario developed within the Safety Assurance of Autonomous Intravenous Medication Management Systems (SAM) project (Sujan et al., 2019a). The SAM project[3] is funded under the Assuring Autonomy International Programme (AAIP)[4], and it is a collaboration between Human Reliability Associates (a human factors and safety consultancy), NHS Digital (an arms-length body of the Department of Health), and clinicians based at Royal Derby Hospital. The project explores safety assurance strategies for novel, highly-automated or autonomous infusion pumps within the intensive care setting. Figure 1 provides an illustration of the intensive care setting.

Fig. 1. Simulated patient in intensive care. The patient is on a ventilator. The stack of infusion pumps is on the left, next to the screen that charts the patient's data.

The motivation for considering the use of AI for intravenous medication management is twofold: to reduce medication errors, and to improve efficiency and

[3] http://www.humanreliability.com/casestudies/sam_project/

[4] https://www.york.ac.uk/assuring-autonomy/

effectiveness. Medication errors are a significant problem for the National Health Service (NHS), and health systems world-wide. A 2018 report estimates that as many as 237 million medication errors occur in England every year, and that these cause over 700 deaths (Elliott et al., 2018). Intravenous medication preparation and administration are particularly vulnerable activities, and therefore such infusion errors represent a considerable burden to patients and the health system (McLeod et al., 2013, Furniss et al., 2019).

In order to reason about the capabilities of automated and autonomous infusion pumps we took inspiration from the automotive domain, where a 6-level taxonomy of driver-vehicle control was developed by the Society of Automotive Engineers (SAE), which ranges from no automation (level 0) through to full automation (level 5). We used this approach and developed analogous levels of automation for infusion pumps, as shown in figure 2. Level 2 represents current smart pump capabilities, where the pump is able to undertake a number of automated checks, e.g. drug and patient identification. At level 5, which represents the scenario of future AI technology considered in this paper, an autonomous infusion pump is able to take clinical guidelines (e.g. for insulin administration) as a starting point, but has the ability to learn and modify these based on continuous monitoring of the patient's physiological response to the drug. We consider the reference scenario described in Table 1.

Table 1. Reference Scenario

Reference Scenario: L5 Infusion Pump
The patient is a 68-year old type II diabetic with sepsis secondary to pneumonia. The patient's blood sugars require insulin control via IV actrapid insulin infusion. Patient identity, nurse identity, prescription and syringe formulation checks are all done by barcode. If checks match, the pump automatically programmes itself to start the infusion, displays medication identity and selects hard and soft programme infusion rate limits without further or final human confirmation. The pump controls the IV infusion rate of insulin in response to continuously measured blood sugar from a central venous sampling device. Within the programmed limits it is able to "learn" the patient's actual insulin requirements and formulate an individualised protocol for the infusion rate based on the sugar readings to optimise sugars control through pre-emptive changes in infusion rates.

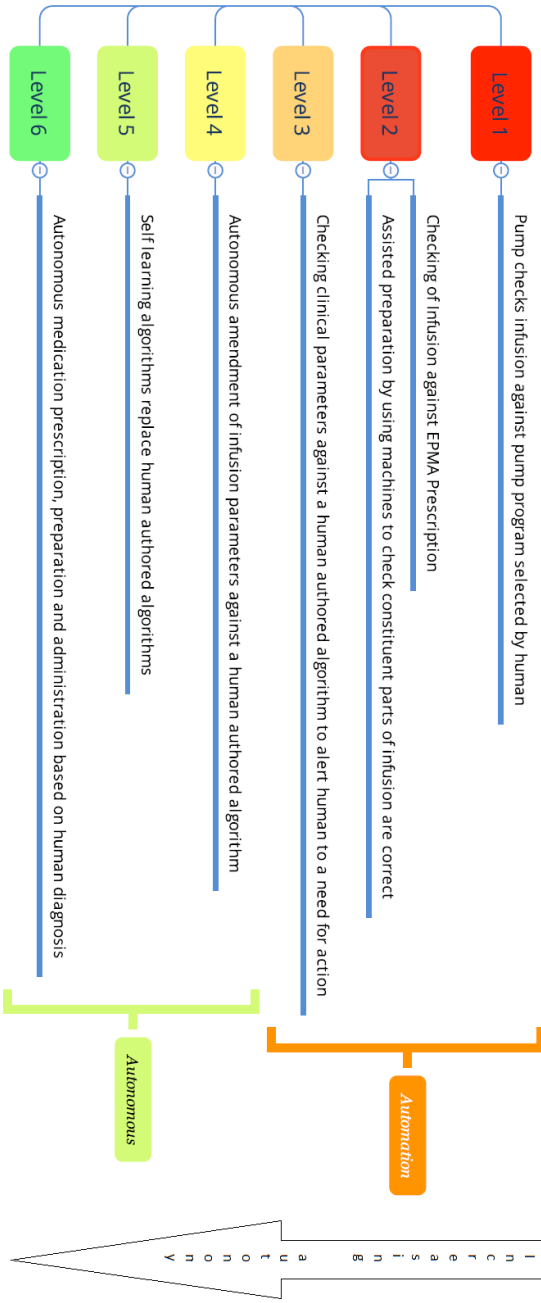

Fig. 2. Levels of Automation - Infusion Pumps

3 Human Factors Analysis

We undertook a human factors analysis of the reference scenario in order to identify human factors challenges that might impact on safe and effective care. The focus of the analysis was the clinical system, which includes consideration of how clinicians interact with the AI infusion pump, other tools and systems that might communicate with the infusion pump and clinicians, the impact on teamwork and the organisation of work, and the impact on communication with patients and patient experience. This socio-technical unit of analysis is shown in figure 3.

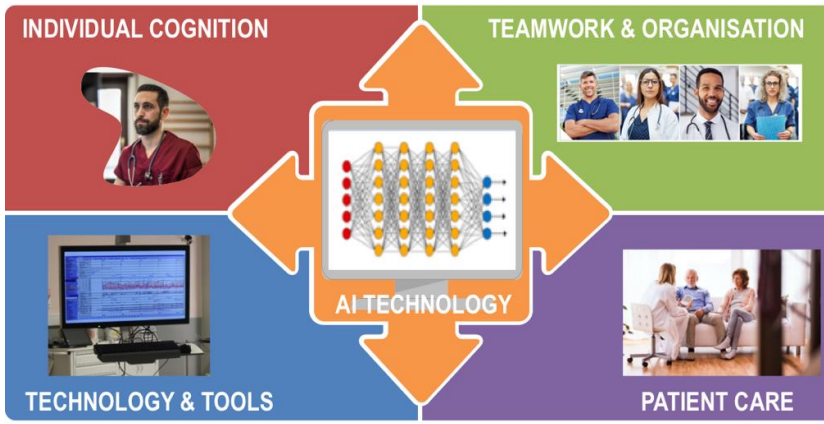

Fig. 3. Socio-technical system unit of analysis

As part of the analysis we undertook a task analysis of the current process as baseline. This was mapped using the Hierarchical Task Analysis (HTA) approach (Stanton, 2006). We undertook a human failure analysis using the Systematic Human Error Reduction and Prediction Approach (SHERPA) methodology (Embrey, 1986). We then mapped the future state that incorporates the autonomous infusion pump. The process map is shown in figure 5 in the appendix. The analysis involved a clinical team consisting of a consultant anaesthetist, an intensive care nurse and a pharmacist. We also interviewed 10 further clinicians about their views on the potential impact of using autonomous infusion pumps in intensive care.

The human factors analysis identified a number of human factors challenges that need to be considered and addressed in order to provide assurance that the AI can be integrated safely into a clinical system, and that the overall service is safe. An overview of the human factors challenges is given in table 2. The table contextualises the identified human factors challenges within the autonomous infusion pump example.

Table 2. Human factors challenges

HF Challenge	Description	Example
Handover	The autonomous system needs to be able to recognise its own performance boundaries, project into the future clinical scenarios that will be beyond its performance boundaries, and identify suitable ways to hand over control to the clinician. Handover includes consideration of: (a) when to hand over; (b) whom to hand over to; (c) what to hand over; and (d) how to hand over.	The patient's blood sugar levels do not respond sufficiently to the insulin given by the autonomous infusion pump. The pump predicts and recognises that it will not be able to control the patient's blood sugar. The pump triggers an alert on the electronic health record, raises an audible alarm, and requests the nurse to take over. The nurse can review the reason for the alert, the history of the pump's insulin management, and its projection into the future, and act accordingly.
Performance Variability	Clinicians need to manage competing organisational priorities and operational demands. They use their experience and judgement to make trade-offs based on the requirements of a specific situation. The autonomous system needs to support rather than constrain this performance variability and adaptive capacity.	The nurse realises that insulin has not yet been prescribed for the patient even though they will likely need it. The nurse goes and finds the doctor, explains the situation, and the doctor issues a verbal medication order and will follow this up with the written prescription later (performance variability). The autonomous system requires an electronic medication order, but allows for a manual override. The autonomous system sends reminders to the doctor with a request for completing the electronic medication order.
Automation bias	When a system works well most of the time, clinicians start to rely on it. In some situations, this can lead to overreliance, for example when the system takes an inappropriate action but the clinician does not recognise this because they trust the system.	Due to sepsis the patient requires tighter control of blood sugar levels than usual. The autonomous system has managed successfully septic patients before but, in this instance, fails to recognise the need for tighter glycaemic control. The autonomous system provides clinician interpretable justification and explanation of its decisions, and the clinician, who has received training on potentially inappropriate behaviours of the autonomous system, is able to spot the discrepancy and act accordingly.

Supervision	Clinicians are both users and supervisors of the autonomous system. They need to understand not only how to operate the autonomous system (e.g. loading a syringe), but also how to recognise potential failure modes or deviations from appropriate behaviour or changes in the environment that might move the autonomous system outside of its design envelope.	The autonomous infusion pump is operating on the sliding scale algorithm for administering insulin. It classifies the patient's response to the current insulin infusion as requiring transition to another scale with 70%, as opposed to 30% for staying within the current scale. The autonomous system initiates and the transition, and activates an "uncertainty marker" to alert the clinician.

3.1 Handover

Handover between clinicians has long been recognised as a safety-critical and particularly vulnerable activity (Sujan et al., 2015b). Handover is not simply the transfer of information from a sender to a more or less passive receiver, but involves collaboration, negotiation and coordination (Sujan et al., 2015c). The introduction of AI and autonomous systems complicates handover even further, as the AI needs to identify appropriate trigger points for handover, it needs to determine the appropriate person or persons to hand over to and understand their information needs, and it needs to use adequate communication channels for the handover. Such trigger points could be the self-detection of an internal fault or the recognition of situations outside of the system's design envelope. For example, should the autonomous infusion pump wait to raise an alert until it fails to control the patient's blood sugar levels or should it communicate its potential failure much sooner to allow clinicians to prepare for taking back control? Should it just raise an alert or should it communicate a history of its actions and a prognosis of the patient's physiological development? Should it sound an audible alarm so that the nurse can pick this up, or should it send a text message to doctors not close to the bedside?

All of these considerations are human factors concerns, and a look at the wider human factors body of knowledge can provide insight into potential approaches for designing adequate handover strategies. For example, the autonomous infusion pump would ideally initiate a form of graceful handover, where the trigger points are determined considering human performance characteristics as well as the specific clinical scenario. The design of the infusion pump should consider the information needs of different types of stakeholders that allows them to build an adequate situation awareness (Endsley, 1995). Alarm prioritisation and alarm

management are further strategies that have been developed in control room operations to prevent operator overload (e.g. EEMUA 191 Alarm Systems[5]), and these should also inform the way the handover between the autonomous infusion pump and clinicians is designed.

3.2 Performance Variability

Within the resilience engineering (RE) community performance variability is regarded as an asset that enables complex systems to deal with disturbances, conflicting goals, and unforeseen situations (Hollnagel et al., 2006). People continuously adapt their behaviour and make trade-offs, often based on some form of subjective risk assessment, and in this way they are able to cope with competing demands, uncertainty, and everyday disturbances such as staff shortages and peaks in demand (Sujan et al., 2015a). This work-as-done (WAD) is necessarily different from work-as-imagined (WAI) by people who design and manage systems (Hollnagel, 2015).

Our human factors analysis provided several examples of such everyday local adaptations. For example, nurses would sometimes administer drugs without a prescription and then chase the doctor to issue a prescription later. They do this depending on the perceived risk category of the drug, the urgency of administering the drug, the availability of the doctor, and their own workflows. This violates the protocol, which requires that a prescription is issued prior to administration of any drug, but it enables smoother functioning of the intensive care unit as a whole, and it can adapt better to patient needs.

Safety assurance of new technology focuses frequently only on the failure modes of the technology and the associated risks. However, from a Safety-II perspective it is equally as important to consider the impact on the resilience abilities of the (clinical) system, i.e. the impact on the ability to anticipate, to adapt, to monitor, and to learn (Hollnagel, 2014).

In the autonomous infusion pump scenario, it is easy to envisage how the static implementation of procedures and protocols might disrupt existing workflows and, in this way, create the need for other workarounds. The design of the infusion pump should consider WAD, e.g. with data collected through observations, interviews and task analysis. There is also a need to look beyond the immediate impact on human – machine (i.e. clinician – infusion pump) interaction, towards the potential impact the introduction of technology has on human – human relationships. Building and maintaining such relationships is an important aspect of

[5] Standard available for a fee from the Engineering Equipment and Materials User Association (EEMUA): https://www.eemua.org/Products/Publications/Print/EEMUA-Publication-191.aspx.

resilient health care (Sujan et al., 2019c). The introduction of technology should not prevent opportunities for building relationships and trust among clinicians.

3.3 Automation Bias

Automation bias describes the phenomenon that people tend to trust and then start to rely on automation uncritically (Parasuraman and Riley, 1997). An interesting recent study in the automotive domain found that even with training and specific instruction on the limitations of an autonomous vehicle, study participant drivers came to rely on the autonomous car within a week, and were spending most of their time on their smartphones or reading (Burnett et al., 2019). Examples of automation bias have also been found in healthcare, for example in mammography reading, where the introduction of a computer algorithm can decrease the performance of radiologists for certain difficult cases, where the algorithm provided incorrect classification (Alberdi et al., 2004, Lyell and Coiera, 2016).

Many, if not most, AI systems will be advertised as having ultra-high reliability, and it is to be expected that in due course clinicians will come to rely on these systems. However, the studies on automation bias suggest that the reliability figures by themselves do not allow prediction of what will happen in clinical use, when the clinician is confronted with a potentially inaccurate system output. It is important that clinicians are informed not only about the accuracy of algorithms, but also about their potential weaknesses and what to look out for.

Guarding against automation bias is not easy. While studies have suggested a number of strategies such as explainability and transparency of decision making, clear accountability and adaptive interfaces and task allocation, the evidence base for these is far from clear (Goddard et al., 2012). Different people might have different mental models and assumptions of the autonomous infusion system, which might be partial and even contradictory, because the behaviour may be too complex for anyone to understand what is going on. Technology developers, healthcare providers and clinicians need to have an awareness of this challenge, and find solutions that work in their specific setting so that users can build a good picture of the behaviour of the autonomous system in a way they can comprehend.

3.4 Supervision

With current infusion pumps (at L2 in our model of automation) clinicians are users of the infusion pump, i.e. they need to know how to load and program the

infusion pump. Failure modes of the infusion pump are fairly limited and reasonably well understood. The training provided to clinicians is about the functionality of the infusion pump and how to use it, e.g. how to navigate the interface.

The situation changes dramatically when we move to L5, because at this level the infusion pump becomes an autonomous system capable of taking decisions independently. The clinician needs to understand not only how to use the infusion pump, but also what potential weaknesses are and how the safe envelope is defined, maintained or breached. Clinicians need to be able to make sense of the pump's actions and provide clinically-based checks. In this sense, the role of the clinician changes from user of a passive pump to that of supervisor of an autonomous system. Consideration needs to be given to how clinicians can fulfil this role, and what kind of novel training needs might arise. It is even conceivable that a new role is created, e.g. that of an AI nurse specialist, who is specifically trained in managing AI and autonomous systems within their care setting.

4 Cross-Domain Discussion

This paper has identified a number of key human factors challenges through consideration of automation and autonomy in healthcare. Although the specific issues highlighted relate to the introduction of autonomous infusion pumps, the general challenges are not unique to this application and domain, and are likely to be more broadly applicable. In this section we discuss the generalisability of a number of the challenges presented through consideration of examples in other domains where the level of autonomy of systems is also increasing. We believe that there is great benefit that can be gained through sharing knowledge between domains on how to address these challenges.

4.1 Handover

The problems associated with handover from autonomous operation to a human operator are well known in other domains and have been widely studied. None more so than in the automotive domain where recent high-profile incidents have highlighted concerns around the use of so-called safety drivers. Current self-driving cars are only capable of driving autonomously under limited conditions such as defined geographical areas, types of roads or specified scenarios and environmental conditions. This means that the vehicle must hand back control to a human driver if the required conditions are not met, or if the car is in a situation that it

cannot resolve safely. Studies such as (Gold et al., 2013) have shown that, depending on the complexity of the situation at handover, it can take up to 8 seconds for a driver to take back full control of the vehicle, particularly if they are distracted at the time of handover. When driving on a motorway, 8 seconds may correspond to over 200 metres travelled. It does not seem unreasonable to take the high-end of this estimate. Given that the handover to a safety driver will often occur because the vehicle is in a difficult or dangerous state it is likely that the situation is complex. The ability of drivers who are not actively engaged in driving the vehicle to avoid distraction is also challenging, as discussed in (Merat et al., 2014). This is an area of active research, but there are certainly strongly held views, such as by Waymo (Waymo, 2018) that ultimately human drivers should be removed, as they cannot be relied upon to react quickly enough to ensure safety. This view has been given additional weight by accidents such as that in Tempe, Arizona in March 2018 (National Transportation Safety Board, 2018), where for various reasons the safety driver was unable to intervene quickly enough to prevent a fatality.

Human factors strategies such as graceful handover, situational awareness and designing with consideration of performance influencing factors, can also be seen to be crucial for ensuring safe handover in autonomous driving, and are also more broadly applicable to other domains.

4.2 Automation bias and impact on working practices

Aircraft have been highly automated for a long time, prompting research to investigate what the consequences of this might be on pilots' ability to fly the aircraft manually if the automated systems fail. This has been particularly motivated by a number of crashes that may have involved some element of de-skilling on the part of the pilots, such as Colgan Flight 3407 in 2009 when 50 people died when the pilots were found to have done the opposite of what they were trained to do when the aircraft entered an aerodynamic stall (National Transportation Safety Board, 2009), see figure 4 for an image of the crash site (Clarence Centre, New York). The paradox is that it would seem that although automation has made it increasingly unlikely that airline pilots will face critical problems during flight, it is also perhaps making it less likely they will be able to cope if such problems do arise.

Fig. 4. Crash site of Colgan Flight 3407 in 2009
(copyright: Bureau of Aircraft Accident Archives)

A study from 2014 (Casner et al., 2014) set out to understand how the prolonged use of cockpit automation has been affecting pilots' manual flying skills. They did this through experiments on 16 Boeing 747-400 airline pilots in a simulator, where they systematically varied the level of automation used to fly routine and non-routine flight scenarios. What they found was that pilots' instrument scanning and manual control skills to be mostly intact, even when pilots reported that they were infrequently practiced. However, when pilots were asked to manually perform the cognitive tasks needed for manual flight (e.g., tracking the aircraft's position without the use of a map display, deciding which navigational steps come next, recognizing instrument system failures), more frequent and significant problems were observed, and this seemed to depend on the degree to which pilots remain actively engaged in supervising the automation. Such observations in a domain where the use of high levels of automation are long established clearly bring knowledge that may be important for domains such as healthcare where high levels of autonomy are novel.

4.3 Supervision

In the maritime domain there are ambitious plans for autonomous operation of ships. For example, Rolls-Royce plan to operate a fleet of unmanned ships across

the world using a small number of operators in shore-based control centres (SCCs), which could be located thousands of miles away[6]. Unmanned shipping does not mean removing humans completely from operations, but moving them to a role more focussed on monitoring and supervision, requiring entirely new kinds of work roles, tasks, tools, training and environments. Crucially, to assure safety, people may need to be able to take some level of control over the ship at any time. Many of the issues relating to supervision for autonomous medical systems are therefore relevant here.

One of the big challenges of such a shift to SCCs is the loss of direct ship-sense. An investigation conducted in (Man et al., 2016) highlighted how critical ship-sense is in ship manoeuvring. They consider how operators in a remote operating centre will be able to effectively perceive the ship's movements and manoeuver the ship without ship-sense since there will be no physical connection between the human and the vessel, and no directly perceived information from the ship's environment. In (Wahlström et al., 2015), an overview of the human factors challenges that might concern future monitoring and control of unmanned ships from SCCs is presented. They identify the challenges through consideration of autonomous and remote operation across a number of domains including aviation, forestry, subway systems, space and military operations, and contrast these to the maritime context. The most prominent issues they identify include information overload, boredom, mishaps during changeovers and handoffs, lack of feel of the vessel, constant reorientation to new tasks, delays in control and monitoring, and the need for human understanding in local knowledge and object differentiation (e.g., in differentiating between help-seekers and pirates).

5 Conclusion

There is significant enthusiasm for the use of artificial intelligence in healthcare as well as in other industries, and there is no shortage of promise by technology developers of how AI can transform overstretched health services and improve patient care. There is also political will to support the development of new technologies with funding and by opening up relatively closed health systems such as the NHS. On the one hand this is good news, because these developments recognise the great potential that AI technologies undoubtedly bring. On the other hand, from a safety assurance perspective there is cause for concern because the evidence base on whether and how the introduction of such technologies might impact on patient safety is very thin. Largely, evaluation studies to date have considered performance of AI on specific tasks, but have neglected the

[6] Rolls Royce video available at:
https://www.youtube.com/watch?time_continue=1&v=vg0A9Ve7SxE

wider impact on clinical systems. One way forward might be to look not at algorithms in isolation, but rather consider the services AI systems are contributing to, and how the introduction of novel technologies will change the ways in which services are provided.

Standards and guidance exist, which could form a starting point for more rigorous safety assurance of AI technologies, such as established standards for risk management of medical devices (ISO 14971) and health information technology (NHS Digital clinical safety standards). However, these standards focus predominantly on technical aspects and do not cover human factors and service issues. In addition, many of the technology developers entering the AI healthcare market do not come from a safety-critical system engineering background and might be largely unfamiliar with existing guidance and best practice.

There is an opportunity for national bodies such as the Chartered Institute of Ergonomics and Human Factors (CIEHF) and the newly established NHSX to raise awareness of human factors and safety challenges for the use of AI in healthcare, and to develop and disseminate appropriate guidance. Funding should be made available not only for the development of AI technologies, but also for their rigours evaluation to ensure we understand from the outset how AI will impact on patient care and patient safety, and how potential hazards and human factors challenges can be addressed.

Appendix

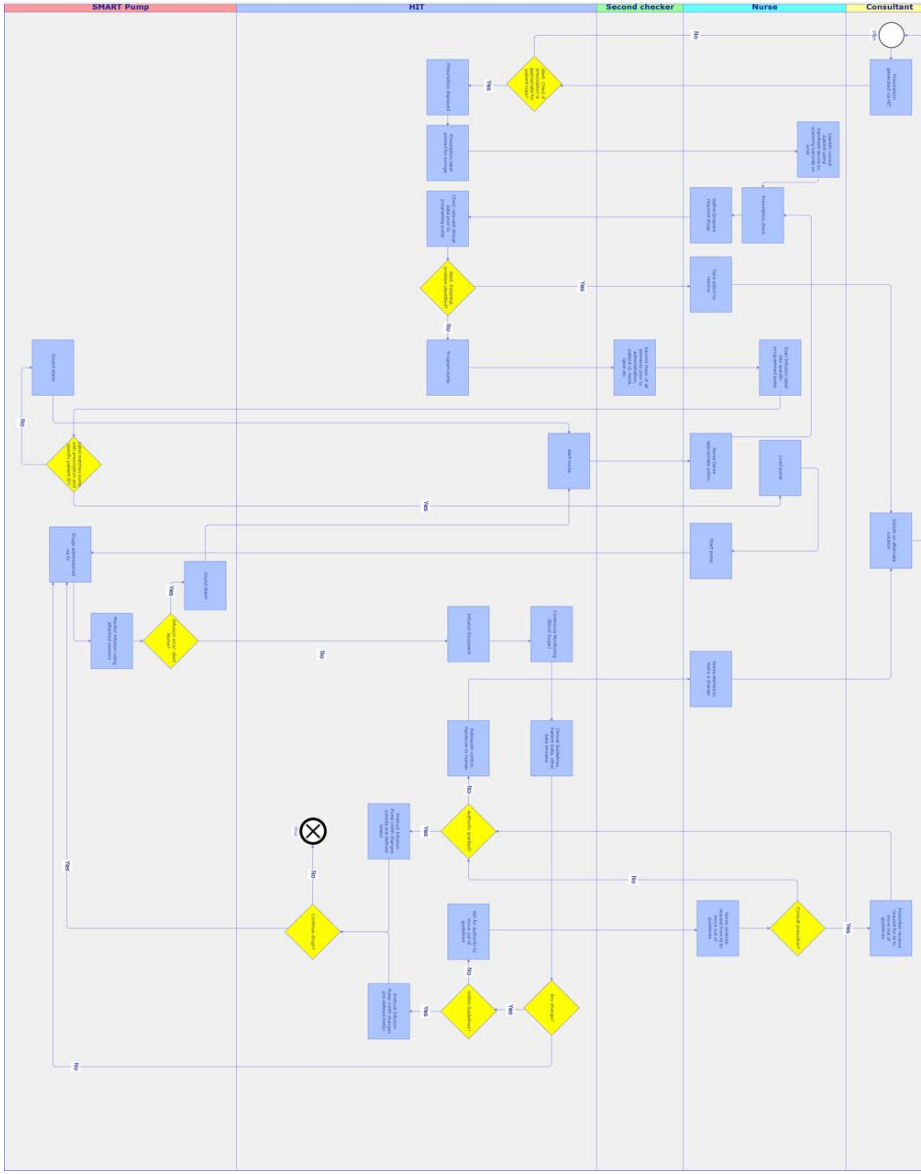

Fig. 5. Process map of infusion process with L5 autonomous infusion pump

Acknowledgments This work is supported by the Assuring Autonomy International Programme, a partnership between Lloyd's Register Foundation and the University of York. The process map in the appendix was developed by Shakir Laher using the NHS Digital SMART software. We are grateful to the clinical team at the Royal Derby Hospital (David Nelson, Matthew Elliott and Nick Reynolds) and the clinical safety team at NHS Digital (Sean White and Shakir Laher) for their contribution to the SAM demonstrator project.

References

ALBERDI, E., POVYAKALO, A., STRIGINI, L. & AYTON, P. 2004. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. *Academic Radiology,* 11**,** 909-918.

AVRAM, R., TISON, G., KUHAR, P., MARCUS, G., PLETCHER, M., OLGIN, J. E. & ASCHBACHER, K. 2019. PREDICTING DIABETES FROM PHOTOPLETHYSMOGRAPHY USING DEEP LEARNING. *Journal of the American College of Cardiology,* 73**,** 16.

BURNETT, G., LARGE, D. R. & SALANITRI, D. 2019. How will drivers interact with vehicles of the future? London: RAC Foundation.

CASNER, S. M., GEVEN, R. W., RECKER, M. P. & SCHOOLER, J. W. 2014. The Retention of Manual Flying Skills in the Automated Cockpit. *Human Factors,* 56**,** 1506-1516.

CHILAMKURTHY, S., GHOSH, R., TANAMALA, S., BIVIJI, M., CAMPEAU, N. G., VENUGOPAL, V. K., MAHAJAN, V., RAO, P. & WARIER, P. 2018. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet,* 392**,** 2388-2396.

COIERA, E. 2018. The fate of medicine in the time of AI. *The Lancet,* 392**,** 2331-2332.

ELLIOTT, R. A., CAMACHO, E., CAMPBELL, F., JANKOVIC, D., ST JAMES, M. M., KALTENTHALER, E., WONG, R., SCULPHER, M. J. & FARIA, R. 2018. Prevalence and economic burden of medication errors in the NHS in England. Sheffield: Policy Research Unit in Economic Evaluation of Health & Care Interventions.

EMBREY, D. 1986. SHERPA: A systematic human error reduction and prediction approach. *Proceedings of the International Topical Meeting on Advances in Human Factors in Nuclear Power Systems.* Knoxville, Tennessee: American Nuclear Society.

ENDSLEY, M. R. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors,* 37**,** 32-64.

ESTEVA, A., KUPREL, B., NOVOA, R. A., KO, J., SWETTER, S. M., BLAU, H. M. & THRUN, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature,* 542**,** 115.

FITZPATRICK, K. K., DARCY, A. & VIERHILE, M. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health,* 4**,** e19.

FURNISS, D., DEAN FRANKLIN, B. & BLANDFORD, A. 2019. The devil is in the detail: How a closed-loop documentation system for IV infusion administration contributes to and compromises patient safety. *Health Informatics Journal***,** 1460458219839574.

GODDARD, K., ROUDSARI, A. & WYATT, J. C. 2012. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc,* 19**,** 121-127.

GOLD, C., DAMBÖCK, D., LORENZ, L. & BENGLER, K. 2013. "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* SAGE Publications

GULSHAN, V., PENG, L., CORAM, M., STUMPE, M. C., WU, D., NARAYANASWAMY, A., VENUGOPALAN, S., WIDNER, K., MADAMS, T., CUADROS, J., KIM, R., RAMAN, R., NELSON, P. C., MEGA, J. L. & WEBSTER, D. R. 2016. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus PhotographsAccuracy of a Deep Learning Algorithm for Detection of Diabetic RetinopathyAccuracy of a Deep Learning Algorithm for Detection of Diabetic Retinopathy. *JAMA,* 316**,** 2402-2410.

HAENSSLE, H. A., FINK, C., SCHNEIDERBAUER, R., TOBERER, F., BUHL, T., BLUM, A., KALLOO, A., HASSEN, A. B. H., THOMAS, L., ENK, A., UHLMANN, L., LEVEL-I, R. S. & GROUPS, L.-I. 2018. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology,* 29**,** 1836-1842.

HOLLNAGEL, E. 2014. *Safety-I and Safety-II,* Farnham, Ashgate.

HOLLNAGEL, E. 2015. Why is Work-as-Imagined different from Work-as-Done? *In:* WEARS, R., HOLLNAGEL, E. & BRAITHWAITE, J. (eds.) *The Resilience of Everyday Clinical Work.* Farnham: Ashgate.

HOLLNAGEL, E., WOODS, D. D. & LEVESON, N. 2006. *Resilience Engineering: Concepts and Precepts,* Aldershot, Ashgate.

JETER, R., JOSEF, C., SHASHIKUMAR, S. & NEMATI, S. 2019. Does the" Artificial Intelligence Clinician" learn optimal treatment strategies for sepsis in intensive care? *arXiv* 1902.03271.

KOMOROWSKI, M., CELI, L. A., BADAWI, O., GORDON, A. C. & FAISAL, A. A. 2018. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine,* 24**,** 1716-1720.

LYELL, D. & COIERA, E. 2016. Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association,* 24**,** 423-431.

MAN, Y., LUNDH, M. & PORATHE, T. 2016. Seeking Harmony in Shore-Based Unmanned Ship Handling - From the Perspective of Human Factors, What Is the Difference We Need to Focus on from Being Onboard to Onshore? *In:* AHRAM, T., KARWOWSKI, W. & MAREK, T. (eds.) *5th Int Conf on Appl Human Factors and Ergonomics AHFE.* Krakow, Poland: CRC Press.

MCLEOD, M. C., BARBER, N. & FRANKLIN, B. D. 2013. Methodological variations and their effects on reported medication administration error rates. *BMJ Quality & Safety,* 22**,** 278-289.

MERAT, N., JAMSON, A. H., LAI, F. C. H., DALY, M. & CARSTEN, O. M. J. 2014. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour,* 27**,** 274-282.

NATIONAL TRANSPORTATION SAFETY BOARD. 2009. Accident report - Loss of Control on Approach Colgan Air, Inc. Operating as Continental Connection Flight 3407 Bombardier DHC-8-400, N200WQ, NTSB/AAR-10/01. Available: https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1001.pdf.

NATIONAL TRANSPORTATION SAFETY BOARD. 2018. Preliminary Report Highway HWY18MH010. Available: https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf.

PARASURAMAN, R. & RILEY, V. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors,* 39**,** 230-253.

SARIA, S., BUTTE, A. & SHEIKH, A. 2018. Better medicine through machine learning: What's real, and what's artificial? *PLoS Med,* 15**,** e10002721.

SEMIGRAN, H. L., LINDER, J. A., GIDENGIL, C. & MEHROTRA, A. 2015. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ,* 351**,** h3480.

STANTON, N. 2006. Hierarchical task analysis: Developments, applications, and extensions. *Appl Ergon,* 37.

SUJAN, M., FURNISS, D., EMBREY, D., ELLIOTT, M., NELSON, D., WHITE, S., HABLI, I. & REYNOLDS, N. 2019a. Critical barriers to safety assurance and regulation of autonomous medical systems. *In:* BEER, M. & ZIO, E. (eds.) *29th European Safety and Reliability Conference (ESREL 2019).* Hannover: CRC Press.

SUJAN, M., FURNISS, D., GRUNDY, K., GRUNDY, H., NELSON, D., ELLIOTT, M., WHITE, S., HABLI, I. & REYNOLDS, N. 2019b. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health & Care Informatics,* 26**,** e100081.

SUJAN, M., HUANG, H. & BIGGERSTAFF, D. 2019c. Trust and Psychological Safety as Facilitators of Resilient Health Care. *In:* BRAITHWAITE, J., HOLLNAGEL, E. & HUNTE, G. (eds.) *Resilient Health Care V: Working Across Boundaries.* CRC Press.

SUJAN, M., SCOTT, P. & CRESSWELL, K. 2019d. Digital health and patient safety: Technology is not a magic wand. *Health Informatics Journal*.

SUJAN, M., SPURGEON, P. & COOKE, M. 2015a. The role of dynamic trade-offs in creating safety—A qualitative study of handover across care boundaries in emergency care. *Reliability Engineering & System Safety,* 141**,** 54-62.

SUJAN, M. A., CHESSUM, P., RUDD, M., FITTON, L., INADA-KIM, M., COOKE, M. W. & SPURGEON, P. 2015b. Managing competing organizational priorities in clinical handover across organizational boundaries. *Journal of Health Services Research & Policy,* 20**,** 17-25.

SUJAN, M. A., CHESSUM, P., RUDD, M., FITTON, L., INADA-KIM, M., SPURGEON, P. & COOKE, M. W. 2015c. Emergency Care Handover (ECHO study) across care boundaries: the need for joint decision making and consideration of psychosocial history. *Emergency Medicine Journal,* 32**,** 112-118.

WAHLSTRÖM, M., HAKULINEN, J., KARVONEN, H. & LINDBORG, I. 2015. Human factors challenges in unmanned ship operations–insights from other domains. *Procedia Manufacturing,* 3**,** 1038-1045.

WAYMO. 2018. Waymo Safety Report - On the Road to Fully Self-Driving. Available: https://storage.googleapis.com/sdc-prod/v1/safety-report/Safety%20Report%202018.pdf.

YEUNG, S., DOWNING, N. L., FEI-FEI, L. & MILSTEIN, A. 2018. Bedside Computer Vision — Moving Artificial Intelligence from Driver Assistance to Patient Safety. *New England Journal of Medicine,* 378**,** 1271-1273.

YU, K.-H. & KOHANE, I. S. 2019. Framing the challenges of artificial intelligence in medicine. *BMJ Qual Saf,* 28**,** 238-241.