Chartered Institute of Ergonomics & Human Factors

Integrating Human-Centred AI in Clinical Practice

A guide for health and social care professionals

Contents

About this guide

Format

The guide is divided into three parts:

Part 1

uses a systems framework to describe general principles to consider for health and social care professionals (hereafter referred to as 'clinicians') when developing, deploying and using AI in clinical systems.

Part 2

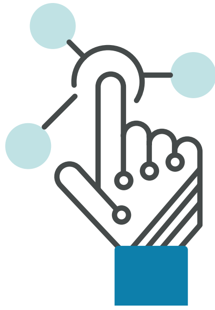
contextualises the general principles through a clinical example.

Part 3

provides links to resources for those who wish to learn more.

Who is this guide for?

This is a guide for designers, developers and users of AI in healthcare. It's written with health and social care professionals in mind – professionals who are involved in the development, deployment and use of AI in clinical practice.



Introduction

What is human-centred AI?

Human-centred AI combines a human factors and systems focus to ensure that any AI serves the needs and values of humans and ensures our wellbeing. Adopting a human factors systems approach to AI involves considering the interaction of people, tools, technology, environment and organisation with AI as part of the wider health and social care system throughout all stages of the AI development and implementation. A human factors systems approach is informed by insights from disciplines such as psychology, anatomy, physiology, engineering and design.

The guide builds on the Chartered Institute of Ergonomics & Human Factors (CIEHF) White Paper 'Human factors and Ergonomics in Healthcare AI' which sets out the need to move beyond a technology-centric view of AI [CIEHF, 2021; Sujan et al., 2022].

Artificial intelligence (AI) refers to machines (including software) that perform functions that normally require human cognition, without direct human aid [WHO, 2021].

Why is human-centred AI important?

From a human factors systems perspective, we should analyse and design for individuals, processes and environments in which the AI is to be embedded and the interactions between them [Sujan et al, 2019].

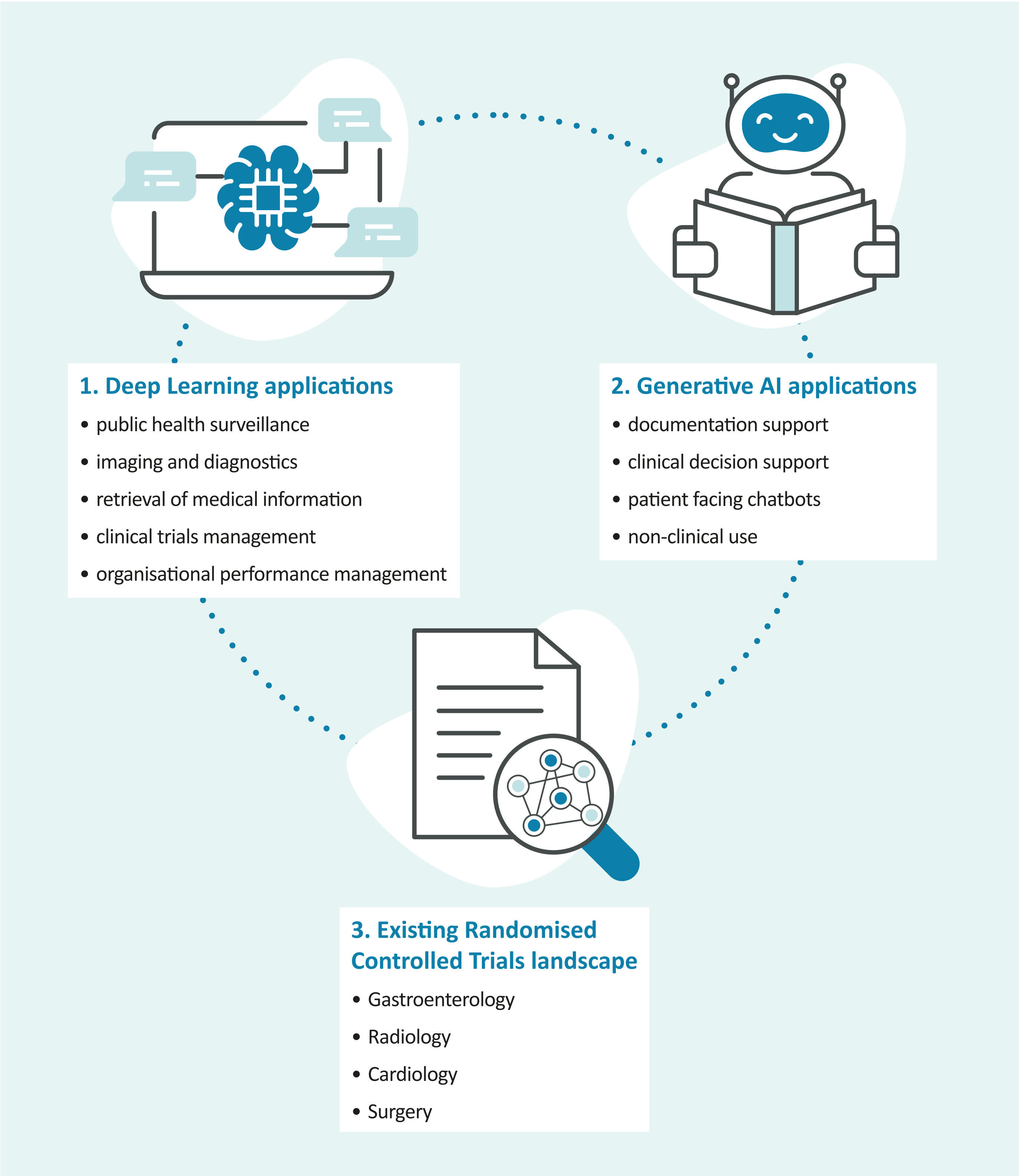
As health and social care professionals, adopting a human factors approach will allow a more inclusive and cohesive design process and an enhanced assessment and audit process, all of which will benefit service users, healthcare professionals and healthcare systems.

Human factors is defined 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 optimise human wellbeing and overall system performance' [IEA, 2000].

How is AI used in health and social care?

AI can be used in many ways in health and social care but the majority of examples to date come from healthcare. The NHS AI in Health and Care award has provided funding to many promising technologies throughout healthcare delivery: in process and triage, diagnostics, clinical decision support, chronic care management and care delivery. Examples of specific applications include the recognition of cancer, identification of undiagnosed spinal fractures, and diagnosis of rare diseases using electronic health record data. **Figure 1** provides examples of types of AI and how they might be used in healthcare.

Figure 1. Examples of Deep Learning and Generative AI application in healthcare and the emerging evidence base



Integrating Human-Centred AI in Clinical Practice: Systems considerations

Chartered Institute of Ergonomics & Human Factors

Human-centred AI is about taking a systems perspective to the design and use of AI technologies. This type of systems analysis can help clinicians and those wishing to integrate AI into their systems to look beyond just the technical characteristics of the technology. This supports better design, which will contribute to better use, better experience and acceptability, and improved patient safety. This list of prompts is not exhaustive, but it illustrates the kind of considerations that should be taken into account.

PEOPLE

Who are the people involved?

What are their needs and their specific characteristics?

What are the relationships between people (both pre- and post-AI implementation)?

ENVIRONMENTS

What do the physical spaces where work is carried out look like? How do they support or hinder the work?

What organisational processes are relevant and how do they impact on the work?

What external influences and pressures should be considered?

TASKS

What tasks are being carried out?

What characterises the tasks?

TECHNOLOGY

What tools are being used as part of the process?

What characteristics do the tools have?

KEY INTERACTIONS

Which sets of interactions among the elements of the work system are particularly critical or important?

Part 1 – General Principles

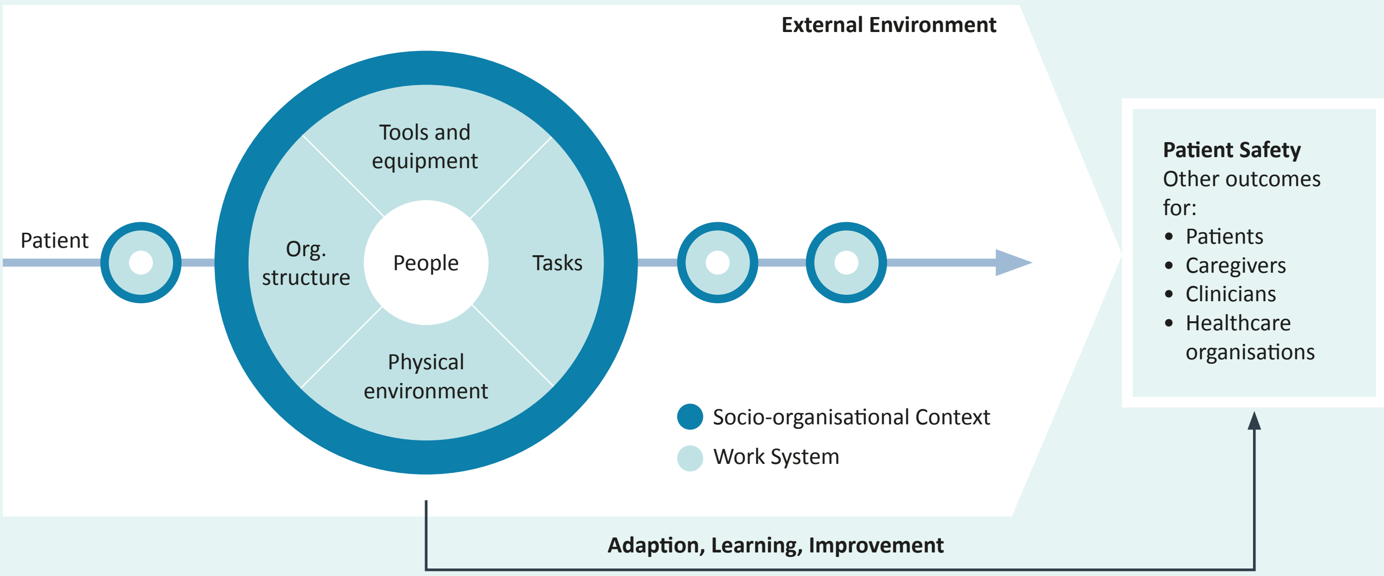
1.1 A Systems framework

We use the term 'service users' hereafter in a broad and inclusive sense, to refer to patients and clients, their families and carers or advocates or any other people receiving or supporting the implementation/realisation of health services in institutional or community settings.

The wider care system is made up of many components including people, tools and technology, tasks, workplace environments and organisational structures. AI is just one part of the wider care system. The human factors Systems Engineering Initiative for Patient Safety (SEIPS) framework [Carayon et al., 2020; NHS, 2022] provides the systems-focused framework for this guide, to enable clinicians to consider the incorporation of human factors into their AI journey. **Figure 2** outlines how SEIPS helps visualise components of the care system. The SEIPS framework can be used to identify elements of a system, describe how those elements interact and affect each other to deliver care processes, and map outcomes against the care process. Depending on the nature and variability of these interactions, different outcomes may arise. Using a systems approach enables us to understand how the use of AI might interact with other elements of the work system. This understanding can be used to design AI that can be used more safely and effectively, in order to deliver better patient, client or service user outcomes and improve healthcare professionals' wellbeing.

Example human factors considerations for each part of the system will be described here using SEIPS 3.0 [Carayon et al., 2020]. Key questions for clinicians and AI developers to address together are framed around these system elements and discussed in Part 1 and in Part 2 of this guide using a clinical example.

Figure 2. SEIPS 3.0 (taken from Carayon et al., 2020)
[www.ncbi.nlm.nih.gov/pmc/articles/PMC7152782/)](www.ncbi.nlm.nih.gov/pmc/articles/PMC7152782/)



1.2 AI use in healthcare: the socio-technical system

Most AI products will not be simply technical interventions but will be both social and technical in nature, i.e. socio-technical interventions. An example might be supporting clinicians administering medicines to acutely unwell patients in the intensive care unit (ICU). Socio-technical interventions impact people, behaviours and social relationships, including how people choose to use the tools and technology to complete their tasks and processes. Such socio-technical interventions may require changes to the surrounding socio-technical system including how people organise their work, the tools and technology they use, the workplace environment and the wider organisational context.

1.2.1 People: Service users

Integration into care pathways

AI products developed for a healthcare setting need to have demonstrable health benefits for service users (BS 30440) [BSI, 2023]. The AI needs to be integrated into the service user's care pathway and add value to that pathway, for that particular service user or future cohorts of similar service users.

Some types of AI in healthcare will be classified under Medical Devices Regulation (MDR) and the In Vitro Diagnostic Medical Devices Regulation (IVDR) AI as a Medical Device (AIaMD). It's important that service users recognise the need – and right – to have appropriate access to training and instructions for use (IFUs) for any AIaMD they are expected to use.

Care pathways are complex interventions for the mutual decision-making and organisation of care processes for a well-defined group of patients during a well-defined period [Vanhaecht, De Witte & Sermeus, 2007]. The aim of care pathways is to enhance the quality of care across the continuum by improving risk-adjusted patient outcomes, promoting patient safety, increasing patient satisfaction and optimising the use of resources.

Goal setting and development of value proposition

Service user priorities

There needs to be an understanding of the service users' priorities and less quantifiable benefits. What might the service user gain or lose from their care pathway or wider care journey through the addition of AI? For example, will interactions between service users and clinicians be increased or reduced? Documenting and analysing these interactions with the involvement of service users is important to ensure elements of the care pathway that add value to the service user are maintained.

Find out more: Mapping the clinical care pathway or the service user journey

The clinical application of AI in the care pathway needs to be amenable to iterative product improvements in response to the changing needs of service users. Ideally, service users should be involved in the co-design from the outset of AI ideation and development, and in the continuous co-production of the clinical care pathway which has the AI solution embedded within it.

Co-design

Co-design in healthcare involves the equal partnership of individuals who work within the system (healthcare professionals), individuals who have lived experience of using the system (service users, patients and their families/carers, patient advocates) and the 'designers' of the new AI tool (in this case AI developers). Co-design involves working together to design a new product, making full use of each other's knowledge, resources and contributions, to achieve better patient outcomes and/or improved system efficiency [Ward et al., 2018]. To be true co-design, service users must be involved from the very beginning and need to be supported to acquire the knowledge, skills and confidence to provide input and feedback. Co-production is the active involvement of citizens in service planning, design and delivery including the direct involvement of users in the production, at least in part, of their own services [Reape & Wallace, 2010].

Find out more: Co-design and co-production

A holistic approach to identifying and addressing service users' needs should be adopted. As part of the co-design and co-production processes, service users should be involved in designing and testing a suite of metrics that can be gathered on an ongoing basis to reflect the user experience of this AI including, for example, service user reported outcome and experience measures.

Find out more: Obtaining feedback from service users

Ethics

In line with the aim of the application of human factors, the development of any AI for use in healthcare needs to be governed by a set of ethical principles, which ensure that the wellbeing of service users is held in the highest regard [IEA, 2000]. UNESCO produced the first global standard on the ethics of AI in November 2021. This framework was adopted by all 193 Member States and stresses that the use of AI systems must respect, protect and promote human rights, freedoms and dignity and not go beyond what is necessary to achieve a legitimate aim [UNESCO, 2021]. The World Health Organization (WHO) recommended stakeholders "continuously and transparently" validate that AI is meeting these ethical principles, rights and laws [WHO, 2021]. The WHO's latest guidance goes further, recommending that ethical principles and rights must be adhered to regardless of benefits that AI can achieve and there should be post-deployment independent validation that the AI continues to meet these obligations [WHO, 2024]. The CIEHF White Paper highlights the principles for ethical AI in healthcare [CIEHF, 2021].

KEY QUESTIONS FOR CLINICIANS AND AI DEVELOPERS TO ADDRESS TOGETHER:

- Are service users and clinicians able to easily interpret the outputs of the AI?
- Does the use of AI maintain or improve the integrity of the current clinical care pathway and its benefits including those less quantifiable benefits? Is there a clearly defined need for the AI?
- Does the use of AI mitigate for any loss in these benefits?
- Does the use of AI add value to the service user's care pathway, what are the expected outcomes and how will they be measured?
- Can AI support healthcare professionals and service users to reimagine the services being provided, e.g., using AI to enable patients and service users to be active partners in their care?
- Have service users had access to appropriate training and IFU for any AIaMD they are expected to use?
- Is there a mechanism for iterative product improvements in the AI in response to the changing needs of service users built into the AI deployment?
- Are all ethical principles and rights adhered to regardless of the benefits that AI can bring?
- How will the AI empower people – both clinicians and service users – and how will they experience the difference?

1.2.2. People: Care team

Impact on current work

All clinicians who will be involved in the deployment and use of the AI need to be engaged from the beginning in mapping out how the current process works, from simple task steps to process governance, and how that process will be improved with the deployment of the AI.

- Task performance is how well the work system factors are functioning to achieve the desired task or process outcome.
- Situational awareness is an understanding of the desired outcome and how variations on the interplay of work system factors might affect the outcome of the task. Situational awareness also includes an understanding of the data sources that can inform knowledge of task performance, as well as understanding and predicting the implications of what data changes tell us about task performance and potential outcomes of the task.

Human factors tools such as Socio-Technical Systems Analysis (STSA) and Task Analysis (TA) can help with this process of understanding the current system and where AI best fits into that system.

Find out more: Socio-technical systems understanding

Impact on roles and responsibilities

Socio-technical interventions may require changes to people's roles and responsibilities. All staff, for example, need to understand and be comfortable with what tasks the AI will contribute to, the impact of any changes from the AI implementation on their role and responsibility or on their use of other information and communication technology (ICT) systems. Ideally, AI should complement the expertise and skills of healthcare staff and take the "task to the next level" [Shneiderman, 2022] rather than AI taking over, or being seen to take over, the work of humans. Schneiderman, in his Human-Centred AI approach, advocates for a 'control centre' approach. This means that those tasks that can easily be automated, e.g. inputting data, are automated. It also means that humans have more opportunity to spend time on more complex and meaningful tasks, such as providing essential oversight and ensuring situational awareness across aspects of task performance. Clinicians will need to be appropriately trained in the use and regulatory aspects of AIaMD.

AI may be deployed to facilitate a discrete task or process. Healthcare teams need to also look beyond the people impacted in that discrete task or process to people involved in other processes that intersect with the work facilitated by the AI, e.g. to include people and services who produce the inputs needed for the AI-facilitated work to be successful or to include people and services whose work will be impacted by outputs from AI-facilitated work.

Addressing needs and continuous monitoring and improvement

Developers and manufacturers should articulate clearly how and under what circumstances the proposed AI product or intervention will satisfy the identified needs and be able to adapt and improve.

- Logic models in developers and manufacturers' 'Theories of Change' can be used for clear articulation. Theories of Change identify a desired outcome or change and work backwards to identify the steps needed to achieve that outcome or change. This articulation should include consideration of the wider STS and incorporate learning and improvement about the STS from existing implementations.
- Feedback loops need to be in place to monitor circumstances where the AI is not working, new hazards or risks are introduced, or to pick up on different healthcare staff using the AI differently when integrating it into their practice. Agreements need to be in place for how all this information is communicated back to developers and how any issues are dealt with.

KEY QUESTIONS FOR CLINICIANS AND AI DEVELOPERS TO ADDRESS TOGETHER:

- Were all staff, including those who support clinicians in their daily duties, involved in the co-design of how and where the AI will be deployed in the work processes and STS?
- Does the use of AI support 'taking the task to the next level' and does it improve system or service user reported outcomes?
- Are there systems and response pathways in place with clear paths of action to produce alerts for when the AI is not working?
- Were all clinical staff appropriately trained in the regulations and requirements surrounding those AIaMD?

1.2.3 Tasks

Task analysis

A good starting point for analysing work systems is to look at the tasks that people are supposed to do (work-as-imagined) as well as what they actually do in practice (work-as-done) [Hollnagel & Woods, 1983]. A thorough understanding of the tasks can then inform us about other elements of the work system such as the people who are involved, the tools and the equipment that are going to be used, the physical spaces where the tasks are carried out, and the procedures and organisational structures that are in place. In this way, we can ensure that we adequately contextualise the potential contribution of AI technologies.

Task analysis (TA) can help to understand the task that the AI product will support and the task of successfully developing and/or deploying the AI. Human factors TA can help define and support the requirements of the AI and what it will be doing. It might be interesting to produce a job specification for what the AI product will do and how it will contribute to overall system performance.

Find out more: Task Analysis

KEY QUESTIONS FOR CLINICIANS AND AI DEVELOPERS TO ADDRESS TOGETHER:

- Do we have a detailed understanding of the tasks that the AI product will support or an 'AI job specification'?
- Is this job specification in line with existing human resource management processes, evidence-based best practice and clinical governance requirements?

1.2.4 Tools and equipment

AI 'stack'

An understanding of the existing tools and technologies used in the specific health or social care environment is important. Human factors methods can be used to help understand the interdependencies between the AI and the existing tools and technologies (referred to as 'AI stack'), as well as those tools that are the most significant in the existing work routines of the people and roles involved.

We need to understand the tools and technologies that will be used to assess how the AI works (more and less formally), how the data from the AI integrates with data from other systems, and how data on service user outcomes and service user safety can be extracted from the AI product in ways that are meaningful to clinicians.

Data quality and governance

Health and social care systems already struggle with information management systems (IMS) infrastructure, with siloed and old 'legacy systems' in places, so how will this state of play cope with the planned AI integration? AI deep machine learning tools offer amazing potential to health and social care to do large scale analysis of service user outcomes. However, data quality, integration and changes to existing infrastructure need to support these successes and strategic aspirations.

Good quality and representative data is necessary to optimise the AI product capabilities. Careful assessment is required for data collection to avoid the risk of introducing a selection bias into the AI. This can happen if the population from which training data is obtained is not sufficiently representative or contains biases within it, or algorithms are not validated in the appropriate populations [Vicente and Matute, 2023]. Sourcing and preparing good quality data for machine learning can take a long time. It is good to think of resources for this activity early in the implementation process.

It is also good to think about the end of the software development lifecycle, and put in place processes to ensure that data can be extracted in a non-proprietary format (particularly for those external facing/open access initiatives).

KEY QUESTIONS FOR CLINICIANS AND AI DEVELOPERS TO ADDRESS TOGETHER:

- Are the existing tools and technologies that support the care pathway understood?
- Have we a shared understanding (mapped out) of how the existing tools and technologies will either be replaced by or interact with the new AI product?
- Have we mapped the defined task, and the upstream and downstream inputs and impacts, of the new AI technology on the existing technology?
- Have we mapped out the flow of data and information between the existing and new tools and technologies and where the AI design/development/deployment sits in this, to ensure the longstanding aspiration of the right information flowing to the right people and systems, at the right time, for the right purposes?

1.2.5 Environment

Physical space for additional resources

In terms of the physical space, we need to consider interactions between the physical space and other needs like data security. Because health deals with 'sensitive' category data, it might be preferred to have physical space for on-site AI deployment rather than in the cloud for some AI systems.

The physical environment refers to physical layout, location and factors such as lighting, noise and temperature.

Where appropriate privacy protection and data-sharing agreements are in place, shared repositories of data in a trusted research environment that can be analysed by multiple stakeholders using different types of AI are a growing feature of healthcare data analytics. Data can be inputted in raw format but most of the time it needs to be cleaned and coded before being shared with others so that the outputs of the data analytics are trustworthy. This may require space for additional hardware equipment and extra office space for data science professionals. Also, if using the AI allows for more service users to be seen by clinicians, the existing physical environment may need to be adapted to cater for this.

External environment and pressure to adopt AI

The external environment includes influences from outside, for example, things from the regulatory, legal, economic, political, cultural or societal contexts. There is a perception in society and many nations globally that 'AI is the way to go' and that it will be used in healthcare and will be useful. We need to be aware of the influence of this societal perception on how AI is introduced into healthcare and the pressure that may be on clinicians to adopt AI uncritically into their clinical practice, before they and their host organisations are truly ready to make the most of this change. These external pressures can negatively affect valid diligence to ensure an appropriate assessment of benefits versus risks from these new technologies.

Government readiness for AI

Governments need the right tools and environment to successfully implement and manage AI in their public services, including healthcare. Successful implementation and management of AI in public services includes looking at national and international data protection issues and legislation, data quality and governance and data interoperability as well as taking account of guidance documents such as this one. The tool, Oxford Insights [2022] Government AI Readiness Index, assesses these conditions in order to help governments better prepare for the sustainable adoption of AI in their existing services.

Significant investment will need to be made at national and international levels to support AI infrastructure and to support research into AI use in healthcare, including the development of trusted research environments. The WHO app 'SARAH' is an example of such an investment in a global resource using AI-powered health information avatars that can engage users 24 hours a day in eight languages on multiple health topics, on any device [WHO, 2024b].

KEY QUESTIONS FOR CLINICIANS AND AI DEVELOPERS TO ADDRESS TOGETHER:

- How will the AI impact on the physical environment including any potential unanticipated consequences or knock-on effects on the socio-technical environment (e.g., structure, procedures, roles and responsibilities)?
- Does your organisation have the infrastructure to support data governance requirements including physical spaces and staff resources?
- Are you aware of and reflecting on any societal pressures which might influence the trust in and adoption of the AI?
- Are there government or national supports available for healthcare organisations embarking on their AI journey?

1.2.6 Socio-organisational context

The socio-organisational environment describes the characteristics of an organisational unit (e.g. a hospital, department, clinic, home or programme). Socio-organisational characteristics include aspects like structure, procedures, roles and responsibilities, relationships and organisational culture. In the deployment of an AI product, healthcare professionals and clinicians need to consider if the knock-on effects on the socio-organisational environment are understood e.g., staffing levels, clinical competency levels.

Readiness for AI

Before setting out on an AI journey, the healthcare organisation should have an awareness of its level of readiness for AI deployment, so that appropriate steps can be taken to ensure successful deployment. There are readiness assessments that have been developed for change in general [e.g. Weiner, 2009] and for change in healthcare [Holt et al., 2010]. The Healthcare Information and Management Systems Society (HIMSS) has developed maturity models for healthcare IT deployment and a self-assessment tool to help organisations measure their strengths and opportunities to advance toward a digital health and 'AI ready' ecosystem [HIMSS, 2023]. Any of these tools could be adapted for AI readiness assessments. They involve, for example, first exploring the organisation's ability to develop robust data governance processes covering the gathering, analysing, transparency and interlinking of key data sources.

Find out more: Organisational Readiness

Governance for AI in healthcare organisations

Governance structures for AI in healthcare organisations should be established to support healthcare Chief Executive Officers (CEO) or Senior Accountable Officer (SAO). Consideration should be given to where governance lies and who are the people best positioned to inform and support CEOs and SAOs on governance e.g., a transdisciplinary set of stakeholders may be needed to cover clinical activity, operations, legal and ethical implications, data governance, quality and patient safety and service user experience. Healthcare organisations supported by these committees must decide on parameters for the application of AI in healthcare, including the considerations discussed below around accountability for the clinical diagnosis supported by AI; data quality and governance going into the AI; what happens is something goes wrong with the AI, etc. Such committees should establish policies, procedures, protocols and guidelines (PPPGs) for the use of AI, environment supports and take a human factors informed approach to consider all aspects of implementing AI into an STS like healthcare.

Healthcare organisation, clinician and AI accountability

An agreement needs to be reached on accountability and where the lines can be drawn between healthcare organisation, clinician and AI accountability. Healthcare organisations have a duty to make sure staff are appropriately trained in the use of any AIaMD. Staff needs change over time, and the deployment of AI within a process or pathway needs to be responsive in design to these changes. Society also changes and the growth of technology and the use of AI is part of these ongoing wider societal changes. The next generation of clinicians may use technology in ways we can't foresee.

Healthcare organisations and developers need to identify and consider relevant professional codes of conduct for clinicians involved in the development and deployment of AI. Clinicians should also ensure that AI developers are aware of the relevant professional codes of conduct for the clinicians who will be using the AI, and ensure that developers are aware that the use of the AI does not create any tensions for clinicians in relation to their codes of professional conduct. The professional bodies may also need to change codes of conduct on the basis of the growing use of AI in healthcare.

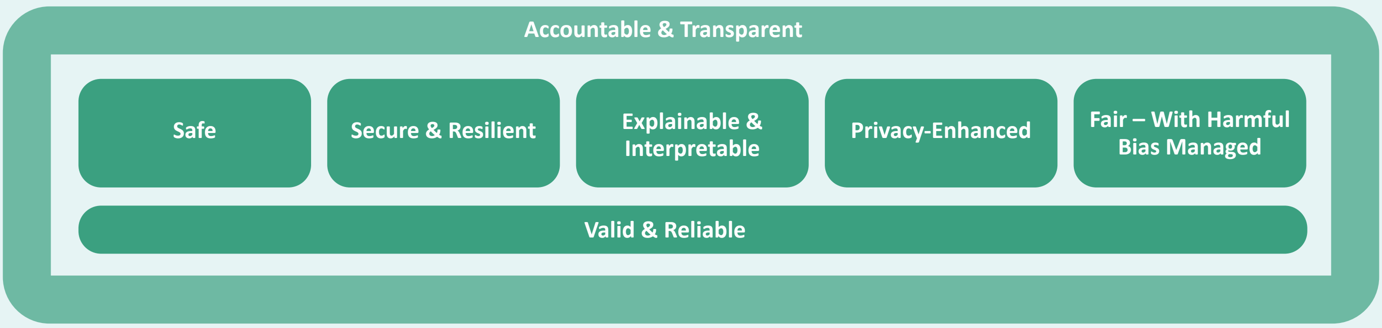
Trust in AI

Both clinicians and service users must trust in AI tools and technologies for them to be effectively deployed in STS. Building a shared understanding of all the aspects of the STS which the AI will work in and how it will work will support trust in the AI tool and technology. How the AI works and reaches decisions should be understood by, and be transparent or explainable to, both clinicians and service users. There are many frameworks for trustworthy AI, including the European Commission Ethics Guidelines for Trustworthy AI [2019] which highlight key characteristics under the categories of lawful, ethical and robust AI including, e.g. respect for human autonomy, prevention of harm, fairness, explicability, human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, environmental wellbeing and finally accountability.

The US National Institute of Standards and Technology (NIST) has developed a Trustworthy and Responsible AI Resource Center which includes guidance on how to develop a risk management framework for AI [NIST, 2023]. Its trustworthy AI framework is outlined in **Figure 3**. Its components are safe, secure, explainable, privacy-enhanced and fair. In its approach, 'Valid & Reliable' is a necessary condition of trustworthiness and is shown as the base for other trustworthiness characteristics. 'Accountable & Transparent' relates to all other characteristics.

There is an overlap between the frameworks for trustworthy AI and ethical and governance frameworks. The UNESCO [2021] and World Health Organization publications [2021, 2024] align with many of the components of trustworthy AI frameworks.

Figure 3. NIST TAI framework (2023)



Standards and regulation

Different jurisdictions have applied different models to the regulation of AI in healthcare. In 2021 the European Commission published the 'EU AI Act: First regulation on artificial intelligence', a proposed law that aims to regulate AI systems in the EU [EU, 2021] which will come into force in 2024. The Act will apply to providers of AI systems within the EU borders, users of AI systems in the EU, and certain AI systems produced externally but utilised in the EU. The Act takes a risk-based approach, categorising AI systems as high-risk, limited-risk, minimal-risk or no-risk, with requirements getting progressively stricter as the risk level rises. The Act emphasises transparency and accountability in the development and deployment of AI systems, aligning with the WHO and UNESCO published ethical and governance principles for AI.

The UK Government produced a white paper with a 'pro-innovation approach to AI regulation' [Gov. UK, 2023]. The white paper proposes principles for a proportionate risk-based approach, delivered through secondary legislation set by regulators. The UK Medicines and Healthcare products Regulatory Agency (MHRA) currently regulates software as a medical device (SaMD) under the existing general medical regulations. The MHRA considers that AIaMD is a subset of SaMD and, as such, is already covered by these regulations. Following a recent public consultation, the MHRA is now introducing a programme of new medical device regulations. This includes the MHRA's 'Software and AI as a Medical Device Change Programme' [2022] which specifically looks at specific features of AIaMD products.

The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) published their standard on Information Technology: Artificial Intelligence Management System (AIMS) in 2023 [BS ISO/IEC 42001:2023]. The standard specifies requirements for establishing, implementing, maintaining and continually improving an Artificial Intelligence Management System (AIMS) within organisations.

Due to the rapid changes that technology including AI is bringing to healthcare, a need has been identified to transition towards collaborative, transparent and inclusive consortiums for real-world data collection. One such example is the Idea, Development, Exploration, Assessment and Long-term monitoring (IDEAL) Robotics Colloquium, which proposes recommendations for evaluation during development, comparative study and clinical monitoring of surgical robots. Multiple perspectives are being considered in this colloquium, including economics, surgical training, human factors, ethics, patient perspectives and sustainability [Marcus et al., 2024].

The NHS has established an 'AI and digital regulations service' which is a cross-regulatory advisory service supporting developers and adopters of AI and digital technologies. The service provides guidance across the regulatory, evaluation and data governance pathways [NHS, 2024].

Find out more: Standards and regulations

KEY QUESTIONS FOR CLINICIANS AND AI DEVELOPERS TO ADDRESS TOGETHER:

- How will the AI impact on the socio-organisational environment including any potential unanticipated consequences or knock-on effects (e.g., relationships, organisational culture)
- Is there a readiness to adopt AI, including considerations around IT infrastructure and data availability?
- Are the appropriate governance structures in place in the healthcare organisation for the application of AI?
- How will the use of the AI products be regulated and under which authority?
- Is there clarity in relation to accountability for healthcare organisations, clinicians and the AI?
- Is there ongoing monitoring that will detect and inform when the AI is not performing to the standard the organisation hopes it will?
- Are there any issues or concerns in relation to trust in the data used for the AI, how the AI works or how the AI will be implemented to be addressed?

1.2.6 Work System and Processes

Work-as-done vs work-as-imagined

In the SEIPS model, the work system and processes are how the work is done and how it flows. Work processes are physical, cognitive, social-behavioural or a combination of these [Karsh et al., 2006]. Work processes can be performed by healthcare professionals, service users and families or collaboratively between professionals and non-professionals [Holden & Carayon 2020].

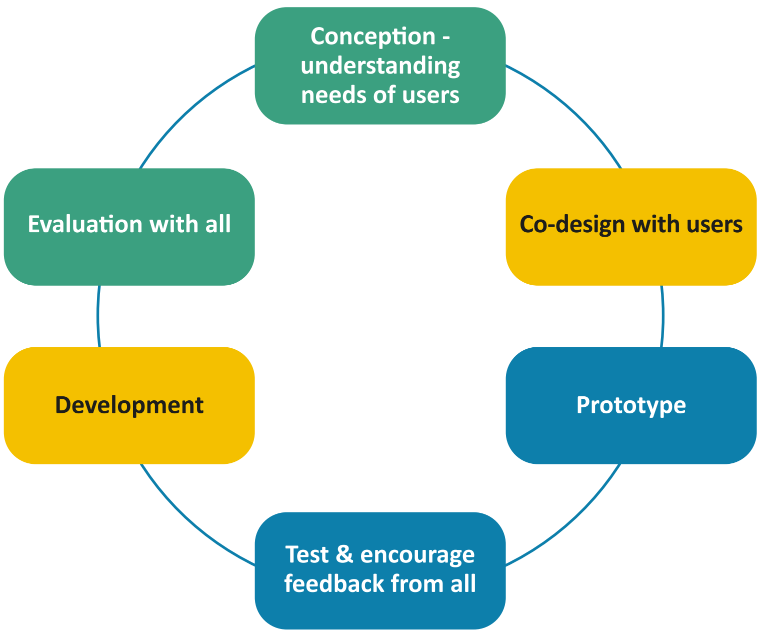
Healthcare policies, procedures, protocols and guidelines (PPPGs) can sometimes map out our abstract understanding of how work processes should happen or are 'imagined' to happen (work-as-imagined). As noted above, it is important when introducing AI to also have a grounded understanding of the reality of how work 'is done' (work-as-done) [Hollnagel & Woods, 1983]. Human factors tools can assist with gaps in understanding between work-as-done (WAD) and work-as-imagined (WAI). Before introducing AI into a care pathway, clinicians should examine their PPPGs in relation to that care pathway and test them out to see if they do reflect reality. If they do not, processes either need to be improved or the PPPGs need to be updated first to reflect the reality, before introducing the new AI product. Sometimes this can also be an iterative process of redesigning processes as you are trialling the new AI product, to ease its co-design and deployment into existing practices.

Human-centred design (HCD)

HCD principles need to be employed throughout the conception, co-design, prototyping, testing, development and evaluation phases of design for projects, technologies and their socio-technical impacts (**Figure 4**).

Human-centred design (HCD) is a well-established design approach to interactive systems development. HCD "aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors and usability knowledge and techniques. This approach enhances effectiveness and efficiency, improves human wellbeing, user satisfaction, accessibility and sustainability; and counteracts possible adverse effects of use on human health, safety and performance" [ISO 9241-210:2019(E)].

Figure 4. Simple human-centred design cycle



[*Find out more: Human-centred Design*]

Project management (PM)

A well-developed suite of procurement, project and implementation methodologies are also already available to support clinicians in deploying new technology. Clinicians can choose to develop their own technology based on in-house expertise and skills, or purchase an existing product, which may be customised.

Project management methodologies include, for example, PRINCE, Agile, Waterfall and Spiral. These methodologies vary between iterative and discrete sequential phases of planning, design, testing and implementation. The advantages and disadvantages of the methodologies centre around time, risk, complexity, flexibility and user engagement.

These factors can have an impact on project costs, user experience, the degree to which the product meets the users' needs and, ultimately, the project success. Human factors methods optimise systems for their users. There are activities across project management methodologies which can benefit from human factors methods such as TA.

Healthcare systems are both complex and dynamic in nature and combining approaches and methods that can support understanding complexity and dealing with constant change and development can support healthcare organisations and staff to not feel overwhelmed with pace of change in relation to AI.

KEY QUESTIONS FOR CLINICIANS AND AI DEVELOPERS TO ADDRESS TOGETHER:

- Has the broader understanding of work done by healthcare professionals, work done by service users and their families, and work done collaboratively by healthcare professionals and service users/families been taken into account?
- Do we have an understanding of both WAD and WAI before the AI is introduced into clinical processes?
- Do our PPPGs need to be revised to build in the new functioning of the AI into the care processes?
- Has the task of developing and implementing the AI been approached in a systematic way?
- Have HCD methods and approaches been employed throughout the lifecycle of the AI development, implementation and optimisation?

1.2.7 Outcomes

Service user experience

Patient Reported Experience Measures (PREMs) need to be developed in relation to the care pathway to ensure that the introduction of AI does not have any negative impact on service users' experience of the care they receive. PREMs need to be routinely gathered and fed back to the developer and users of AI to ensure that the use of the AI is not negatively impacting perceptions of care.

Patient experience surveys are more effective if patients feel empowered and have the knowledge, skills and confidence to provide honest feedback. This is referred to as health literacy. The WHO adopted Nutbeam's [1998] definition of health literacy as the ability of individuals to gain access to, understand and use information in ways which promote and maintain good health for themselves, their families and their communities.

AI can play a role in empowering service users, and it is important that healthcare organisations drive this forward. In turn, AI can be used effectively to gauge patient experience to complement routinely gathered PREMs.

Clinician experience

Human wellbeing, and in healthcare this means both service users and staff, is a key goal of human factors. Staff wellbeing, including how comfortable healthcare staff are with using AI or automation and its supervision in their clinical setting, how educated and confident they are in the use of AI and how they understand what the AI product can or cannot do, are all important considerations from the outset.

Safety and reliability

AI for healthcare needs to be designed and deployed with safety and reliability in mind at all times. Outcomes are much broader than just the performance of the AI, and they include safety for both service users and staff, service user and staff experience, and staff wellbeing. National and international guidelines exist for ensuring the quality and safety of service user care. Healthcare staff should carry out an assessment of the relevant standards of care in relation to the area where the AI will be deployed and develop a statement of how the new process of work with the AI will meet these standards. These can be the national standards of healthcare below which any service should not fall, e.g. the Care Quality Commission (CQC) in England or Health Information and Quality Authority (HIQA) standards in Ireland. The standards may also refer to those outlined by private accreditation bodies, e.g. Joint Commission International (JCI). There may also be international evidence-based guidelines in relation to best practice for the clinical service or standards from the various professional bodies.

The AI deployment should also consider how to meet the standards relevant to the care pathway where the AI is being deployed, e.g. the National Kidney Foundation Kidney Disease Outcomes Quality Initiative (NKF KDOQI)™ Clinical Practice Guidelines for Chronic Kidney Disease. The AI needs to be able to respond timely and effectively to current, new and emerging clinical care issues to enable safe and reliable care to be provided. Support will also need to be given to clinical service user safety managers to understand risks, incidents and change management in relation to updating existing, or introducing new, AI products.

Human factors tools can support both proactive and reactive safety processes. Proactive processes include hazard and risk analysis at different levels appropriate to the AI deployment, e.g. immediate clinician use; knock-on effects on different parts of the clinical micro, macro- and meso- level [after Bronfenbrenner's levels of a system, 1974]. Human factors tools appropriate to the different levels should be chosen, e.g.:

- Assessing clinical micro-system risk – hazard and operability study (HAZOP).
- Assessing meso-system risks – safety-critical task analysis (SCTA) [CIEHF, 2023].
- Assessing macro-system risk – developing safety case (SC) for the introduction of the AI [Sujan & Habli, 2021].

Reactive processes include safety event responses, complaints, and learning from deaths. Human factors tools can be used to understand the interplay of work system factors that resulted in the safety event and to identify safety actions to mitigate against future failings. NHS England's Patient Safety Incident Response Framework (PSIRF) [2022] incorporates human factors tools, such as those to understand everyday work risks and those to develop safety actions using the SEIPS adaptation of the Human Factors Intervention Matrix (HFIX) and the companion inequality Feasibility, Acceptability, Cost, Effectiveness, Sustainability (iFACES), a human factors approach to quantify the value of identified actions. Proactive and reactive safety processes should be in a continual learning feedback loop.

The Lucian Leap Institute (2024) has produced a report on 'Patient Safety and Artificial Intelligence: Opportunities and Challenges for Care Delivery' with some key recommendations for ensuring patient safety around the following:

- Serve and safeguard the patient.
- Learn with, engage and listen to clinicians.
- Evaluate and ensure AI efficacy and freedom from bias.
- Establish strict AI governance, oversight and guidance.
- Engage in collaborative learning across healthcare systems.

We believe this guide can help clinicians in fulfilling these recommendations.

Find out more: Safety and reliability

KEY QUESTIONS FOR CLINICIANS AND AI DEVELOPERS TO ADDRESS TOGETHER:

- Are safety and reliability taken into account when the AI is designed and deployed?
- Have risk assessments taken place relevant to the different levels of the system?
- Have outcome measures been agreed from the outset including those measures broader than the quantifiable performance of the AI, e.g., service user safety and experience and staff wellbeing and experience?
- Have standards of care and service user safety been a feature of the development and deployment of the AI product, both at the 'individual AI' and at the different system levels of micro-, meso- and macro-system?
- Has the issue of compliance with standards and flexibility to respond to new standards been addressed including compliance with medical device regulations where applicable?
- Have the clinical service safety managers had additional training or information to understand both risk and incident management in relation to the AI product?
- Are service user experience measures (PREMs) routinely collected when the AI is deployed?
- Have agreements been reached on how the AI will change in relation to feedback from service users, proactive and reactive safety processes, and safety data?

Part 2 – Putting it into Practice

2.1 Example description

There is a strong push within the NHS towards the fast and widespread adoption of AI. It is hoped that the use of highly automated or autonomous infusion medication management systems can help reduce the estimated 237 million medication errors that occur in the NHS every year.

We use the term 'patient' in this case study. Patients in intensive care units (ICU) are, by default, very ill. Patients can be on life support machines, such as ventilators, and they typically require a significant number of drugs. Some of these drugs are given intravenously via an infusion pump. The infusion pump controls the flow of the drug. A patient in ICU can receive as many as twelve drugs concurrently via infusion pumps, which are assembled in infusion pump stacks.

Figure 5 shows a simulated patient in the ICU setting. The large screen displays the charting software, which records pertinent patient data. To the right of the large screen is a stack of infusion pumps. A doctor (or clinician with prescribing privileges) prescribes a drug as part of the patient's treatment plan, and a nurse then needs to prepare the drug syringe, load the infusion pump with the drug syringe and then program the infusion pump to run at the required infusion rate for a specific duration.

Figure 5: Simulated service user in intensive care (photo credit: Nick Reynolds)

This process of IV infusion management is error prone. Common hazards include administration of the wrong drug or wrong dose, as well as administration of a drug to the wrong patient. In addition, there can be delays to the adjustment of drug dose in response to changes in the patient's condition in the current manual setup, because nurses and doctors can be busy with other activities.

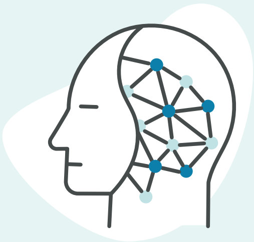
The autonomous infusion pump needs to communicate with the patient's electronic health record (EHR), it requires input from sensors providing information about the patient's vital signs and relevant physiological measurements, and it still requires nurses to load syringes into the pump. For a detailed analysis of this example please see Furniss et al. [2020].

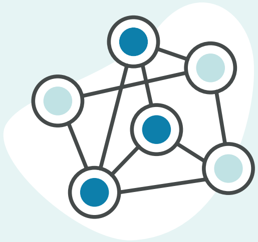
The development and introduction of an autonomous infusion pump has three key aims:

- Improve patient safety.
- Provide more personalised care (i.e. faster and more accurate response to the patient's specific physiological status).
- Reduce clinician workload.

2.2 Systems considerations

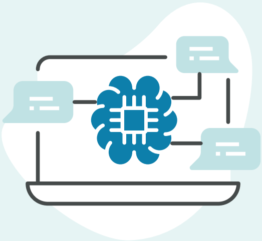
The prompts used in this example are not exhaustive, but they illustrate the kinds of systems considerations that clinicians should take into account, be it as active co-designers of the AI technology or as potential deploying organisations. It is important to emphasise that the individual elements of the work system (people, environments, tasks and technology) should not be considered in isolation, but that the main focus is on understanding key interactions between the different elements.

Work system element	Example prompts	Autonomous infusion pump
People 	Who are the people involved? What are their needs and their specific characteristics? What are relationships between people (both pre- and post-AI implementation)?	Patients can receive multiple infusions. They are often in a vulnerable state following surgery and major illness. Patients appreciate close contact with nurses. Patients do not want their close contact with nurses to change when autonomous technology is introduced. Family members: Adjusting lines and infusions often means a nurse asking family members to move away from the patient. The more often this happens, the more upsetting it can be. Autonomous, or even just remote operation of infusion pumps, can reduce these interruptions between patients and family. Nurses as primary healthcare professional contact for patients: They have close contact with their patients. Nurses need to communicate with doctors and specialists, which can take different forms, e.g. face to face, via electronic documentation, etc. Doctors, surgeons, other specialists who oversee and manage patient care: These roles are often removed from the patient. They typically discuss diagnostic and treatment plans, such as medication administration, interventions and referrals for review by other clinicians.

Work system element	Example prompts	Autonomous infusion pump
Environments 	What do the physical spaces where work is carried out look like? How do they support or hinder the work? What organisational processes are relevant and how do they impact on the work? What external influences and pressures should be considered?	If nurses have to look after several patients with the support of autonomous technology, the physical environment should enable nurses to maintain a line of sight with their patients to enable them to pick up on any changes or signs of deterioration. Additional technology might also add furniture clutter making it harder to navigate the patient's bedside. This can make it harder to access pumps and equipment to ensure they are working as intended. Sleep is very important for recovery and therefore the need for fewer manual adjustments to infusion pumps is likely to improve the environment at times such as overnight during quiet hours. The introduction of autonomous technology could be regarded by individuals in management roles as an opportunity to change the organisation of work (the activities, resources – human, financial and technical – and constraints involved in achieving an objective) with an expectation that nurses can manage more patients. This might increase nurses' cognitive load and take nurses away from the bedside, resulting in worse outcomes for staff, patients and the organisation. External pressures and lack of AI maturity of organisations can create unrealistic expectations about the extent to which autonomous technology can improve the efficiency and effectiveness of care.

Work system element	Example prompts	Autonomous infusion pump
Tasks 	What tasks are being carried out? What characterises the tasks?	Most of the tasks are related to the infusion, but providing emotional support is something that is done alongside the infusion task. How to maintain the emotional support when the infusions are done by autonomous technology requires consideration. Tasks are well documented and procedures exist. There is variability in how tasks are carried out, including the order in which they are carried out. Autonomous technology should support rather than hinder such flexibility. For example, an autonomous infusion pump might require that a prescription is recorded electronically on the system. The manual approach provides flexibility in as far as infusion can be started based on a verbal instruction, which can be entered into the system later. Preparing infusions can require complex dose calculations. This can be simplified with standardised drug sizes / strengths or pre-filled syringes. Autonomous technology can incorporate safety checks.

Work system element	Example prompts	Autonomous infusion pump
Technology 	What tools are being used as part of the process? What characteristics do these have?	The EHR can be difficult to use. As the basis for operating autonomous technology, it might enforce a particular order of steps, which might not fit actual practice in all circumstances. Computers might not always be accessible. There might be login issues. Infusion pump interfaces can be difficult to navigate and it can be hard to determine whether autonomous technology is working as expected. Medications might not come in expected and standardised doses.

Work system element	Example prompts	Autonomous infusion pump
Key interactions 	Which sets of interactions among the elements of the work system are particularly critical or important?	Patients value the close and compassionate contact with nurses **(people)**, which is often provided and maintained when nurses are monitoring or changing infusions **(tasks)**. The introduction of autonomous technology **(technology)** might lead to well-intentioned attempts to increase the number of patients nurses care for **(organisational environment)** to deal with the demand **(external influences)**. However, this might take nurses away from a patient's bedside, potentially affecting their ability to build and maintain close relationships (**people)**. There might not be sufficient space **(physical environment)** to accommodate more patients in such a way that nurses can always maintain a line of sight **(tasks)**. The monitoring and adjusting of infusions provide nurses with an opportunity to informally assess patients (e.g., the way they look, their level of alertness) to build situation awareness **(tasks)**. With the introduction of autonomous technology **(technology)**, nurses might have to rely on information and alerts provided to them via interfaces (technology), which have to be meaningful. Nurses might also require further training opportunities to better understand the potential limitations of the technology **(organisational environment)**.

2.3 Human-centred AI

Human-centred AI is about taking a systems perspective to the design and use of AI technologies. This type of systems analysis can help clinicians and those wishing to integrate AI into their systems to look beyond just the technical characteristics of the technology. This supports better design, which will contribute to better use, better experience and acceptability, and improved patient safety. Human-centred AI aims to ensure that the use of AI is aligned to human values.


L-CC
1:20
L-CC
1:21
R/06
-0-
-0-
90mm
W:4096
L:2048
zoom: 1.22
18
23.07.2020
D:100023-4225
DW-41
Group 1
90mm
W:4096
L:2048
zoom: 1.22
24
Settings
View
Open

Part 3 – Resources for further reading

This section provides a collection of useful references for further reading.

Human factors principles	Use case / Reference
A. Goal setting and development of value proposition	Department of Health & Social Care (2021) A guide to good practice for digital and data-driven health technologies, requires "having a clear value proposition", which includes understanding user needs and defining the outcome. (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#principle-1-understand-users-their-needs-and-the-context). BS 30440 requires identification of healthcare needs during the inception phase (BSI, 2023) From Deloitte, AI readiness for government – the 'Why' is important here – ambition, alignment, approach (Deloitte, 2020) Organizational readiness for artificial intelligence in health care: insights for decision-making and practice (Alami et al., 2021) STSA framework called CUBE used in an acute hospital setting which highlights importance of goals (Geary et al., 2022) Z-Inspection® is a holistic process used to evaluate the trustworthiness of AI-based technologies at different stages of the AI lifecycle. It focuses, in particular, on the identification and discussion of ethical issues and tensions through the elaboration of STS scenarios. It uses the general European Union's High-Level Expert Group's (EU HLEG) guidelines for trustworthy AI.
B. Mapping the clinical care pathway or the service user journey	Application of participatory ergonomics to the redesign of the family-centred rounds process (Xie et al., 2015) Norm-Based Approach to Incorporate Human Factors into Clinical Pathway: Reducing Human Error and Improving Patient [Service User] Safety (Tehrani et al., 2018)
C. Co-design and co-production	Using Co-Design to Develop a Collective Leadership Intervention for Healthcare Teams to Improve Safety Culture (Ward et al., 2018) Co-Design of a Trustworthy AI System in Healthcare: Deep Learning Based Skin Lesion Classifier (Zicari et al., 2021) Co-production – Effective implementation and monitoring of telehealth and telecare in Ireland: learning from international best practice. (National Disability Authority, 2018) Planning for participation (NHS England, 2015)
D. Obtaining feedback from service users	Patient [Service user] and public involvement to build trust in artificial intelligence: A framework, tools, and case studies (Banerjee et al., 2022) Public patient [service user] views of artificial intelligence in healthcare: A nominal group technique study (Musbahi et al., 2021) Patient [service user] apprehensions about the use of artificial intelligence in healthcare (Richardson et al., 2021)

Human factors principles	Use case / Reference
E. Socio-technical systems understanding	Hollnagel's Functional Resonance Analysis Method (FRAM) applied to understand transitions of care (O'Hara et al., 2020) https://www.sciencedirect.com/science/article/abs/pii/S0003687020300168 FRAM to study the response to the deteriorating service user following emergency abdominal surgery (Sujan et al., 2022). https://www.sciencedirect.com/science/article/abs/pii/S0003687021002556 The Systems Theoretic Accident Model and Processes/Systems Theoretic Process Analysis (STAMP/STPA) (Leveson, 2012) applied to care transitions (Carman et al., 2021). https://www.sciencedirect.com/science/article/abs/pii/S0003687020302878 The Systems Theoretic Accident Model and Processes/Systems Theoretic Process Analysis (STAMP/STPA) (Leveson, 2012) applied to risk in the sepsis treatment process (Kaya, 2021). https://www.sciencedirect.com/science/article/abs/pii/S0003687021000557 The Royal Academy of Engineering (RAEngineering) joined forces with the Royal College of Physicians (RCP) to advocate for a robust understanding and consideration of people, systems, design and risk. These principles were informed from work on improving the safety of service users using oral methotrexate for the treatment of rheumatoid arthritis (Ward et al., 2010). https://www.sciencedirect.com/science/article/abs/pii/S0003687009001653 The Integrated Resilience Attribute Framework was developed by Anderson et al., 2020 to guide researchers in researching resilience, especially the linkages between resilience at different scales of time and space across the whole healthcare system. https://www.sciencedirect.com/science/article/abs/pii/S0003687020300727 The STS analysis CUBE has been employed to explore the use of RFID in tracking precious specimens (Geary et al., 2021); https://www.sciencedirect.com/science/article/pii/S0003687022000825?via%3Dihub#bib3 The STS analysis CUBE has been employed to explore the use of AI for risk management in healthcare (McDonald et al., 2021) https://www.mdpi.com/1660-4601/18/23/12572 The STS analysis Cube has been used for risk management of healthcare acquired infections (Ward et al., 2024) https://www.tandfonline.com/doi/full/10.1080/00140139.2024.2396527

Human factors principles	Use case / Reference
F. Task Analysis	AHRQ Resources on Task Analysis Applying cognitive task analysis to health services research
G. Organisational Readiness	Healthcare Information and Management Systems Society (HIMSS) Intel AI Readiness Model Deloitte AI readiness framework Governmental level: Oxford Insights' Government AI Readiness Index
H. Standards and regulations	Section 11 Gov UK guide to good practice for digital data-driven health technology. **UK** Care Quality Commission (CQC) **Ireland** Information and Quality Authority (HIQA) *National Standards for Safer Better Healthcare* (2012) *Trust and medical AI: the challenges we face and the expertise needed to overcome them* (Quinn et al., 2021) *Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review* (Hond et al., 2022) *Assessment of Adherence to Reporting Guidelines by Commonly Used Clinical Prediction Models From a Single Vendor. A Systematic Review* (Lu et al., 2022) International Code of Practice (ICoP) for Telehealth Services. A quality benchmark that will transform digital health and care. The Code incorporates ISO/TS 13131 Health Informatics – Quality Planning Guidelines for Telehealth Services (2018/19) ISO 13131:2021. Health informatics – Telehealth services – Quality planning guidelines National Institute for Health and Care Excellence (NICE) Evidence Standards Framework for Digital Health Technologies (March 2019) (Appendix 3) NHS England Clinical Risk Management (2020) DCB0160: Clinical Risk Management: its Application in the Deployment and Use of Health IT Systems - NHS Digital 0160252018impguid (5).pdf –the implementation guide for DCB0160

I. Human-Centred Design (HCD)	**Relevant Standards** International Standard 15288, 2015. Systems and software engineering – system life cycle processes International Standard 24748, 2016. Systems and software engineering – life cycle management – Part 4: systems engineering planning. Human-centred design standards • ISO 9241-210:2019 Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems • ISO 9241-220: 2019 Ergonomics of human-system interaction — Part 220: Processes for enabling, executing and assessing human-centred design within organisations • ISO 9241-171:2008 Ergonomics of human-system interaction — Part 171: Guidance on software accessibility • ISO/TS 18152:2010 Ergonomics of human-system interaction — Specification for the process assessment of human-system issues Lessons learnt from applying a human-centred design process to develop one of the largest mobile health communication programmes in the world (Chamberlain et al., 2022) Transforming healthcare with AI: The impact on the workforce and organisations (McKinsey & Company 2020) Human Centred Design: Recorded webinar Dr Shelly Jeffcott from the Scottish Ambulance Service, covering the power and significance of design in all human factors work and the principles of human-centred design, which aim to support humans in the system. Human Centred Design | Turas | Learn (nhs.scot)
I. Safety and reliability	*An improved approach for failure mode and effect analysis involving large group of experts: An application to the healthcare field* (Liu et al., 2018) *Becoming a High Reliability Organization: Operational Advice for Hospital Leaders* (Hines et al., 2008) *Evidence Brief: Implementation of High Reliability Organization Principles* (Veazie et al., 2019) *Role of Artificial Intelligence in Patient [Service User] Safety Outcomes: Systematic Literature Review* IHI: Developing process, outcome and balancing measures

References

Bronfenbrenner, U. (1974). Developmental research, public policy, and the ecology of childhood. *Child Development*, 45 (1), 1-5.

BSI (2023) BS 30440 Validation framework for the use of AI within healthcare – Specification

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

CIEHF (2021) *Human Factors and Ergonomics in Healthcare AI*. White Paper.

CIEHF (2023) *How to carry out human factors assessments of critical tasks: Guidance for COMAH establishments*.

EU Commission (2021) AI Act.

European Commission (2019) *Ethics guidelines for trustworthy AI*.

Furniss, D., Nelson, D., Habli, I., White, S., Elliott, M., Reynolds, N., & 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*, 103113.

Gov UK (2023) A pro-innovation approach to AI regulation.

HIMSS (2023) Your roadmap to digital health maturity.

Holden RJ, Carayon P. SEIPS 101 and seven simple SEIPS tools. BMJ Quality & Safety 2021;30:901-910.

Hollnagel, E., & Woods, D. D. (1983). Cognitive Systems Engineering: New wine in new bottles. *International Journal of Man-Machine Studies*, 18(6), 583–600.

Holt, D. T., Helfrich, C. D., Hall, C. G., & Weiner, B. J. (2010). Are You Ready? How Health Professionals Can Comprehensively Conceptualize Readiness for Change. *Journal of General Internal Medicine*, 25 Suppl 1(Suppl 1), 50–55.

IEA (2000) What is Ergonomics (HFE)?

ISO (2019) ISO 9241-210:2019. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems

ISO/IEC 42001 (2023). Information technology – Artificial intelligence – Management system

Khanna N. N., Maindarkar M. A., Viswanathan V., Fernandes J. F. E., Paul S., Bhagawati M., Ahluwalia P., Ruzsa Z., Sharma A., Kolluri R., Singh I. M., Laird J. R., Fatemi M., Alizad A., Saba L., Agarwal V., Sharma A., Teji J. S., Al-Maini M., Rathore V., Naidu S., Liblik K, Johri A. M., Turk M., Mohanty L., Sobel D. W., Miner M., Viskovic K., Tsoulfas G., Protogerou A. D., Kitas G. D., Fouda M. M., Chaturvedi S., Kalra M. K., Suri J. S. (2022) Economics of Artificial Intelligence in Healthcare: Diagnosis vs. Treatment. *Healthcare* (Basel). Dec 9;10(12):2493. doi.org:10.3390/healthcare10122493. PMID: 36554017; PMCID: PMC9777836.

Karsh B-T, Holden RJ, Alper SJ, et al. (2006) A human factors engineering paradigm for patient [service user] safety: designing to support the performance of the healthcare professional. *BMJ Quality & Safety Health Care* 15 Suppl 1:i59-65.doi:10.1136/qshc.2005.015974

Lucian Leape Institute (2024) Patient Safety and Artificial Intelligence: *Opportunities and Challenges for Care Delivery*. Boston: Institute for Healthcare Improvement. (Available at ihi.org)

Marcus, H. J., Ramirez, P. T., Khan, D. Z. et al. (2024) The IDEAL framework for surgical robotics: development, comparative evaluation and long-term monitoring. *Nature Medicine* 30, 61–75.
NHS England (2022) Patient Safety Incident Response Framework

NHS England (2022) SEIPS quick reference guide and work system explorer

NHS (2023)

NHS (2024) Home - AI and Digital Regulations Service for health and social care

NIST (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0).

Nutbeam, D. (1998). Health Promotion Glossary. *Health Promotion International*, 13 (4): 349-364. doi: 10.1093/heapro/13.4.349

Oxford Insights (2022) Government AI Readiness Index.

Reape A. and Wallace L. M. (2010) *What is co-production?*

Shneiderman B. (2022) *Human-Centered AI*. Oxford University Press.

Sujan, M., Furniss, D., Grundy, K., Grundy, H., Nelson, D., Elliott, M., White, S., Habli, I., & Reynolds, N. (2019). Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health & Care Informatics*, 26(1), e100081.

Sujan, M., & Habli, I. (2021). Safety cases for digital health innovations: can they work?. *BMJ Quality & Safety*, 30(12), 1047–1050.

Sujan M., Pool R., Salmon P. (2022) Eight human factors and ergonomics principles for healthcare artificial intelligence *BMJ Health & Care Informatics* 2022;29:e100516.

UNESC) (2021) Recommendation on the Ethics of Artificial Intelligence

Vanhaecht K, De Witte K, Sermeus W. (2007) *The impact of clinical pathways on the organisation of care processes*. PhD dissertation, Belgium: KU Leuven

Vicente, L., Matute, H. (2023) Humans inherit artificial intelligence biases. Scientific Reports 13, 15737.

Ward, M. E., De Brún, A., Beirne, D., Conway, C., Cunningham, U., English, A., Fitzsimons, J., Furlong, E., Kane, Y., Kelly, A. et al. (2018) Using Co-Design to Develop a Collective Leadership Intervention for Healthcare Teams to Improve Safety Culture. *International Journal of Environmental Research and Public Health*, 15, 1182.

Weiner, B. J. (2009) A theory of organizational readiness for change. Implementation Science 4, 67.

World Health Organization (WHO) (2021) June 28. Ethics and governance of artificial intelligence for health: WHO Guidance

World Health Organization (WHO) (2024) Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models.

World Health Organization (WHO) (2024b) WHO unveils a digital health promoter harnessing generative AI for public health

Authors

Marie E Ward, St James's Hospital Dublin, Centre for Innovative Human Systems, Trinity College Dublin, The University of Dublin and Irish Human Factors and Ergonomics Society

Mark Sujan, University of York, Human Factors Everywhere and Health Services Safety Investigations Body

Rachel Pool, Patient Safety Implementation, NHS England

Kate Preston, University of York

Huayi Huang, Centre for Population Health Services, University of Edinburgh

Angela Carrington, Medication Safety, HSC Northern Ireland

Nick Chozos, Independent Safety and Human Factors Consultant

Contributors

Members of the CIEHF Special Interest Group on AI in Healthcare including:

Aimee Ferguson, BA(Hons), MSc, PhD. Researcher at the University of Strathclyde, Glasgow

Barry Kirby, BEng(Hons), C.ErgHF, CEng, FCIEHF, MIET. Managing Director K-Sharp

Brittany Anderson-Montoya, Lead Human Factors Specialist, Teladoc Health Inc

Emma Howie, Clinical Surgery (NHS Lothian) and the Surgical Sabermetrics group (University of Edinburgh)

Giuseppe Frau, Deep Blue

Haotian Yi, Loughborough University

Matthew Woodward, The Healthcare Improvement Studies (THIS) Institute, University of Cambridge

Raquel Santos, Head of Human Factors Engineering and Deputy-director of Information Systems and Technology, Learning Health, Luz Saúde, Portugal. Universidade Católica Portuguesa, Center for Interdisciplinary Research in Health, Portugal

Thomas Jun, Professor of Sociotechnical System Design at Loughborough University

Tom Stocker, Director – Care Cascades Ltd

External contributors:

Adam Khimji, Birmingham Women's and Children's NHS Foundation Trust; Honorary Assistant Professor, University of Nottingham, School of Medicine

Bernardo Neves, Learning Health, Luz Saúde, Portugal. Internal Medicine Department, Hospital da Luz Lisboa, Portugal

Claire Donohoe, Gastrointestinal and General Surgery, St James's Hospital, Dublin, Ireland

Eilidh Gunn, University of Edinburgh

Neil Lawrence, Software and AI, Innovative Devices, Medicines and Healthcare products Regulatory Agency, UK

Paul Campbell, Software and AI, Innovative Devices, Medicines and Healthcare products Regulatory Agency, UK

Sean White, NHS ENGLAND – X26

Tiberius Pereira, Patients for Patient Safety Ireland

Una Geary, Quality and Safety Improvement Directorate, St James's Hospital Dublin, Ireland

Chartered Institute
of Ergonomics
& Human Factors