Chartered Institute
of Ergonomics
& Human Factors

WHITE PAPER

Human Factors and Ergonomics in Healthcare AI

FOREWORD

This White Paper sets out a human factors and ergonomics (HF/E) perspective on the use of artificial intelligence (AI) applications in healthcare. It represents a significant body of work, led by Mark Sujan and ably supported by a range of highly skilled professionals and international thought leaders.

Its aim is to promote systems thinking among those who develop, regulate, procure, and use AI applications, and to raise awareness of the role of people using or affected by AI.

The CIEHF Digital Health & AI Special Interest Group was set up in 2020 to describe and promote the role and contribution of HF/E in the design and use of digital technology and AI in healthcare.

The group also aims to build a community of practice, and its members include Chartered Ergonomists, AI researchers, clinicians, healthcare managers, technology developers, and individuals working in the standardisation sector.

The Chartered Institute of Ergonomics & Human Factors (CIEHF) received its Royal Charter in 2014 to recognise the uniqueness and value of the scientific discipline and the pre-eminent role of the Institute in representing both the discipline and the profession in the UK.

This includes the protected status of "Chartered Ergonomist and Human Factors Specialist" with the post-nominal C.ErgHF awarded to practising Registered Members and Fellows who are among a group of over 500 elite professionals working at a world-class level.

We are rightly proud of what this document has achieved and believe it has an important role to play as AI increases its presence and impact in the health sector.

Dr Noorzaman Rashid, CEO CIEHF

NH2
HS
OH
HN
F
NH
N
H

CONTENTS

Executive Summary

This White Paper outlines eight key human factors and ergonomics (HF/E) principles and methods relevant to the design and use of artificial intelligence (AI) in healthcare.

These eight HF/E principles can be used by the reader to reflect on their own practice and experience with AI, be that as developer of the technology, user, regulator, or research funder. The objective is to start a conversation and to raise awareness of the role of people who use or are affected by the use of AI.

AI is seen as the next IT step in addressing healthcare challenges. If done well, AI could support clinicians in their decision making, it could generate predictions aimed at improving inefficiencies in the management of care processes, and it might radically transform the way care is provided and accessed.

Diagnostic devices using AI are leading the way, e.g., AI applications interpreting radiological images to identify pneumonia or to differentiate COVID-19 from other types of chest infections. Other examples of healthcare AI applications include the use of patient-facing chatbots, mental health applications, ambulance service triage, sepsis diagnosis and prognosis, patient scheduling, planning of resources, quality improvement activities, and even the development of COVID-19 vaccines.

However, the aspiration of using AI to improve the efficiency of health systems, and to enhance patient safety, patient experience and staff wellbeing is currently weakened by a narrow focus on technology that contrasts people and AI ("human vs. machine"), and by a limited evidence base of AI in real-world use.

It is likely that the real challenges for the adoption of AI will arise when algorithms are integrated into healthcare systems to deliver a service in collaboration with healthcare professionals as well as other technology. It is at this health system level, where teams consisting of healthcare professionals and AI systems cooperate and collaborate to provide a service, that HF/E challenges will come to the fore.

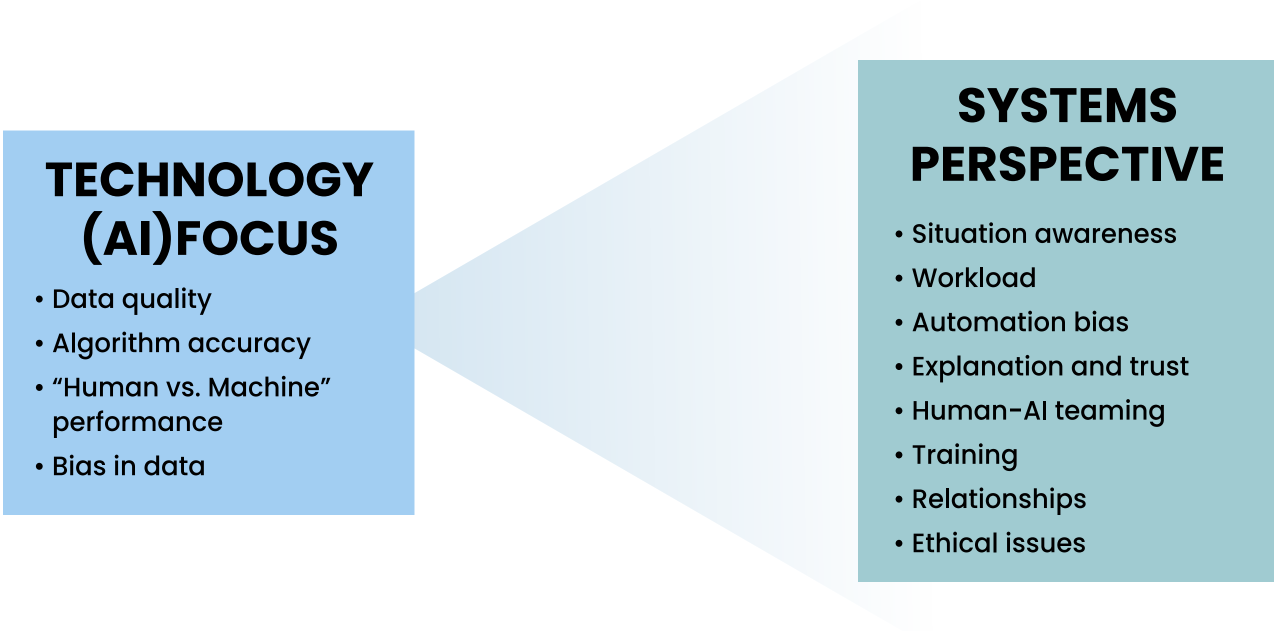
It is, therefore, important to understand the principles of human factors and ergonomics. HF/E is a scientific discipline that is concerned with the design of sociotechnical systems to improve overall system performance, safety and the wellbeing of people. From this perspective, HF/E provides theories and methods to support the design and use of AI during its lifecycle as part of the wider system.

Designers and developers of AI, individuals with responsibility for procuring AI applications, regulators, and bodies funding research and development need to move beyond the technology-centric view, and instead approach AI from a systems perspective, i.e., to consider from the outset the interaction of people with AI as part of the wider clinical and health system (Figure 1).

These activities need to be underpinned by education in and support with HF/E, which healthcare professionals and organisations can tap into.

The vision of the Chartered Institute of Ergonomics and Human Factors, and of its national and international partners, is that greater awareness and the effective application of HF/E theories and approaches in practice are key enablers for embedding AI successfully in health systems.

The stated objective of this White Paper – to start a conversation about the role of people in the design and use of healthcare AI – is modest, yet important because all of the stakeholders in the health sector need to work together in order to move forward the ambitious project of healthcare AI. However, it seems prudent to note what this White Paper does not intend to deliver: it does not aspire to be a comprehensive introduction to HF/E, nor to provide a how-to guide for practitioners. Links to further reading and useful resources are provided as bibliography, and CIEHF has launched in July 2021 a Human Factors Healthcare Learning Pathway, which provides in-depth education on the principles discussed in this White Paper.



TECHNOLOGY (AI)FOCUS

- Data quality
- Algorithm accuracy
- “Human vs. Machine” performance
- Bias in data

SYSTEMS PERSPECTIVE

- Situation awareness
- Workload
- Automation bias
- Explanation and trust
- Human-AI teaming
- Training
- Relationships
- Ethical issues

Figure 1: Broadening the scope - from technology (AI) focus to systems perspective

This White Paper has identified eight HF/E principles that should be taken into consideration in the successful use of AI in healthcare. These are:

SITUATION AWARENESS

Design options need to consider how AI can support, rather than erode, people’s situation awareness

WORKLOAD

The impact of AI on workload needs to be assessed because AI can both reduce as well as increase workload in certain situations.

AUTOMATION BIAS

Strategies need to be considered to guard against people relying uncritically on the AI, e.g., the use of explanation and training.

EXPLANATION AND TRUST

AI applications should explain their behaviour and allow users to query it in order to reduce automation bias and to support trust.

HUMAN-AI TEAMING

AI applications should be capable of good teamworking behaviours to support shared mental models and situation awareness.

TRAINING

People require opportunities to practise and retain their skill sets when AI is introduced, and they need to have a baseline understanding of how the AI works.

RELATIONSHIPS BETWEEN STAFF AND PATIENTS

The impact on relationships needs to be considered, e.g., whether staff will be working away from the patient once more and more AI is introduced.

ETHICAL ISSUES

AI in healthcare raises ethical challenges including fairness and bias in AI models, protecting privacy, respecting autonomy, providing benefits and minimising harm.

BOX 1: Artificial Intelligence and Machine Learning

The term AI, in the widest sense, refers to the science and engineering of intelligent computer systems. For the purpose of this White Paper, we focus specifically on AI applications that use machine learning (ML) for a given task. ML refers to the use of computer algorithms that learn from data, through:

(a) Supervised learning approaches, where data is labelled before being presented to the algorithm, and the algorithm aims to minimise an error function, i.e., the algorithm approximates the output and adjusts itself to minimise the difference between its own output and the correct output.

(b) Unsupervised learning, where the algorithm independently discovers patterns in the data.

(c) Reinforcement learning, where algorithms learn based on the optimisation of a reward function, i.e., the algorithm is given input data and then produces an output that can be measured (using a number, quantity or parameter) to indicate the quality of the output, which is then used by the algorithm to adjust its behaviour to maximise the measure (reward) and thus improve the quality of the output.

A frequently used ML approach is deep learning. Deep learning typically refers to the use of artificial neural networks with many layers. Artificial neural networks have been used for decades, but the increase in computing power has given rise to much larger and more deeply layered artificial neural networks in recent times.

The use of AI is backed by policy makers

Expectations for the use of AI in healthcare are high[1]. The focus has been on the potential health and economic benefits that the widespread adoption of AI can bring. This has been underpinned by the establishment of new dedicated bodies and by significant government funding to facilitate and speed up the development of the adoption of AI in health services. For example, in the UK, NHSX[2] has been set up specifically to accelerate the digital transformation in the National Health Service (NHS), and it is home to the NHS AI Lab, which is backed by £250m in government funding. In a much-noted speech in January 2020, the then Secretary of State for DHSC (the Department of Health and Social Care) set out the vision for a digital NHS, remarking that technology was essential for modern health systems to meet the challenges they are facing[3].

Health systems are struggling with rising costs, staff shortages and burnout, an increasingly elderly population with more complex health needs, and health outcomes that often fall short of expectations. Information technology (IT) has been used to help address some of these problems, such as expediting and widening knowledge and information transfer. For example, specialists can collaborate across different sites and even internationally to help manage complex needs and improve outcomes. At the same time, the introduction of IT has frequently created additional problems, such as care delivery disruption, patient harm and clinician burnout, due to poor interoperability, usability and lack of regard to implementation within a system, i.e., without due consideration of the relationship between the IT and other system elements[4].

AI is seen as the next IT step in addressing health and care challenges. If done well, AI could support clinicians in their decision making, it could generate predictions aimed at improving inefficiencies in the management of care processes, and it might radically transform the way care is provided and accessed. For example, the use of AI applications could help address the shortage of skilled clinicians e.g., by screening radiological images

to identify those that require further consideration by a human expert or by providing diagnostic decision support[5]. Similarly, AI tools using natural language processing could digest and synthesise the hundreds of thousands of papers published in the healthcare literature every year, and provide healthcare workers with up-to-date information to support the adoption and spread of evidence-based practice. AI technology could also help improve efficiency of logistical processes, for example by using AI scheduling that targets supply chain inefficiencies and resource shortages that undermine the quality of health and care, ultimately affecting patient outcomes. The aspiration is, therefore, that the successful adoption of AI technologies will contribute to increasing healthcare sustainability and outcomes.

AI is starting to be used in all areas of healthcare, with diagnostics leading the way

Another popular application domain are AI-driven chatbots, e.g., patient-facing symptom checkers to assist self-diagnosis and triage. Chatbots enable patients to describe symptoms in natural language in an interactive way, guided by questions and prompts by the AI. The AI can then advise on the most likely diagnoses and provide guidance as to the urgency of seeking specialist help. Such AI symptom checkers appear to be received favourably by patients, with one study suggesting that 90% of people surveyed felt that the symptom checker provided them with useful information[9].

Mental health is a further area that has received significant interest from the AI community. Applications have been developed that aim to identify depression, e.g., from voice interactions with the application or from posts on social media, or to deliver cognitive behavioural therapy via an AI chatbot[10].

The potential areas for the application of AI in healthcare are seemingly endless, with other promising developments being tested, for example, in ambulance service triage, sepsis diagnosis and prognosis, patient scheduling, planning of resources and quality improvement activities. Longer-term, the use of AI to enable precision medicine, i.e., treatments and health management that are data-driven and that consider variability in people's genes and personal circumstances, holds significant promise. In addition, there is much excitement about the use of AI to support drug and vaccine development, especially given recent experiences with the COVID-19 pandemic. Vaccine development typically takes years to complete, but the use of AI techniques to support the development of COVID-19 vaccines, e.g., by decoding the genetic makeup of the virus, resulted in the availability of crucial vaccines in record time.

The UK government is funding healthcare AI innovations

In the UK, the NHS AI Lab supports the piloting of AI applications through the AI Award programme[11], which has a dedicated budget of £140m over a three-year period. It covers AI applications at four different phases of development, including feasibility studies, clinical evaluation, real-word testing, and initial adoption in the NHS. In 2021, 38 AI applications across the different phases were chosen for funding via the second round of the AI award (Figure 2 on page 10). While the majority of applications are in the early phases, five applications are at the stage of initial adoption. The majority of AI applications are in the area of diagnostics, particularly at the initial adoption stage (phase 4).

The real challenge for AI is the integration into clinical reality

Evaluation studies of a number of AI algorithms have produced encouraging results, with many studies suggesting that the performance of the AI was at least as good as that of human experts, for example in the detection of skin cancer[12] or the detection of diabetic retinopathy[13]. However, looking across these studies, the focus of the evaluation is usually on the performance of the AI on a narrowly defined task. The use of AI technologies in clinical practice, such as deep learning (see Box 1), is still in its infancy, and healthcare professionals are still unfamiliar with them. The evidence base to date reflects this, and is weakened by it: the evaluation is typically undertaken by the technology

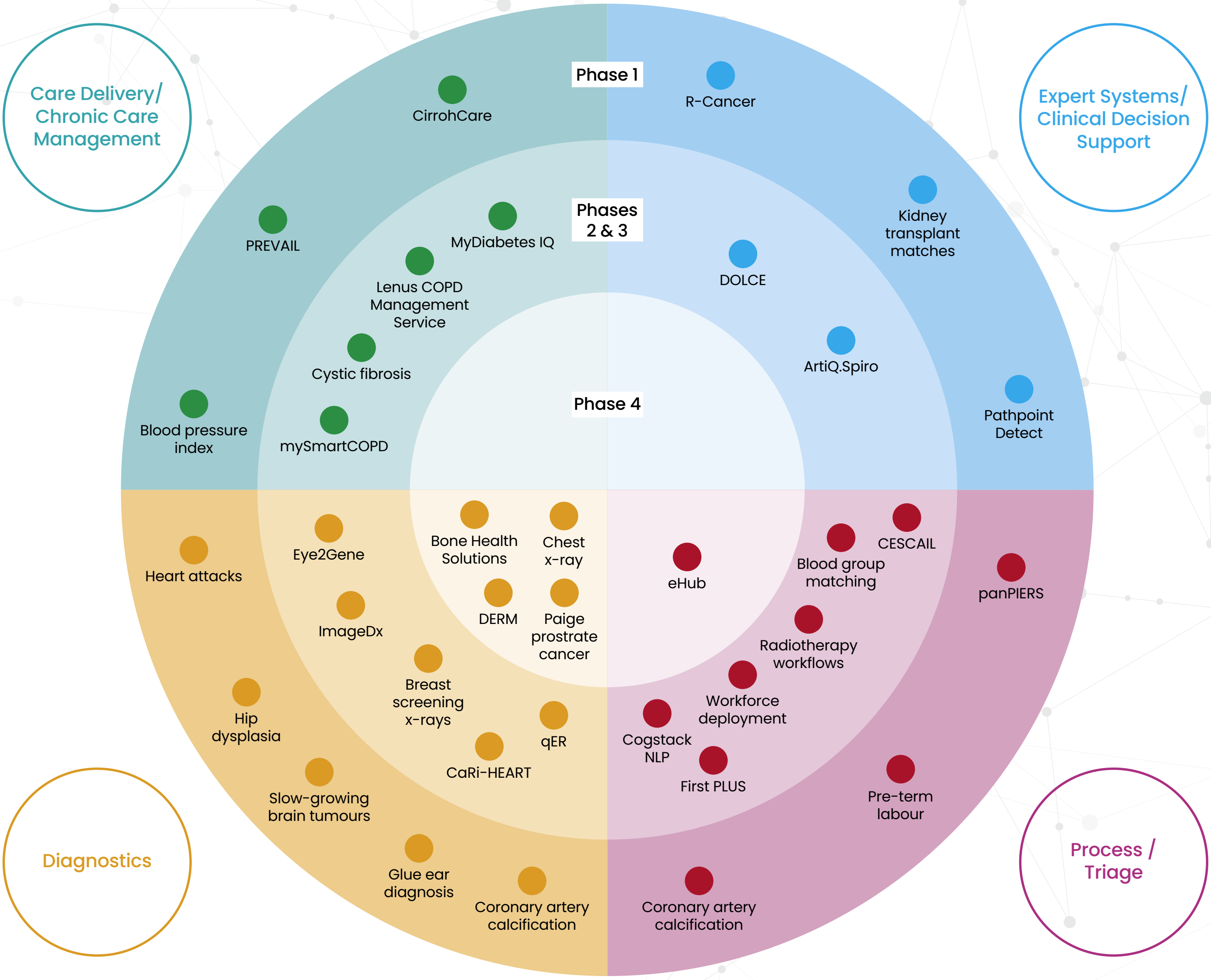


Figure 2: AI applications funded through the second AI Award 2021

developers, and independent evaluation remains the exception; the number of human participants tends to be small; and prospective trials are still infrequent. A systematic review published in the BMJ concluded that claims that AI outperforms humans are likely to be overstated given the limitations in study design, reporting and transparency, and the risk of study bias[14]. Similar results were found in a review of clinical decision support systems using machine learning, which demonstrated that there was no robust evidence that such systems improve the performance of clinicians when used to augment rather than replace human intelligence[15]. An as yet rare example of an independent external validation study found that a widely used sepsis prediction model had significantly lower accuracy than that described from the internal validation[16]. In addition, the sepsis prediction model created many false alerts, thus contributing to alert fatigue.

It is likely that 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. It is at this clinical system level, where teams consisting of healthcare professionals and AI systems cooperate and collaborate to provide a service, that HF/E challenges will come to the fore[17] (see the example in Box 2).

BOX 2: Real-world challenges of a Deep Learning algorithm for detecting diabetic retinopathy

The management of diabetes is a challenge for health systems world-wide, and diabetic eye disease is among the leading causes for vision impairment. Researchers at Google Health developed a system using deep learning to detect diabetic retinopathy. Evaluation of the system and comparison with human experts in a retrospective study found that the system has accuracy levels comparable to that of specialists. The Google Health researchers complemented their initial pre-clinical evaluation study with a noteworthy observational study in clinical environments following the adoption of the system across clinics in Thailand[18]. The findings of this study highlight many of the socio-technical aspects that emerge only once the AI is embedded in the real world. Examples include:

- High degree of variation in the process of eye screening across the different clinics
- Variability regarding the physical environment including light conditions affecting the quality of the fundus photos
- High rejection rate of images by the AI system due to inadequate image quality even though human readers were able to screen the images
- Additional workload of staff as they were trying to retake photos several times to ensure suitable image quality, and increased waiting times due to this
- Increased numbers of non-essential referrals due to the inability to process low-quality images causing frustration among patients
- Significant backlogs during periods where the internet speed dropped causing delays, patient frustration and increased stress levels among staff

BOX 3: HF/E Principles

Interactions
HF/E considers the interactions of people with other elements of the system, because outcomes are produced by these interactions rather than by any single element in isolation.

Design
HF/E aims to improve outcomes through system design rather than by telling people to behave differently.

Human well-being and overall system performance
HF/E considers a broad range of outcomes including human well-being and quality of working life (e.g., satisfaction, stress, fatigue), and system performance (e.g., safety, effectiveness).

Further information and resources are available from the CIEHF webpages (www.ergonomics.org.uk)

HF/E takes a systems view

Human factors and ergonomics (HF/E) has been defined by the International Ergonomics Association as "the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and methods to design in order to optimize human well-being and overall system performance"[19] (see Box 3).

Clinical systems or work systems are sociotechnical systems. Sociotechnical systems are systems, where people interact with one another in pursuit of shared goals, where they use tools and technology, engage in tasks, are governed by procedures and formal as well as informal rules, and where they move and work in a physical environment of specific characteristics. Work systems are situated within the context of wider professional, legal and societal rules and expectations. The multitude of people and interactions within the sociotechnical work system produce the care processes and other processes, such as cancer screening and infection prevention and control, which shape outcomes.

Understanding how clinical systems work comes from taking time to look at the elements of the system and how they interact with each other. There are several HF/E frameworks that can help adopt a systems perspective. The Systems Engineering Initiative for Patient Safety (SEIPS) is a frequently used framework that has been designed specifically for healthcare[20],[21] . SEIPS can be used to describe how the elements of a work system interact, and how these interactions deliver processes, which in turn lead to outcomes (see Figure 3).

The most recent version of SEIPS (referred to as SEIPS 3.0) puts the patient at the centre of the socio-technical system. The patient's health and wellbeing define a pathway that can span many years and that can include encounters with many different health and social care providers, each of which can be understood as a different work system within its own socio-organisational context.

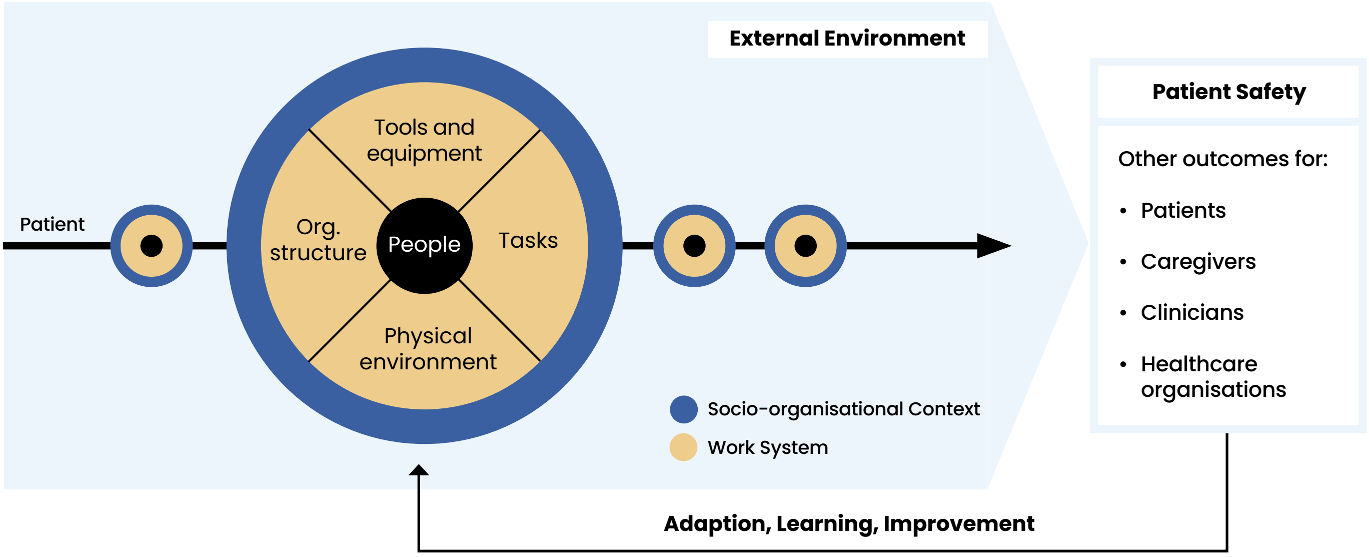


Figure 3: The Systems Engineering Initiative for Patient Safety (SEIPS) model

A broad vision for the integration of HF/E into healthcare is provided by the Model for the Integration of Ergonomics in Healthcare Systems (MIEHS)[22]. This model addresses health systems properties, such as coverage, robustness, integrity and resilience at the micro, meso and macro level.

The unit of analysis is wider than the technology

Using a systems perspective, it becomes clear that the introduction of new technology, such as an autonomous infusion pump in intensive care (Figure 4) or an AI-based ambulance call centre triage system, does not simply replace a function previously carried out by a human, but can fundamentally change the work system (Box 4).

There is a shared responsibility for the performance of healthcare systems that goes beyond front-line workers, and includes stakeholders at the supervisory, managerial regulatory and policy levels. When considering the use of AI in healthcare, this includes not only end-users, but also designers, developers, regulators and those who set and implement design guidelines and standards.

Figure 4: Simulated patient in intensive care (photo credit: Nick Reynolds)

BOX 4: HF/E approaches to study the development and introduction of an autonomous infusion pump

It has been estimated that as many as 237 million medication errors occur in England every year, and that these cause over 700 deaths[23]. Intravenous medication preparation and administration are particularly error prone. The introduction of highly automated and ultimately autonomous IV medication management systems might contribute to reducing these error rates by taking over functions previously carried out by clinicians, such as safety cross-checks (e.g., patient identity and prescription), calculating infusion rates and independently adjusting infusion parameters based on the patient's physiology. A large UK-US study found that whether or not infusion technology successfully improves patient safety depended largely on the specific context of implementation within the clinical system[24].

Systems approach

Interviews with patients, healthcare professionals, and individuals with responsibility for procurement, IT integration and training can provide rich insights into the work system. Taking a systems approach helps to anticipate and explore potential implications for the wider clinical system:

- Will clinicians' skills related to medication administration be affected?
- What new skills do clinicians require, e.g., how to tell if an autonomous infusion pump is working correctly?
- How will the relationship between clinicians be affected? The autonomous system replaces the practice of double checking by a second nurse, which often serves also as an opportunity for teaching and discussion.
- How will the relationship between patients and clinicians be affected? The use of autonomous infusion pumps could provide nurses with more time to spend with patients – or nurses might be spending more time managing and supervising autonomous systems away from the patient's bedside.
- How will the autonomous system interact with other systems, e.g. other autonomous infusion pumps or the electronic health record, and what will be the impact on the overall IT infrastructure (e.g., in case of failures)?
- What is the impact on the medication administration task, e.g., does the autonomous system reduce clinician workload by taking over parts of the task or does it increase workload, e.g., due to monitoring and administration requirements?
- How does the autonomous system impact clinician situation awareness, if clinicians do not manage infusion settings by themselves any longer? Is the autonomous system able to exchange situation awareness with clinician? Can clinicians easily tell what the system is doing and what kind of situation awareness it has?
- What is the impact on the perception of job roles, e.g., on the nursing role? Will nurses be regarded as autonomous clinicians who manage and supervise autonomous infusion pumps potentially away from the bedside, or will nurses' roles change towards more personal caring task with less responsibility and authority for managing medications?

Modelling the socio-technical system

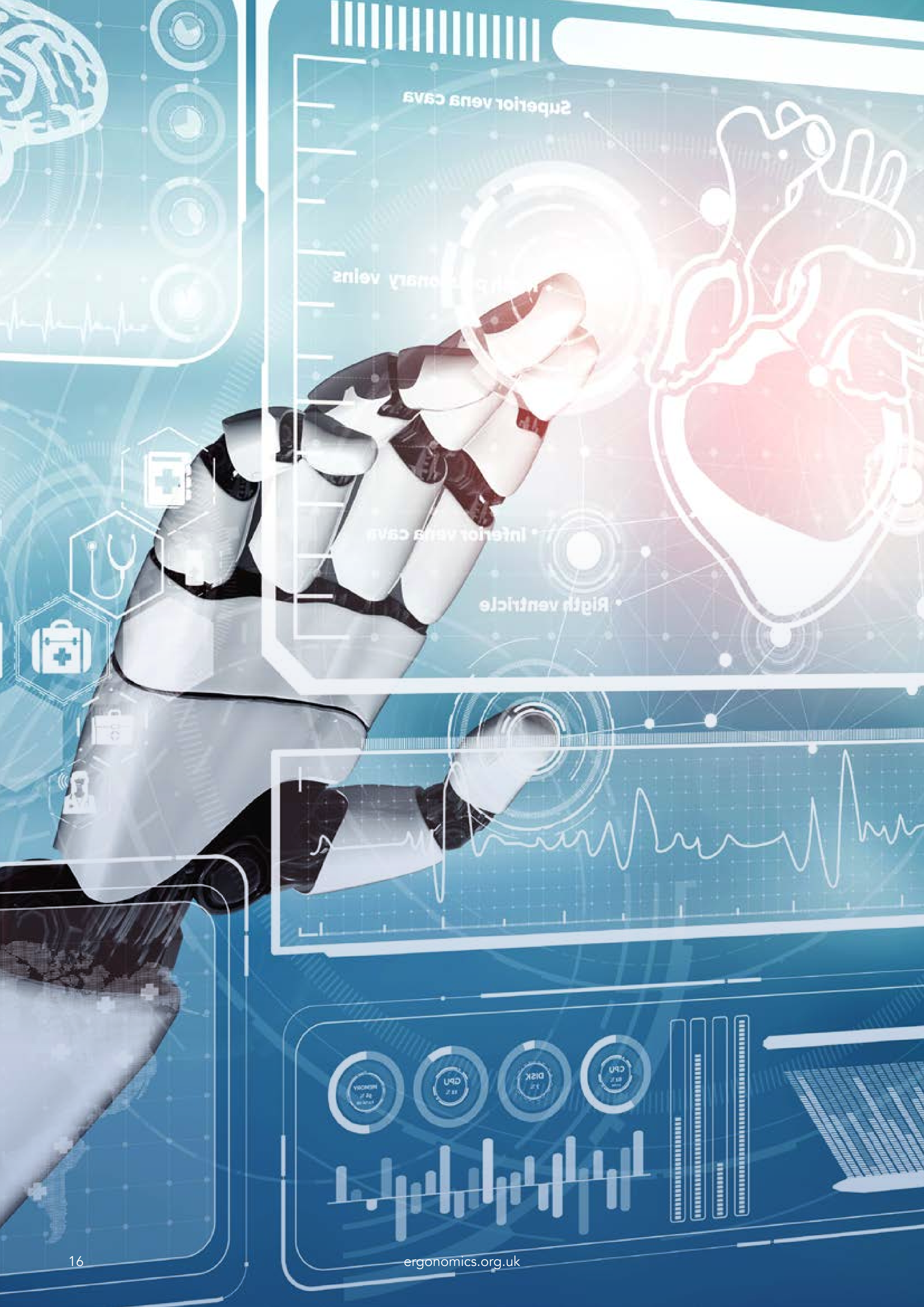
Infusion pumps are used in many different settings, such as intensive care units and emergency departments, and every setting (even, for example, two intensive care units) has its own unique characteristics. The definition of the operating environment can, therefore, be challenging for developers of AI, and usually this results in the application domain being bounded very narrowly during the development of the AI. A more fruitful approach is to define the operating environment as the clinical system within which the AI will be used (e.g., interactions with people, other tools and systems, other tasks that might be relevant, characteristics of the physical work environment etc.). It is important that the definition of operating scenarios is done based on operational realities (i.e., how work is actually carried out, or "work-as-done") rather than through an abstract view of what should be done in principle (how work has been designed or specified, or "work-as-imagined"). The HF/E toolbox offers many different methods that have been developed for understanding and representing work-as-done, such as task analysis (TA) methods[25] (Box 10), work domain analysis (WDA)[26], and systems-theoretic process analysis (STPA)[27]. A relatively recent approach is the Functional Resonance Analysis Method (FRAM)[28], which supports the analyst in reasoning about interactions in everyday clinical work[29]. For example, the use of FRAM encourages consideration of not just the algorithmic performance (e.g., whether the infusion pump can control a patient's blood sugar levels by giving insulin), but also of how the autonomous infusion pump communicates with nurses and doctors as well as other systems, such as the electronic patient record. This provides a more realistic representation of the complexity of the operational environment in healthcare settings.[30]

HF/E emphasises human-centred design, co-design and co-production of systems

Established HF/E design principles and standards aim to ensure that technology can be integrated meaningfully into the complex operational reality of healthcare systems. Human-centred design is a process for developing interactive systems that focuses on the users and their needs and requirements, and which includes HF/E approaches and methods to make systems usable and useful. Human-centred design processes[31] enhance effectiveness and efficiency, improve human wellbeing and user satisfaction, accessibility and sustainability, and counteract possible adverse effects of use on human health, safety and performance.

Human-centred design starts with the identification of stakeholders and their needs, and then actively involves the stakeholders in the co-design and co-production of technology and systems. Co-design and co-production are based on the partnership of people who work within the system, people who have lived experiences of using the system (i.e., patients and their families or carers), and the designers of technologies and other elements of the work system, making best use of each other's respective knowledge, resources and contributions[32].

An important aspect of co-design and co-production is the development of measures to determine what success would look like from the perspective of the multiple stakeholders involved. For the NHS in England there exists a "guide to good practice for digital and data-driven health technologies[33]" that sets out principles for determining the value proposition and concomitant outcome measures for healthcare organisations, staff and patients. Ensuring a range of measures to determine the value of AI to healthcare will not be easy, but HF/E can play a role here in helping to understand the interplay of components of the work system, and mapping out outcome, process and balancing measures, i.e., any potential knock-on effects of AI on task performance, social relationships among staff, relationships with patients and their families, trust, roles and responsibilities and work processes.


Superior vena cava
Inferior vena cava
Rigth ventricle

The use of AI presents challenges both old and new

When automation started to be deployed at scale in industrial systems in the 1970s and 1980s, HF/E research on "automation surprises" and the "ironies of automation"[34] explained some of the problems that appeared following the introduction of automation (see Box 5). The fundamental fallacy is the assumption that automation might replace people, but in reality the use of automation changes what people do. Fast forward to the present day, we can identify the potential for similar ironies in the current approach to the development and deployment of AI applications in clinical systems. There is a danger that developers and decision-makers adopt reductionist and simplistic perspectives that "human errors" can be designed out with the introduction of more reliable AI systems that replace human tasks.

However, "intelligent" systems also present completely new challenges that were not as relevant in the design of traditional automated systems. Intelligent agents have the ability to augment what people do in ways that were

BOX 5: The fatal Uber/Volvo crash - classic ironies of automation in a modern context

In 1983, Lisanne Bainbridge published a short paper describing some "ironies of automation", which went on to become one of the most influential contributions in the area of human – machine interaction. Bainbridge argued that as a by-product of the introduction of automation as a means to reduce reliance on humans, people are still left to perform those tasks, which designers could not work out how to automate, and are expected to manage those situations, which were not foreseen or where the automation fails. This can result in several additional, potentially critical ironies:

- When people need to intervene and take over from the automation, they will most likely need to apply precisely those skills, which the automation had been taking over from the human, i.e., people will be less practised at those skills, and they will have to respond under time pressure.
- People build their understanding of different situations and strategies for their effective management by performing the relevant tasks and by receiving feedback on their effectiveness. When people do not have these opportunities anymore, they will have less knowledge about a situation when they must take over from the automation.
- Automation frequently shifts the role of the human to that of monitor and supervisor of automated systems. People need to be able to determine, therefore, what the correct and expected behaviour of the automation is. However, considering that the automation was introduced to overcome human limitations, this monitoring task might be difficult to achieve. In addition, people are very poor at continuous monitoring and are prone to monitoring fatigue.

In March 2018 an Uber self-driving Volvo XC90 test vehicle hit and killed a pedestrian pushing a bicycle across the road. The vehicle initially failed to correctly classify the pedestrian with a bicycle and project their path as a potential collision risk. Self-driving vehicles have a safety driver[35] who has at all times the responsibility to take over from the automation in case of anomalies and safety risks. In this instance, the safety driver was using their phone to watch a TV show, and did not have their eyes on the road, while the automation failed to alert the driver to the obstacle in the road ahead[36]. This is a tragic illustration of Bainbridge's ironies at play in modern systems. Peter Hancock concluded that "if you build vehicles where drivers are rarely required to respond, then they will rarely respond when required."[37]

BOX 6: AI can support different stages of human information-processing

Information acquisition
AI supports or takes over sensing and recording of input data. Data might be organised and highlighted, but raw data is preserved.

Information analysis
AI supports or automates the integration and analysis of data, e.g., for the purpose of prediction.

Decision selection
AI offers decision options or decides independently.

Action implementation
AI supports the implementation of an action or executes it independently.

not possible when machines simply replaced physical work. The interaction with interconnected AI-based systems could potentially develop more into a relationship between people and the AI, especially where the AI has means of expressing something akin to a personality via its interfaces[38]. Social aspects will become much more relevant, as well as mutual understanding of expected behaviours and norms. We might think of, for example, the seemingly ubiquitous voice-enabled virtual assistants (e.g., Amazon's Alexa or Apple's Siri) that aim to deliver a realistic and natural social interaction experience. Examples from the health and social care sector might include mental health chatbots and assistive robots. These relationships between people and technology, along with social, cultural and ethical aspects have much greater importance for AI-based systems than for traditional automation. Healthcare professionals, patients and AI will increasingly collaborate as part of the wider clinical system.

There are different design options for embedding AI in clinical systems

In the automation literature taxonomies of levels of automation have been proposed. The purpose of these taxonomies is to help designers determine which functions should be automated and to what extent[39]. These levels of automation range from no automation through to full automation without human input, via intermediate stages. The main utility of such taxonomies is that they highlight how different choices of automation can affect human performance by impacting on, for example, workload, situation awareness, trust and reliance, and skills. From a systems perspective, criteria to determine the scope of automation need to be wider than just technical feasibility

Levels of automation taxonomies can be combined with models of human information processing, which break down human activity into a number of stages, such as information acquisition (i.e., perception and sensing), information analysis (i.e., understanding and sensemaking), decision selection (i.e., weighing up options, constraints and consequences), and action implementation (Box 6). This can help practitioners think about different types of human – AI interaction, e.g., in the case of medical devices using machine learning (Box 7).

Such simple classifications might be helpful to derive insights about how different levels of automation can affect human performance under different situations (Box 8).

Recent thinking on human-centred AI extends the one-dimensional levels of automation taxonomies by introducing separate levels of autonomy for people and AI. This creates a two-dimensional autonomy matrix. The point of this approach is to emphasise that increasing the autonomy of AI does not necessarily have to diminish the autonomy of people. Instead, AI might augment what people do, in a way granting autonomy to both humans and AI rather than either or[43].

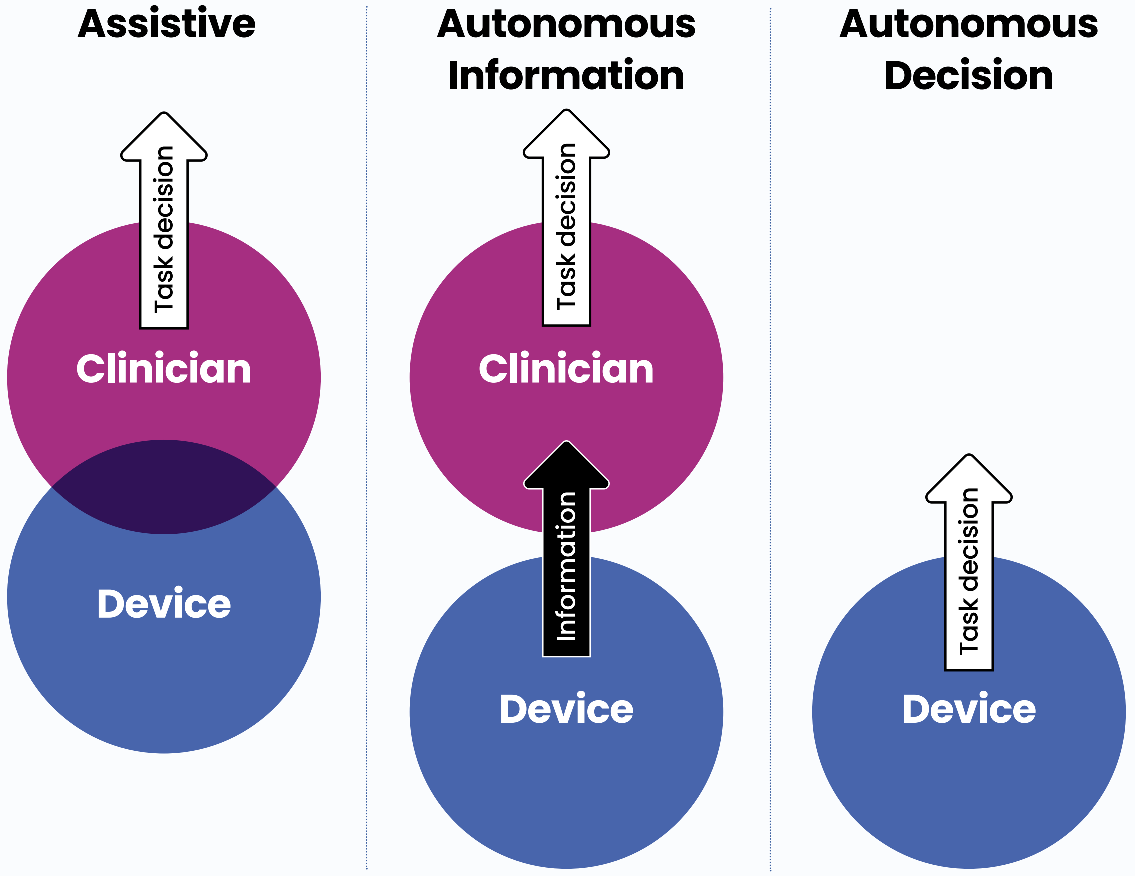
BOX 7: A framework for classifying ML-based medical devices into different levels of autonomy[40]

The combination of stages of human information processing and of levels of automation can be used to derive simple classification frameworks for AI, e.g., for medical devices using machine learning.

Assistive devices
characterised by an overlap between the device and clinician. For example, for breast cancer screening, both identify possible cancers, however clinicians are responsible for making decisions on what should be followed up, and therefore must decide whether they agree with AI marked cancers.

Autonomous information
characterised by a separation between what the device and the clinician each contribute to the activity or decision. An example is an ECG that monitors heart activity, interprets the results, and provides the information, such as quantifying heart rhythm, which clinicians can use to inform decisions on diagnosis or treatment.

Autonomous decision
where the device provides the decision on a clinical task that can be enacted by the device or the clinician. An example is the IDx-DR diabetic screening system in the US that can detect diabetic retinopathy. General practitioners can act on positive findings and refer those patients to specialists for diagnosis and treatment, without having to interpret retina photographs themselves.



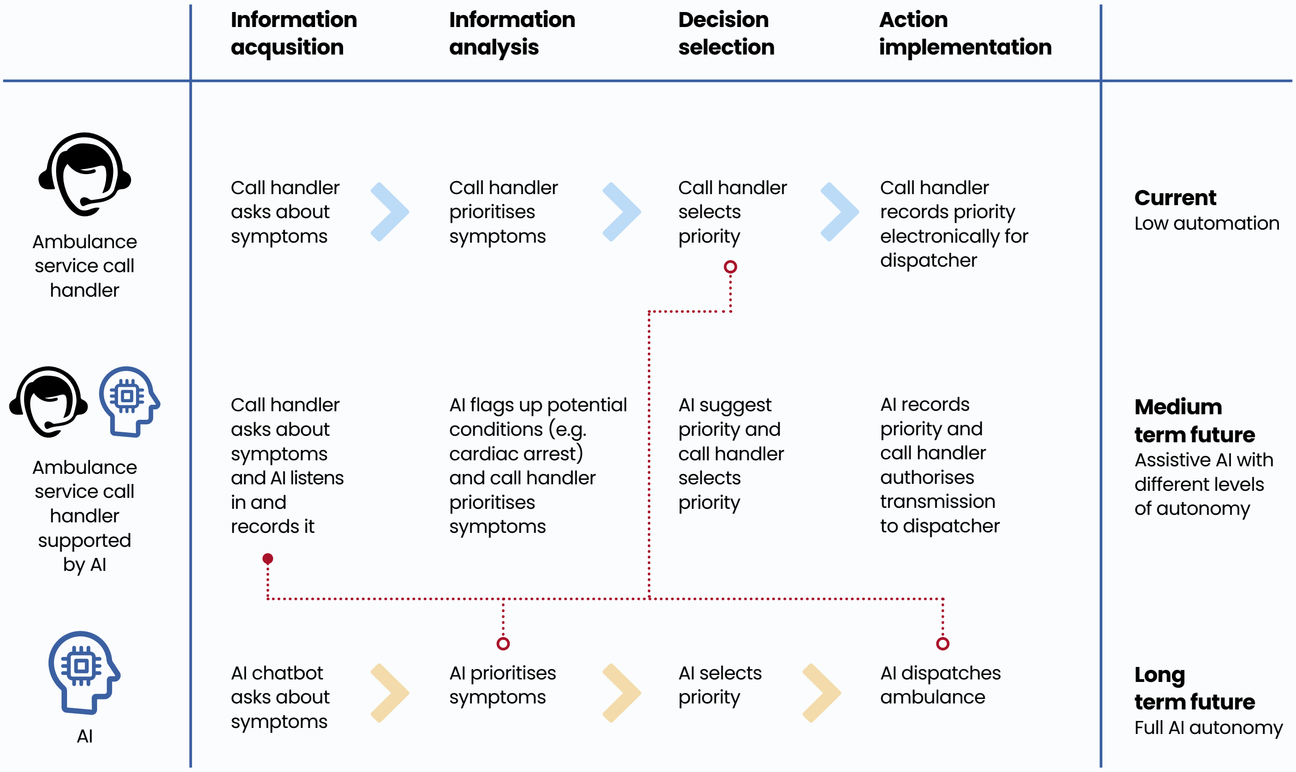
LEVELS OF AUTOMATION

BOX 8: Different design options for embedding AI in the ambulance service call centre

Ambulance service call handlers provide first-line contact to the public and ensure safe and timely response of emergency services. Recognition of critical conditions, such as cardiac arrest, is vital so that cardiopulmonary resuscitation can be delivered immediately (e.g., bystanders receiving telephone instructions), and delays to the arrival of ambulance crews can be minimised. However, recognition of cardiac arrest is difficult, and a significant number of cases are missed.

The introduction of AI could help to improve recognition rates of critical conditions[41]. There are different design options that could be considered.

Call handlers do not currently have AI support, i.e., the level of automation is low. In the longer-term future, a potential scenario might be to have full AI autonomy, i.e., the level of automation across all stages of information processing would be high[42]. More realistic scenarios in the medium term could consist of different levels of automation for each of the information processing stages. One example might be to have AI assistance during the call when asking the caller questions (i.e. information acquisition) while the AI records and transcribes it; to have AI autonomy in prioritising symptoms (i.e., information analysis), but then to have the AI only suggesting a priority (i.e., decision selection), which the call handler can review and accept or reject accordingly; and finally the AI might autonomously implement the action and dispatch an ambulance.



	Information acqusition	Information analysis	Decision selection	Action implementation	
Ambulance service call handler	Call handler asks about symptoms	Call handler prioritises symptoms	Call handler selects priority	Call handler records priority electronically for dispatcher	**Current** Low automation
Ambulance service call handler supported by AI	Call handler asks about symptoms and AI listens in and records it	AI flags up potential conditions (e.g. cardiac arrest) and call handler prioritises symptoms	AI suggest priority and call handler selects priority	AI records priority and call handler authorises transmission to dispatcher	**Medium term future** Assistive AI with different levels of autonomy
AI	AI chatbot asks about symptoms	AI prioritises symptoms	AI selects priority	AI dispatches ambulance	**Long term future** Full AI autonomy

●····○ Designers are free to “pick and mix” levels of automation, e.g., they might choose full automation for information analysis and action

AMBUL

A systems perspective (e.g., the SEIPS model) and the levels of automation taxonomies are useful tools to help designers and decision-makers understand better the potential impact of using AI in healthcare, and the options that are available regarding the design of interactions. In this section we draw particular attention to some core HF/E principles that are relevant when considering the design, implementation and use of AI: **situation awareness, workload, automation bias, explanation and trust, human – AI teaming, training, relationships and ethical issues.**

Situation awareness (SA)

Healthcare workers need to have a good understanding of the ongoing situation, including patients, their needs and specific circumstances, their past medical history and their current treatment plan, which then informs decisions about future actions and management of the patient journey. This dynamic understanding of the ongoing situation is referred to as situation awareness (SA). SA is also important for patients, e.g., their health literacy, their understanding of the need to share information, escalate concerns, ask questions and their ability to provide informed consent. Consideration of the impact on SA is important when designing, implementing and using AI. This includes:

- What SA requirements people as well as AI systems have (e.g., what each needs to know to operate effectively);
- How people and AI systems share SA-related information with one another (i.e., what information, when and in what format); and
- How people can understand what the AI is doing and what SA it has.

BOX 9: The 3-level model of situation awareness applied to anaesthesia

Anaesthesia is a highly dynamic, complex and safety-critical domain. AI has been used to support depth of anaesthesia monitoring, control of anaesthesia delivery, event prediction, ultrasound guidance, and pain management[45]. It is vital for patient safety and patient outcomes that anaesthetists have good SA[46]:

LEVEL 1
Perception of elements in current situation
The anaesthetist needs to have awareness of the patient's vital signs (e.g., heart rate, breathing rate, oxygen saturation), test and investigation results, treatments given, and the actions of other team members.

LEVEL 2
Comprehension of current situation
The anaesthetist needs to integrate and synthesise the data to enable them to explain what is going on, e.g., a sudden change in vital signs.

LEVEL 3
Projection of future status
The anaesthetist can anticipate likely future developments based on this understanding, such as the patient's likely physiological response to certain drugs and dosages.

More formally, SA has been defined as a state of knowledge derived from a dynamic process at three ascending (but not necessarily linear) levels, which includes perception of relevant data and cues (Level 1), comprehension of their meaning (Level 2), and projection into the future (Level 3) to enable decision making and selection of actions based on this understanding[44] (Box 9). SA can be affected by work systems elements, such as individual experience and training (people), the task and the physical environment, organisational factors that give rise to different levels of stress, workload and fatigue, and, of course, the technologies that people use.

More recently, the distributed situation awareness (DSA) model emphasises the systems perspective on SA[47]. According to the DSA model, SA is distributed around the sociotechnical system, and is built through interactions between agents, both human (e.g., clinicians) and non-human (e.g., AI systems). Critically, the SA required for successful performance is not held by any one agent alone, and different agents have different views on the same situation, which need to be compatible. A central component of the DSA model is, therefore, the assertion that non-human agents, such as AI, can be considered to be situationally aware. This has significant design implications, not least that the SA requirements of the AI need to be explicitly considered as well as how the AI can share SA-related information with both people and other technologies.

Situation awareness is developed and updated dynamically through SA transactions. These exchanges can be human-to-human, human-to-non-human, and non-human-to-non-human (e.g., interaction between a monitoring device, such as an ECG, and the AI). For example, human agents transact with other human agents via communications and non-verbal signals. Technological agents transact with human agents through displays, alarms and warnings, text-based communications, signs, symbols and other aspects relating to their state. Human agents transact with technological agents through data input and technological agents transact with each other through data transfer and communication protocols. How SA is exchanged between people and AI, and between AI and other systems is thus a critical design consideration. In turn, this requires an understanding of the SA requirements of all components of the clinical system.

The use of AI applications can both enhance and erode SA. Often, the introduction of automation (including AI) aims to reduce the user's workload (see next section), but also reduces SA[48]. One reason for this is that the user becomes removed from the primary information-processing loop, so is presented with a summary of information in the form of the AI recommendation. This can make it difficult to appreciate the context for this recommendation and the factors contributing to it (see "Explanation and Trust" section). Similarly, when asked to work in a supervisory role, people can easily divert their attention to other tasks, or their workload can be so low that their SA is diminished. An example of this is the Uber-Volvo accident described in Box 5, whereby the safety driver was engaged with another, not driving-related task on their mobile phone, and hence was unaware of the pedestrian crossing the road ahead.

It is, therefore, important to assess systematically the potential impact of different types of AI design. Challenges to SA due to inappropriate AI design choices include (see also subsequent sections) increases in workload due to additional data requirements and complexity, and the potential for overreliance on the AI, which can lead to poor monitoring of important issues. In addition, while AI applications and autonomous agents can achieve impressive accuracy figures, it is not straightforward to determine what the AI or autonomous agents should communicate to clinicians, and how, during normal operation to enable the clinician to maintain SA. This cannot be addressed by looking simply at one AI application in isolation, because clinicians might be interacting with many applications (e.g., multiple autonomous infusion pumps) concurrently, and the design of communication has to consider human information needs and limitations.

Design options for AI need to be explored specifically with SA and SA requirements of each agent in mind. This requires good understanding of the domain, the work system, and the task requirements. Such an understanding can be achieved through task analysis[49] (Box 10), especially methods that focus on cognitive aspects, such as cognitive work analysis or applied cognitive task analysis[50]. This type of analysis might help anticipate potential problems when AI is embedded into a clinical system, and it can help explore the effects of different levels of AI support and autonomy.

BOX 10: Task Analysis

Task analysis (TA) is a HF/E framework or process that can be used to understand and represent what people do (or are supposed to do) in order to achieve the overall goals, with a view to identifying problems and proposing potential improvements. TA has been a cornerstone of HF/E for decades and is used in many different work contexts, including designing computer interfaces for air traffic controllers, ensuring safe staffing levels in control rooms in the nuclear and petrochemical industries, improving ambulance dispatch, reducing errors in maintenance tasks and many more. With the introduction of automation and the increasing complexity of socio-technical systems, novel TA methods have been developed that place greater emphasis on cognitive aspects of work. TA can be used to identify SA requirements in terms of what kinds of information system actors need, and how this information should be integrated to support decision-making. There are over 100 different TA methods, many of which are tailored to fit specific purposes. At their core, all of the TA methods share a similar structure or process, which includes data collection of work tasks and their demands, representation in a format suitable for subsequent analysis, and then analysis of the tasks and the development of suggestions for improvement.

BOX 11: Different types of SA assessment approaches

SA requirements analysis
Methods used prior to SA studies to determine what SA should comprise during the task under analysis and therefore what should be assessed during the study.

Freeze-probe-recall methods
he most popular form of SA assessment method, freeze-probe-recall methods involve freezing simulations of the task under analysis and using pre-defined probes to assess participants SA immediately prior to the freeze.

Real-time probe methods
Real-time probe methods involve querying participants in real time regarding their awareness of task-related information (without freezing the situation).

Observer-rating methods
Methods involving the use of Subject Matter Experts (SMEs) to rate participants' SA based on observing their behaviours during task performance.

Performance measures and process indices
Methods that involve measuring SA-related aspects of performance or assessing the cognitive processes involved in developing and maintaining SA (e.g., eye tracking, concurrent verbal protocols).

Team SA measures
Methods that involve specifically the assessment of team SA during task performance.

Systems SA measures
Methods that attempt to assess DSA, with the overall system taken as the unit of analysis..

Post-task subjective rating methods
Methods that are administered post-task and involve asking participants to provide subjective ratings of their own SA during the task under analysis.

AI applications, and especially autonomous agents, need to build some form of SA, too. An autonomous infusion pump needs to know if the patient receives other medications that might affect the patient's physiology and response. These medications might come via other infusion pumps or they might be given by the clinician. The saying "if it's not documented, it didn't happen" applies here with critical consequence: if there are relevant activities going on that are not documented and communicated to the AI (e.g., autonomous infusion pump), then as far as the AI is concerned, these literally did not happen, because the application has no way of knowing about it. The results could be catastrophic.

There are different methods to assess and to measure (human) SA[51] (see Box 11). These methods should be applied throughout the design lifecycle of a product or service. A frequently referenced method is the situation awareness global assessment technique (SAGAT). This approach aims to measure SA directly during simulations of relevant tasks or scenarios. The simulation scenario is interrupted (frozen) at different points in time, and participants are then probed for their SA using questions relating to each of the three SA levels, based on an initial SA requirements analysis. SAGAT requires an in-depth development of simulation scenarios, but can provide useful insights about the impact of different technological interventions on SA.

Workload

The intention when introducing AI applications is often to reduce workload of healthcare professionals. An example might be the use of an AI chatbot and symptom checker, which can elicit information from patients, record, analyse and summarise this for the general practitioner (GP), before the patient has their appointment with their GP. This can support the GP in getting an overview quickly and help them in focusing on the most likely diagnosis.

However, the use of AI can also lead to situations where workload is increased unintentionally. This is a well-known phenomenon familiar to those who have worked with electronic health records, many of which have a reputation for imposing significant administrative burden (see Box 12). AI applications require data, and applications that cannot sense and perceive independently need to be fed these data – in many cases data will have to be input manually. AI applications can also increase workload (and frustration) when they are difficult to use. A mundane example familiar to anyone who possesses a voice-enabled AI virtual assistant on a smart speaker is the time-consuming interaction with such devices when engaging in anything other than simple questions or instructions. We can anticipate the frustrations that might arise from using voice-enabled AI applications in settings such as operating theatres and intensive care units.

Workload can be defined as the demand placed on a person when carrying out an activity. For the purpose of assessment, workload is usually broken down into physical workload and mental (or cognitive) workload. The latter is particularly relevant for many types of activities that take place in healthcare settings, e.g., monitoring of patient parameters, formulating diagnoses and developing treatment plans. Workload is an experience, i.e., it is dependent on the person. This means that the same situation can be experienced differently in terms of workload by different people, e.g., by a novice as opposed to an experienced individual. The level of mental workload experienced by people is a key element in the safety, reliability and efficiency of socio-technical systems[53]. Inappropriate levels of workload (both too high and too low) can contribute directly to errors, and cause stress and fatigue. Increased levels of stress and fatigue over prolonged periods of time can adversely affect staff health and wellbeing.

There are many methods available to assess and measure workload in different situations[54] (Box 13). These can be usefully applied to assess the impact of different design options when embedding AI in clinical systems. It is important to consider the different users and stakeholders, and the full range of tasks they might be engaged in.

BOX 12: 4,000 clicks - physician interaction with electronic medical records

A research team in a community hospital emergency department in the US set out to assess physician productivity when using an electronic medical records system[52]. The study found that physicians can spend around 40% of their time on data entry, and less than 30% on direct patient care. The research team estimated that during a busy 10-hour shift, physicians might perform as many as 4,000 mouse clicks. The findings highlight risks around data entry errors, repetitive stress injuries, eye strain and muscle pain, as well as mental fatigue from prolonged repetitive point-and-click tasks.

BOX 13: Types of workload assessment approaches

Time-based workload assessment
A task inventory is developed, and the time required to carry out individual activities is recorded or estimated. This can be compared with the time and resources available (e.g., whether activities need to be carried out concurrently.

Task performance assessment
The performance of carrying out the primary activity (e.g. making a diagnosis) is measured when secondary activities are introduced (e.g., doing data entry).

Subjective workload assessment
Staff are asked about their experience of workload when carrying out activities. This category of techniques is the most frequently used, and several well-known techniques exist, e.g., NASA Task Load Index (NASA TLX), which uses six dimensions (mental demand, physical demand, temporal demand, performance, effort, frustration).

Psycho-physiological assessment
Physiological measurements of operators are taken while they carry out an activity. Measurements might include heart rate and metabolic rate.

Automation bias

Automation bias (or automation-induced complacency) describes the phenomenon that people tend to trust and then start to rely on automated systems uncritically[55]. For example, we know from recent accidents involving partially automated cars, as well as from simulator studies, that people can come to rely on the autopilot, and spend their time on their smartphones rather than monitoring traffic[56],[57]. Long before the advent of self-driving cars, the aviation sector suffered accidents, which were attributed to the failure of pilots disengaging the autopilot in case of malfunction because they were not monitoring important flight indicators any longer. Such attribution after an accident is always subject to hindsight bias, and it is, arguably, not particularly helpful to regard automation bias as the cause of an accident. It is more productive to be attentive to patterns and potential breakdowns in the interaction between people and automation, and to consider this in the design. Similar situations might arise in healthcare, for example in radiology where the introduction of AI could lead to overreliance on the AI interpretation of radiological images. This has been observed in studies that looked at traditional clinical decision support systems[58].

AI applications will usually be advertised as having ultra-high reliability, and it is to be expected that in due course healthcare professionals will come to rely on these systems. This brings with it the risk of automation bias. Studies on automation bias suggest that the accuracy figures of AI applications in isolation do not allow prediction of what will happen in clinical use, when the clinician is confronted with a potentially inaccurate system output[59].

It is not only important that clinicians are informed about the accuracy of algorithms, but that the design supports the interaction and coordination between people and AI. Strategies that can be exploited include design for situation awareness (see corresponding section), explanation of AI decision making (see section on Explanation and Trust), and provision of training about potential limitations of AI and what to look out for (see section on Training).

Explanation and trust

The Artificial Intelligence Committee recommended strongly to the UK parliament that AI applications provide satisfactory explanation of their decisions: "… it is not acceptable to deploy any artificial intelligence system which could have a substantial impact on an individual's life, unless it can generate a full and satisfactory explanation for the decisions it will take…. In cases such as deep neural networks, where it is not yet possible to generate thorough explanations for the decisions that are made, this may mean delaying their deployment for particular uses until alternative solutions are found"[60].

While the AI research and development community is rightly concerned with making algorithms explainable, there is limited agreement on how best to achieve this. One approach is to use algorithms that produce models, which are inherently interpretable. These are usually simpler, linear models, and they might not perform as accurately on a wide range of tasks compared with algorithms that produce more complex models, such as deep learning methods. Making complex, inscrutable models intelligible to people involves mapping the complex model to a simpler model for the purpose of explanation.

Explanation is often implemented as a static post-hoc account of how a decision by an AI system was made, e.g., by highlighting which features in the data the algorithm used. This assumes that the AI is capable of abstracting and articulating lower-level features in the data into something higher-level that is meaningful to people. In addition, AI algorithms tend to work on the basis of correlations within datasets, while human explanations tend to work on the basis of causation[61]. It is well-known that correlation does not equal causation, and yet humans might misinterpret the output of AI as "causal", i.e., to assume that, given an output, the AI is proposing that there is a belief that x causes y - when, in reality, it is only pointing out that there is a relationship between these two.

For these reasons, explanation might be more helpfully thought of as a social process, and as a conversation between different agents[62]. This means that (a) explanation needs to be interactive allowing for dialogue and questions, and (b) what is considered a sufficient and satisfactory explanation will depend on the context and the specific circumstances of the actors (e.g., a patient might benefit from a different type of explanation compared with a healthcare professional).

To this end, AI systems should facilitate, where appropriate, questions and actions by the user in order to provide explanation (Box 14). The AI should be able to not only indicate "how" and "why" decision are made, but also allow users to interrogate the decision, e.g., allowing users to present "counter-factual" questions (e.g., "what happens if this value is changed or removed or added?"), and to allow a dialogue between user and AI application.

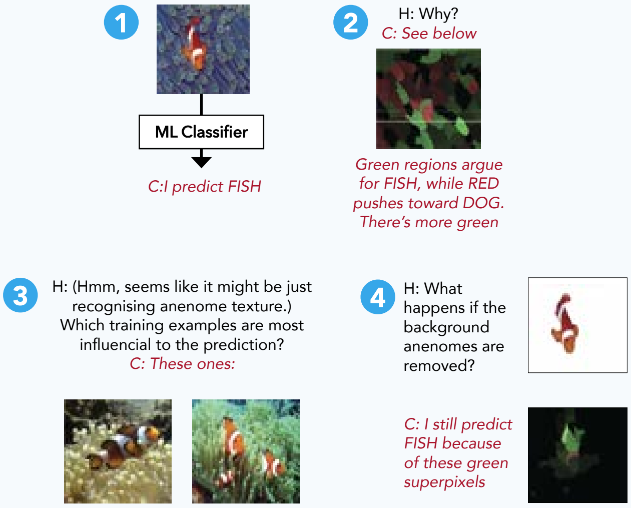
This approach of interactive, user-centred explanation does not simply provide explanation once a decision is made, but includes people as active participants in the decision-making process. This can support situation awareness, potentially reduce unwarranted overreliance on AI, and enhance human-AI teaming.

BOX 14: Interactive explanation of AI decision-making (adapted from Weld & Bansal, 2019)[63]

Explanation can be thought of as a dialogue. An AI system that explains its decisions in an interactive way should enable the user to ask questions and perform exploratory actions, such as:

- Changing the comparator (or "foil"), i.e., asking the AI why not a different class or outcome was chosen.
- Asking for more detail.
- Asking for the rationale for a decision.
- Query the sensitivity of the model, i.e., asking what minimal changes in the data might have led to a different decision.
- Perturbing the data to investigate the effect on the decision and on the explanation.

The example below illustrates some of these principles using a mock dialogue between a user and an AI classifier system that distinguishes between fish and dogs on images.


1
ML Classifier
C:I predict FISH
2
H: Why?
C: See below
Green regions argue for FISH, while RED pushes toward DOG. There's more green
3
H: (Hmm, seems like it might be just recognising anenome texture.) Which training examples are most influencial to the prediction?
C: These ones:
4
H: What happens if the background anenomes are removed?
C: I still predict FISH because of these green superpixels

Human – AI Teaming

The design of AI cannot be considered in isolation from the people who are going to use it or who will be affected by it. As AI applications become more advanced and powerful, we should assume that the AI will be used and will be working within teams of healthcare (and other) professionals and patients. We might not wish to think of AI as the new team member, but their computational capabilities means AI tools will have a much more active and dynamic role within teams than previous automation. Human – AI teaming is, therefore, another critical consideration when designing, implementing and using AI, but the value of teaming AI capabilities with human expertise, intuition and creativity is often not fully exploited[64], and the challenges of human – AI teaming in clinical practice can go unappreciated. Hence, it is important to consider what is known about teams and teamwork, and how this knowledge can be used to optimise the performance of human – AI teams. Indeed, in other contexts, experimental research suggests that the use of AI might adversely impact the performance of high-performing teams[65].

Team members engage in two forms of activity: teamwork and taskwork[66]. Teamwork includes work that requires individuals to interact or coordinate behaviour to achieve shared goals. Taskwork includes work where team members perform individual tasks separate from other team members. With this in mind, where people and AI coordinate their behaviour to achieve healthcare-related goals, they should be considered undertaking teamwork.

Teamwork models provide a good overview of some of the key aspects that require consideration for human – AI teaming. According to the Big Five model of teamwork[67], effective teams require five key teamwork behaviours: leadership, mutual performance monitoring, back-up behaviour, adaptability, and team orientation. In the logic of this model, teams will function effectively if they display these behaviours along with three critical supporting mechanisms: shared mental models, mutual trust, and closed-loop communication.

The Big Five behaviours are important considerations when designing AI systems. For example, mutual performance monitoring represents the ability to keep track of fellow team members' work, while carrying out their own work. When designing AI and the procedures and protocols that should be followed, it is important to ensure that human team members are able to understand the AI system's roles and responsibilities within the team, what the AI is doing and why, and whether the AI is performing effectively. Perhaps more challenging, with advanced AI systems it is important that the AI will be able to understand human team members' roles and responsibilities, and what they are doing and how they are performing. This is particularly the case for dynamic AI applications, which will be able to take over human tasks when people are at risk of being overloaded.

The Big Five supporting mechanisms are also important to consider. Shared mental models are achieved when team members have a shared understanding of the team's goals and tasks, as well as an understanding of individual team member's tasks and how the contribute to the team's goals. Research suggests that teams coordinate their work more effectively and achieve better overall performance when they use shared mental models[68]. As AI becomes more advanced, it will be important to ensure that an appropriate mental model is shared across both human and AI team members.

Training

Expertise is built through frequent exposure and training. Training and the potential for skill degradation are particularly relevant in healthcare, where healthcare professionals take pride in their skill sets. With the introduction of AI applications and autonomous systems, many tasks currently undertaken by clinicians would routinely be taken over by the AI. For example, radiologists are trained by doing large numbers of image reviews, for many different issues, the majority of which are relatively routine - will the expertise of radiographers deteriorate when they are exposed only to certain images specifically selected by an AI application rather than the broad range of images they currently train on day by day? Will ambulance service call handlers lose the ability to recognise ineffective breathing sounds associated with cardiac arrest when the AI routinely does it for them?

In the case of the introduction of autonomous infusion pumps, we might ask whether clinicians will be able to maintain their skills of making complicated drug dose calculations for infusions. We already know that smart pumps have, in some cases, led to deterioration of such skills in some nurses[69].

Maintaining such core skills is important in order to provide healthcare workers with the confidence to override and take over from AI applications. Regular manual involvement of people might be one strategy (referred to as "human-in-the-loop", see also section on Situation Awareness), but it could be hard to implement in practice. This might have to be backed up by specifically designed training, such as simulation.

Healthcare workers will become not only users of AI applications but also co-workers and supervisors of such applications (see section Human – AI Teaming). Healthcare workers need to understand not only how to use the AI, but also what potential weaknesses are and how the safe envelope is defined, maintained or breached. They need to be able to make sense of the AI application's decisions and actions, and provide clinically-based checks where appropriate.

These aspects should be addressed in the design, e.g., through interactive interfaces that enable explanation (see section on Explanation and trust). Such design efforts can be complemented by training that provides people with a baseline understanding of how a specific AI application works so that they are able to identify limitations and problems. Training in AI might include issues such as which aspects of the system the user can configure, and how to spot potential inaccuracies or problems. Better understanding of the limitations of AI applications can also help with problems of over-reliance.

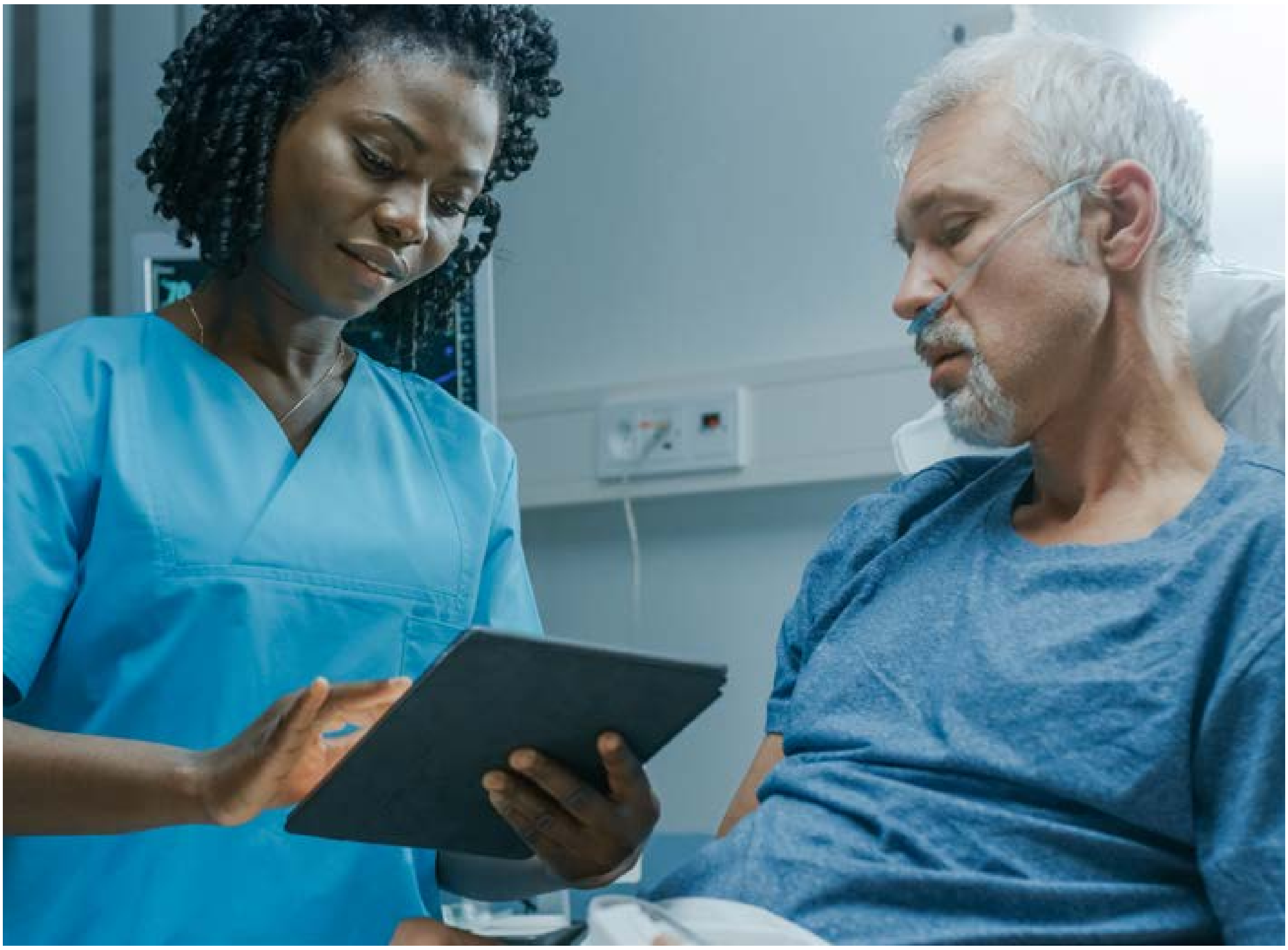
These approaches to training are intended to enable clinicians to practise their skills and to understand the behaviour and limitations of the AI they work with. However, while adequate training is an essential part of quality improvement, it should not shift away the focus from the need for good design of technology and effective systems-level interventions.

Relationships between staff and patients

Expectations at policy level are that the introduction of AI into healthcare can contribute to "giving back clinicians the gift of time"[70], i.e., that AI takes over tasks so that clinicians can focus on their patients. If done well, clinicians could spend more time interacting with patients and their relatives, for example explaining aspects of their treatment to them. This is of great importance to patients, who generally value close relationships with their healthcare professionals.

However, it is also conceivable that when AI takes over certain medical tasks, the experience of patients might deteriorate. For example, when adjusting infusions, nurses generally interact with patients, they pick up on their needs, and they can provide comfort and companionship. From a systems perspective, it is important to consider not only the medical task, which the AI application, such as an autonomous infusion pump takes over, but also the impact on relationships and the psychosocial needs of patients. There is the very real danger that rather than giving clinicians back time, they might get pulled into other activities instead. These activities could potentially be away from the patient's bed side as less manual interaction is required. In this case, the introduction of AI might not lead to more humanised care, but rather leave patients feeling more isolated.

In the UK, the guidance on "Shared decision making[71]" issued by the National Institute for Health and Care Excellence (NICE) applies also to care that is provided with the use of AI. Care (involving AI) needs to support the ongoing relationship between patients and healthcare providers, and should recognise patients as experts in their own care, who should be partners in shared decision making in relation to their care. The consideration of patient SA and how the AI supports this will be critical for shared decision making.



Ethical issues

The consideration of HF/E in the development and use of AI, and in particular for healthcare applications, must include the serious ethical challenges such systems may have for patient safety, security, privacy, and fairness and equality. These challenges may apply differently depending on the type of system, the AI algorithms and data, and the way they are used. For systems that are classed as medical devices, many of these issues may be addressed in part by regulatory processes and standards (e.g., IEC 62366-1[72]), but for AI this is a fairly new and emerging area, and it is reasonable to assert that regulation and standards are lagging behind, and there are significant challenges that manufacturers, regulators and users are facing in relation to the principles of HF/E.

Several of these questions may concern the principles of operation of the AI system, and as such how it was designed, developed and tested. There is much ongoing work to address major ethical issues, including reducing and avoiding bias in AI models, the protection of privacy, and establishing and maintaining the trust of staff, patients and the public in the use of AI. Another dimension to consider is the need for "appropriate feedback" and transparency from the AI system, aiming to help the user understand the basis for the results provided (see Explanation and trust). This, in turn, requires a certain level of technical competence for the user, so that they are capable to make this kind of assessment (see Training).

Given the potential breadth of ethical issues and the variation in complexity we could see across the range of applications, it makes sense to consider a principle-based approach that identifies some high-level areas and objectives that should be incorporated in these systems and their development and use. Table 1 (page 36) provides an example of a set of principles for AI systems[73] and their application to healthcare AI systems (see Box 15 for an example of EU guidance). Arguably, many of these ethical issues concern all types of healthcare applications, regardless of whether they employ AI or not. However, the use of AI has additional and potentially novel effects on many of these, in an increasingly complex landscape.

BOX 15: EU Ethics Guidelines for Trustworthy AI

At European level, the High-Level Expert Group on AI published "Ethics Guidelines for Trustworthy AI"[74]. These guidelines aim to steer developers and deploying organisations towards ethical and robust healthcare AI. At the outset, a Fundamental Rights Impact Assessment (FRIA) should be undertaken:

(1) Does the AI system potentially negatively discriminate against people on the basis of, for example, sex, race, colour, ethnic or social origin, genetic features, language, religious belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation?
(2) Does the AI system respect the rights of the child, for example with respect to child protection and taking the child's best interest into account?
(3) Does the AI system protect the rights to privacy, including personal data relating to individuals in line with GDPR?
(4) Does the AI system respect the freedom of expression or assembly?

The Ethics Guidelines introduce four ethical principles (respect for human autonomy, prevention of harm, fairness and explicability), and then operationalise through these through seven key requirements:

(1) Human agency and oversight,
(2) Technical robustness and safety,
(3) Privacy and data governance,
(4) Transparency,
(5) Diversity, non-discrimination and fairness,
(6) Societal and environmental well-being, and
(7) Accountability.

In order to support the translation of these requirements into practice, the guidance is accompanied by an Assessment List for Trustworthy AI (ALTAI)[75], which organisations can use as part of self-assessment. The Assessment List consists of a set of questions and prompts to consider for each of the seven key requirements.

TABLE 1: Principles for ethical AI

PRINCIPLE	DESCRIPTION
Transparency	This refers to aspects such as explainability, understandability and disclosure. For medical applications, AI transparency concerns both healthcare professionals and patients as users, in terms of explaining and justifying the output of the AI.
Justice and fairness	AI use raises issues such as fairness, consistency, inclusion and openness to challenge. For a healthcare AI application, this may for instance concern the data sets that are used by the AI and their potential to discriminate against individuals or groups.
Non-maleficence	Prevention from harm is central to any healthcare application. AI raises particular issues as potential adverse events have not yet been well identified and are complicated by changes over time in AI components.
Responsibility	Clarity is required about who is morally (and legally) responsible for AI decisions.
Privacy	Privacy and confidentiality, predominantly of patient data, are potentially significant issues given the need for large training sets for AI.
Beneficence	The developers of healthcare AI applications should be able to justify claims about the anticipated benefits from using their technology, and in particular be able to make a case for these on a risk-benefit basis.
Freedom and autonomy	Individuals who contribute data should retain some control to prevent data distribution without their consent. AI should be designed to support users in making their own decisions about healthcare.
Trust	Trust is fundamental to relationships in healthcare, and is fostered when healthcare AI can be seen to be consistently performing in a manner that meets public expectations and is also transparent.
Sustainability	The development of healthcare AI should take account of protecting the environment and also the wider commercial ecosystem in terms of innovation and competition.
Dignity	Dignity refers to how people are treated, and in particular, the extent to which their human rights are respected. The impact of AI on the delivery and experience of healthcare is relevant to people's dignity.
Solidarity	Solidarity is to do with ensuring that the AI system does not threaten social cohesion, and is responsive to the needs of vulnerable persons.

CONCLUSIONS

In this White Paper we set out the Chartered Institute of Ergonomics and Human Factors and its national and international partners' vision for Human Factors and Ergonomics in the design and use of AI applications in healthcare. There are already many examples that suggest the potential of AI to radically transform the way healthcare is accessed and delivered. However, the aspiration of using AI to improve the efficiency of health systems, and to enhance patient safety, patient experience and staff wellbeing is currently weakened by a narrow focus on technology that contrasts people and AI ("human vs. machine"), and by a limited evidence base of AI in real-world use.

It is time to move beyond this technology-centric view, and instead approach AI from a systems perspective, i.e., to consider from the outset the interaction of people with AI as part of the wider clinical and health system. The HF/E considerations discussed in this White Paper – situation awareness, workload, automation bias, explanation and trust, human-AI teaming, training, the relationships between staff and patient, and attention to ethical issues – will be crucial in delivering AI-enabled healthcare that is trustworthy and sustainable.

Designers and developers of AI should consider HF/E as a core contribution to good product design. This should be reinforced by regulatory expectations that HF/E best practices have been followed. Those procuring and using AI applications should request evidence that the AI will work with and for people within their specific settings. Funders of research and development should support this shift in focus by pushing for multi-disciplinary studies that contain HF/E as an essential discipline. In the development of reporting guidelines for AI studies there is starting to be a consensus around the need for inclusion of rigorous HF/E practices and evidence[76].

These activities need to be underpinned by education in and support with HF/E, which healthcare professionals and organisations can tap into. In the last few years, some healthcare organisations have begun to employ clinically embedded HF/E professionals to work alongside healthcare professionals every day to apply HF/E methods and principles. HF/E professionals who are embedded within an organisation have the ability to instigate and sustain systems-based interventions. The Chartered Institute of Ergonomics and Human Factors has launched in July 2021 a "Healthcare Human Factors Learning Pathway" in collaboration with Health Education England. The learning pathway is intended to support both academic and learning-at-work routes to achieve the accredited status of Technical HF/E Specialist. This professional approach can support healthcare professionals and healthcare organisations to effectively apply HF/E theories and approaches in practice – a key enabler for embedding AI successfully in health systems.

REFERENCES

1 Joshi I, Morley J. Artifiical Intelligence: How to get it right. Putting policy into practice for safe data-driven innovation in health and care. London: NHSX; 2019.

2 https://www.nhsx.nhs.uk/

3 https://www.gov.uk/government/speeches/better-tech-not-a-nice-to-have-but-vital-to-have-for-the-nhs

4 Department of Health. Making IT work: harnessing the power of health information technology to improve care in England. London2016.

5 Mollura DJ, Culp MP, Pollack E, et al. Artificial Intelligence in Low- and Middle-Income Countries: Innovating Global Health Radiology. Radiology 2020;297:513-20.

6 For a comprehensive review see: Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. Nature medicine, 25(1), pp.44-56.

7 Muehlematter, U.J., Daniore, P. and Vokinger, K.N., 2021. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. The Lancet Digital Health.

8 Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence 2021;3:199-217.

9 Meyer, A.N., Giardina, T.D., Spitzmueller, C., Shahid, U., Scott, T.M. and Singh, H., 2020. Patient Perspectives on the Usefulness of an Artificial Intelligence–Assisted Symptom Checker: Cross-Sectional Survey Study. Journal of medical Internet research, 22(1)

10 Fitzpatrick KK, Darcy A, Vierhile M. 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 2017;4:e19.

11 https://www.nhsx.nhs.uk/ai-lab/ai-lab-programmes/ai-health-and-care-award/

12 Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115.

13 Ruamviboonsuk P, Krause J, Chotcomwongse P, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. npj Digital Medicine 2019;2:25.

14 Nagendran, M., Chen, Y., Lovejoy, C.A., Gordon, A.C., Komorowski, M., Harvey, H., Topol, E.J., Ioannidis, J.P., Collins, G.S. and Maruthappu, M., 2020. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ, 368.

15 Vasey B, Ursprung S, Beddoe B, et al. Association of Clinician Diagnostic Performance With Machine Learning–Based Decision Support Systems: A Systematic Review. JAMA Network Open 2021;4:e211276-e

16 Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Internal Medicine 2021.

17 Sujan, M., Furniss, D., Grundy, K., Grundy, H., Nelson, D., Elliott, M., White, S., Habli, I. and Reynolds, N., 2019. Human factors challenges for the safe use of artificial intelligence in patient care. BMJ Health & Care Informatics, 26(1)

18 Beede E, Baylor E, Hersch F, et al., 2020. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 1–12.

19 https://iea.cc/what-is-ergonomics/

20 Holden, R.J., Carayon, P., Gurses, A.P., Hoonakker, P., Hundt, A.S., Ozok, A.A. and 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), pp.1669-1686.

21 Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly M. SEIPS 3.0: Human-centered design of the patient journey for patient safety. Applied Ergonomics 2020;84:103033.

22 Rodríguez Y, Hignett S. Integration of human factors/ergonomics in healthcare systems: A giant leap in safety as a key strategy during Covid-19. Human Factors and Ergonomics in Manufacturing & Service Industries 2021

23 Elliott, R., Camacho, E., Campbell, F., Jankovic, D., St James, M.M., Kaltenthaler, E., Wong, R., Sculpher, M. and Faria, R., 2018. Prevalence and economic burden of medication errors in the NHS in England. Rapid evidence synthesis and economic analysis of the prevalence and burden of medication error in the UK.

24 Blandford A, Dykes PC, Franklin BD, et al. Intravenous Infusion Administration: A Comparative Study of Practices and Errors Between the United States and England and Their Implications for Patient Safety. Drug Safety 2019;42:1157-65

25 Kirwan, B., & Ainsworth, L. K. (1992). A guide to task analysis: the task analysis working group: CRC press.

26 Naikar, N. (2013) Work domain analysis: Concepts, guidelines, and cases. CRC press.

27 For a more comprehensive review see Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C. and Jenkins, D.P., 2017. Human factors methods: a practical guide for engineering and design. CRC Press.

28 Hollnagel, E. (2012). FRAM, the functional resonance analysis method: modelling complex socio-technical systems: Ashgate Publishing, Ltd.

29 For a cross-industry review of how FRAM has been used, see: Patriarca, R., Di Gravio, G., Woltjer, R., Costantino, F., Praetorius, G., Ferreira, P., Hollnagel, E., 2020. Framing the FRAM: A literature review on the functional resonance analysis method. Safety Science, 129

30 Furniss, D., Nelson, D., Habli, I., White, S., Elliott, M., Reynolds, N. and Sujan, M., 2020. Using FRAM to explore sources of performance variability in intravenous infusion administration in ICU: A non-normative approach to systems contradictions. Applied Ergonomics, 86.

31 See for example ISO 9241-210:2019 Ergonomics of Human – System Interaction - Part 210: Human-centred design for interactive systems

32 Ward ME, De Brún A, Beirne D, et al. Using Co-Design to Develop a Collective Leadership Intervention for Healthcare Teams to Improve Safety Culture. International Journal of Environmental Research and Public Health 2018;15

33 https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

34 Bainbridge, L., 1983. Ironies of automation. Automatica, 19(6), pp.775-779.

35 In this case, a company employee. In commercially available vehicles with advanced driver support (e.g., Tesla's autopilot) the responsibility lies with the driver who has overall responsibility for safe use of the vehicle regardless of whether in self-driving mode or not.

36 Stanton, N. A., Salmon, P. M., Walker, G., Stanton, M. (2019). Models and Methods for Collision Analysis: A Comparison Study based on the Uber collision with a pedestrian. Safety Science. 120, 117- 128.

37 Hancock, P.A. 2019. Some pitfalls in the promises of automated and autonomous vehicles, Ergonomics: 1-31.

38 ISO/TR 9241-810:2020. Ergonomics of human – system interaction. Part 810: Robotic, intelligent and autonomous systems.

39 Parasuraman, R., Sheridan, T.B. and Wickens, C.D., 2000. A model for types and levels of human interaction with automation. IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 30(3), pp.286-297.

40 Lyell D, Coiera E, Chen J, Shah P, Magrabi F. How machine learning is embedded to support clinician decision making: an analysis of FDA-approved medical devices. BMJ Health & Care Informatics 2021;28:e100301

41 Danish researchers tested an AI tool to recognise cardiac arrest, which had higher sensitivity than call handlers, with slightly lower specificity (Blomberg SN, Folke F, Ersbøll AK, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation 2019;138:322-9). Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation, 138, 322-329). However, a subsequent randomised controlled trial found no significant improvement in call handlers' ability to recognise cardiac arrest when supported by the AI tool (Blomberg SN, Christensen HC, Lippert F, et al. Effect of Machine Learning on Dispatcher Recognition of Out-of-Hospital Cardiac Arrest During Calls to Emergency Medical Services: A Randomized Clinical Trial. JAMA Network Open 2021;4:e2032320-e).

42 Taking the thought of full autonomy further, once technology has advanced thus far, there might not be the need for ambulance service call centres anymore, as we might all be fitted with wearable and implantable sensors that continuously monitor our vital signs, predict our physiological state into the future, and pre-alert the ambulance service even before we develop a critical condition.

43 Shneiderman, B., 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human–Computer Interaction, 36(6), pp.495-504.

REFERENCES

44 Endsley, M.R., 1995. Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors, 37(1), pp.32-64.

45 Hashimoto, D.A., Witkowski, E., Gao, L., Meireles, O. and Rosman, G., 2020. Artificial intelligence in anesthesiology: current techniques, clinical applications, and limitations. Anesthesiology, 132(2), pp.379-394.

46 Wright, M.C., Taekman, J.M. and Endsley, M.R., 2004. Objective measures of situation awareness in a simulated medical environment. BMJ Quality & Safety, 13(suppl 1), pp.i65-i71.

47 Stanton, N. A., R. Stewart, D. Harris, R. J. Houghton, C. Baber, R. McMaster, P. Salmon, G. Hoyle, G. Walker, M. S. Young, M. Linsell, R. Dymott, and D. Green. 2006. Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology, Ergonomics, 49: 1288-311.

48 Onnasch, L., Wickens, C.D., Li, H. and Manzey, D., 2014. Human performance consequences of stages and levels of automation: An integrated meta-analysis. Human factors, 56(3), pp.476-488.

49 Vicente, K.J., 1999. Cognitive work analysis: Toward safe, productive, and healthy computer-based work. CRC Press.

50 Militello, L.G. and Hutton, R.J., 1998. Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. Ergonomics, 41(11), pp.1618-1641.

51 Golightly, D., 2015. Situation awareness. In Evaluation of human work. 4th edn. Taylor & Francis Group, Boca Raton, pp.549-563.

52 Hill, R. G., Sears, L. M., & Melanson, S. W. (2013). 4000 Clicks: a productivity analysis of electronic medical records in a community hospital ED. The American journal of emergency medicine, 31(11), 1591-1594.

53 Gregoriades, A., & Sutcliffe, A. (2007). Workload prediction for improved design and reliability of complex systems. Reliability Engineering & System Safety, 93:4, pp. 530-549

54 For an overview see: Chapter 8 – Mental Workload Assessment Methods, in Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C. and Jenkins, D.P., 2017. Human factors methods: a practical guide for engineering and design. CRC Press.

55 Parasuraman, R. and Riley, V., 1997. Humans and automation: Use, misuse, disuse, abuse. Human factors, 39(2), pp.230-253.

56 CIEHF. Understanding misuse of partially automated vehicles – A discussion of NTSB's findings of the 2018 Mountain View Tesla Crash. (http://bit.ly/MisusePartiallyAutomatedVehicles)

57 Burnett G, Large DR, Salanitri D. How will drivers interact with vehicles of the future? London: RAC Foundation; 2019.

58 Alberdi E, Povyakalo A, Strigini L, Ayton P. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. Academic Radiology 2004;11:909-18.

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

60 "AI in the UK: Ready, Willing and Able?," report, UK Parliament (House of Lords) Artificial Intelligence Committee, 16 April 2017; https:// publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm.

61 Miller, T. (2017) Explanation in Artificial Intelligence: insights from the Social Sciences, arXiv:1706.07269.

62 Weld DS, Bansal G. The challenge of crafting intelligible intelligence. Commun ACM 2019;62:70–9

63 Weld DS, Bansal G. The challenge of crafting intelligible intelligence. Commun ACM 2019;62:70–9

64 Saenz MJ, Revilla E, Simón C. Designing AI Systems With Human-Machine Teams. MIT Sloan Management Review 2020;61:1-5

65 Zhang G, Raina A, Cagan J, McComb C. A cautionary tale about the impact of AI on human design teams. Design Studies 2021;72:100990

66 Salas E, Cooke NJ, Rosen MA. On Teams, Teamwork, and Team Performance: Discoveries and Developments. Human Factors 2008;50:540-7

67 Salas E, Sims DE, Burke CS. Is there a "Big Five" in Teamwork? Small Group Research 2005;36:555-99

68 Scheutz M, DeLoach SA, Adams JA. A Framework for Developing and Using Shared Mental Models in Human-Agent Teams. Journal of Cognitive Engineering and Decision Making 2017;11:203-24

69 Nemeth CP, Brown J, Crandall B, Fallon C. The Mixed Blessings of Smart Infusion Devices and Health Care IT. Military Medicine 2014;179:4-10

70 https://www.gov.uk/government/speeches/better-tech-not-a-nice-to-have-but-vital-to-have-for-the-nhs

71 https://www.nice.org.uk/guidance/ng197/resources/shared-decision-making-pdf-66142087186885

72 IEC 62366-1:2015 Medical Devices: Application of usability engineering to medical devices

73 Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

74 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

75 https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

76 Vasey B, Clifton DA, Collins GS, et al. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. Nature Medicine 2021;27:186-7

ACKNOWLEDGEMENTS

Authors:

Dr Mark Sujan, Human Factors Everywhere Ltd, UK

Prof Chris Baber, University of Birmingham, UK

Prof Paul Salmon, University of the Sunshine Coast, Australia

Rachel Pool, NHS England, UK

Dr Nick Chozos, Adelard, UK

Contributors:

Dr Carlos Aceves-Gonzalez, Universidad de Guadalajara, Mexico

Prof Matthew Cooke, Capgemini, UK

Carlos Manuel Escobar-Galindo, University of Nottingham, UK

Clive Flashman, Patient Safety Learning, UK

Dr Giuseppe Frau, Deep Blue, Italy

Dr Richard Hawkins, University of York, UK

Helen Hughes, Patient Safety Learning, UK

Dr Gabriel Ibarra-Mejia, University of Texas at El Paso, US

Dr Kubra Kaya, Istanbul Medeniyet University, Turkey

Barry Kirby, K Sharp, UK

Irma Cecilia Landa-Avila, Loughborough University, UK

Dr Krzysztof Laudanski, Society for Health Care Innovation and University of Pennsylvania, US

Prof Paul Lewis, Swansea University, UK

Associate Professor Farah Magrabi, Macquarie University, Australia

Dr Riccardo Patriarca, La Sapienza University, Italy

Dr Yordán Rodríguez, Universidad de Antioquia, Colombia

Prof Wendy Rogers, Macquarie University, Australia

Tom Stocker, NHS England and NHS Improvement, UK

Rob Turpin, BSI, UK

Dr Baptiste Vasey, Oxford University, UK

Dr Chris Vincent, PDD, UK

Dr Helen Vosper, Robert Gordon University, UK

Dr Marie Ward, Trinity College Dublin / St. James's Hospital Dublin, Ireland

John Welch, University College London Hospitals NHS Foundation Trust, UK

Sean White, NHS Digital, UK

This work was supported in part by a demonstrator grant from the Assuring Autonomy International Programme (AAIP), a partnership between Lloyd's Register Foundation and the University of York.

OUR PARTNERS

We are grateful to our national and international partners who have supported and contributed to this work:

The British Standards Institution (BSI) is the UK National Standards Body, sharing knowledge, fostering innovation and instilling best practices across organisations. The strategic focus of BSI in the healthcare sector covers medical device regulation and data-driven technology initiatives.

The Society for Health Care Innovation (SHCI) was established by a diverse group of healthcare professionals caring about finding practical and context-sensitive healthcare solutions to improve lives. SHCI aims to empower individuals by supporting their natural talents and abilities to innovate solutions to problems. SHCI acts as the conduit, enabler and inspiration.

Assuring Autonomy International Programme (AAIP) is a partnership between Lloyd's Register Foundation and the University of York to assure the safety of robotics and autonomous systems worldwide.

The Irish Human Factors and Ergonomics Society (IHF/ES) is a professional society representing this discipline across Ireland. In Ireland, there is a strong interest in HF/E of health systems, but also very strong medical devices, health technology and software development industries.

The Latin American Network of Ergonomics and Human Factors in Healthcare Systems (RELAESA) was set up in November 2019 to raise awareness and increase demand and integration for Human Factors and Ergonomics (HF/E) discipline in the healthcare sector across Latin America. RELAESA aims to build a community of practice with ergonomists and healthcare practitioners from different disciplines such as pharmacists, nurses, healthcare managers, clinicians and psychologists.

Patient Safety Learning is a charity and independent voice for improving patient safety. It harnesses the knowledge, insights, enthusiasm and commitment of health and social care organisations, professionals and patients for system-wide change and the reduction of avoidable harm.

Human Factors Everywhere Ltd provides ergonomics input to applied research projects and offers consultancy and training in ergonomics across a range of safety-critical industries.

The Australian Alliance for Artificial Intelligence in Healthcare (AAAiH) brings together almost 100 national and international partners and stakeholders in academia, government, consumer, clinical, industry organisations, and peak bodies to translate frontier artificial intelligence (AI) technologies into real-world health services.

Chartered Institute
of Ergonomics
& Human Factors

www.ergonomics.org.uk
ciehf@ergonomics.org.uk

Designed for CIEHF by Connect Communications – **connectmedia.cc**