Hamburg University of Applied Sciences (HAW Hamburg)
Faculty of Life Sciences

Adoption of AI-driven Clinical Decision Support Systems: A Checklist for Healthcare Providers Based on a Narrative Review of AI Evaluation Resources and Expert Interviews

Bachelor's Thesis
in the study programme Health Sciences (B.Sc.)

submitted by
Sergej Kucenko
Matriculation number: [redacted]

Hamburg
on 26 May 2025

1st Examiner: Prof Dr York Zöllner (HAW Hamburg)
2nd Examiner: Prof Dr Walter Leal (HAW Hamburg

Abstract

Background: Artificial intelligence (AI) has the potential to support healthcare professionals and improve patients' outcomes. AI-based clinical decision support systems (CDSSs) are reported to be particularly promising. To ensure the suitability of an AI-CDSS and prevent negative impacts, healthcare providers should ask the 'right' questions before adoption. However, there is yet no evaluation tool for AI-CDSS adoption publicly available. This thesis aimed to (1) identify guidelines and evaluation tools applicable to the adoption of CDSS and AI, and (2) synthesise AI-CDSS adoption considerations in a checklist for healthcare providers.

Methods: Trustworthy AI evaluation tools were previously identified in a scoping review by the author and colleagues. Guidelines and evaluation tools for other pre-identified categories of AI-CDSS adoption considerations were searched in PubMed, Scopus, and Google. Additional data was collected through four semi-structured interviews with experts who have backgrounds in medicine, bioethics and law, and the social science of the internet. The interviews were analysed using thematic analysis, while items from each literature source were categorised to summarise and structure AI-CDSS adoption considerations.

Results: A total of 76 literature sources, published between 2011 and 2025 and originating mainly from developed countries, were included. The majority of these sources focused on trustworthy AI or AI maturity, though guidance and evaluation tools related to other adoption categories were also identified. Their items were synthesised into a list of 227 AI-CDSS adoption questions covering the following categories: (1) regulatory and legal compliance, (2) utility, (3) trustworthy AI, (4) economic aspects, (5) usability, (6) workflow integration, (7) AI maturity, and (8) vendor reliability, support, and agreements. The expert interviews verified considerations covered by the list and helped to identify the most relevant ones. They also provided guidance on the development of an AI-CDSS adoption checklist with 20 questions.

Conclusions: The checklist integrates findings from 76 literature sources and four expert interviews. It can support both the decision whether an AI-CDSS should be adopted and the deployment of a system. While feedback on the checklist has been received from three experts and incorporated, a Delphi study involving a larger number of experts from diverse disciplines would enhance its usefulness. Furthermore, the checklist's practicality needs to be tested in the real-world, and it should be updated as the use of AI in healthcare continues to evolve.

Keywords: Artificial intelligence (AI), Clinical decision support systems (CDSS), adoption, checklist, AI ethics

Preface

I have been interested in the application of emerging technologies in healthcare for over three years now because I am excited about their potential and concerned about their risks. Tools such as artificial intelligence can contribute to improved health outcomes and reduced health disparities, but they can also exacerbate health disparities and cause severe harm if their development and use are not responsible. I have been fortunate to have Dr María Villalobos-Quesada, Prof Lisa Soleymani Lehmann, and Dr Fengxiang He as my supervisors and collaborators in three ongoing literature reviews focused on AI bias bounties, trustworthy AI evaluation tools in healthcare, and checklists applicable to trustworthy AI evaluations for clinicians. The latter two have inspired the idea to extend the scope beyond trustworthy AI and develop an adoption checklist for AI-CDSS in this thesis. It would be oversimplified to say that these two reviews were the main sources of inspiration for this thesis, because I attribute the inspiration to conduct research on AI in healthcare in large part to the supervisors behind these reviews. Dr Villalobos-Quesada opened the world of trustworthy AI in healthcare to me when we began working on our scoping review, and the regular exchange with her and Lisa Lehmann has reinforced the initial interest in responsible health technology that had been sparked two years ago.

I also want to thank everyone who has contributed to my excitement about academic research during my time at the Charité, the Leiden University Medical Centre, TU Dresden, and, of course, my home university, HAW Hamburg. I cannot stress enough how fortunate I am that all my experiences in academic research settings have been positive, and so the people I crossed paths with and who showed interest in my academic journey are one of the main reasons why I was looking forward to working on this thesis and plan to continue gaining experience in research. I was also excited to ask Prof Walter Leal to be the second examiner of my thesis, who was one of the first people to contribute to my interest in academic research.

I want to thank Prof York Zöllner for sparking my interest in health economics, for his interest in my academic plans, and, of course, for supervising this thesis and for his patience with all my questions. I also want to thank Dr Jessica Morley for supporting my work on this thesis. Her taking the time to answer my questions as a researcher from outside of my university is one of those positive experiences that contribute to my interest in research and digital health.

Finally, I want to thank my friend Milos for his continuous social support and my lovely parents for everything they have done for me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI……Artificial intelligence
AI-CDSS……...Artificial intelligence-based clinical decision support system-(s)
AI HLEG……European Commission's High-Level Expert Group on AI
ALTAI……Assessment List for Trustworthy AI
CDSS……...Clinical decision support system(-s)
CE……...Conformité Européene
EU……..European Union
DL……..Deep learning
EHR……Electronic health record(-s)
GDRP……..General Data Protection Regulation
MDR……Medical Device Regulation
ML……Machine learning
NLP……..Natural language processing
IEEE……Institute of Electrical and Electronics Engineers
ISO……..International Organization for Standardization
OECD……..Organisation for Economic Co-operation and Development

1. Introduction

Artificial intelligence (AI) has the potential to transform healthcare by improving patients' outcomes, increasing efficiency, reducing clinician burnout, and accelerating research. Its applications span treatment suggestions, patient engagement and adherence, diagnostic accuracy, and administrative tasks (Davenport & Kalakota, 2019). However, premature adoption of AI systems can have negative consequences for patients and healthcare providers. For instance, an algorithm used in hospitals in the United States to identify patients who would benefit most from "high-risk care management" programmes was found to be racially biased. By predicting healthcare costs rather than chronic health conditions, the algorithm under-identified Black patients who needed additional care, as they had lower healthcare costs than White patients due to unequal access to care, despite being sicker (Obermeyer et al., 2019). Other concerns of AI use in healthcare include patient safety, interpreting algorithm output and explaining it to stakeholders (patients, clinicians, healthcare providers), privacy considerations, and accountability for mistakes (Davenport & Kalakota, 2019).

These concerns, along with the rapid advancement of AI technology, highlight the need to implement measures that ensure the trustworthiness of these systems. Efforts in this area have shifted from "what to how", with stakeholders from the academic, commercial, public, and nonprofit sectors designing practical AI ethics tools that build on previously published AI ethics principles and guidelines (Morley et al., 2020; Ayling & Chapman, 2021). Aside from considerations related to the trustworthiness of an AI system, healthcare organisations planning to adopt AI should also consider its economic value, usability, integration with their IT infrastructure and workflows, and the reliability of AI vendors.

One use case of AI with high potential for improved health outcomes is its integration into clinical decision support systems (CDSS). These tools warrant focus, as they play a crucial role in supporting healthcare professionals with clinical decisions, and the adoption of AI-driven CDSS—with its potential and challenges—is already taking place (Elhaddad & Hamam, 2024). For example, Wilson et al. (2023) found that an AI decision support tool increased the rate of palliative care consultation for hospitalised patients and reduced 60- and 90-day readmissions, with statistical significance. Furthermore, Radić et al. (2022) stated: "Among all AI applications, clinical decision support systems (CDSS) are most likely to improve patient outcomes in the next 5–10 years." Such AI tools are also expected to reduce healthcare spending. Population ageing and expensive treatment methods drive costs, and already in 2018, the potential for savings in German healthcare expenditure for that year through rule-based and AI-based

clinical decision support (CDS) was estimated at EUR 1.4 billion (Hehner et al., 2018). To ensure the suitability of an AI-driven CDSS for a specific healthcare facility and prevent unintended consequences, healthcare providers should ask the 'right' questions prior to adopting such a system. Guidelines for healthcare providers planning to acquire an AI system (Joshi & Cushnan, 2020), guidelines and evaluation tools for specific adoption considerations such as AI trustworthiness (Scott et al., 2021; Szabo et al., 2022), and recommendations from works on AI-CDSS (Elhaddad & Hamam, 2024; Higgins & Wilson, 2025) may be used to make informed adoption decisions. However, to the best of the author's knowledge, there is yet no evaluation tool to support the adoption of AI-based CDSS publicly available.

Thus, the aim of this thesis is to synthesise publicly available guidelines and tools that can support adoption decisions for AI in healthcare and use findings from their analysis and from expert interviews to provide a checklist of considerations that healthcare providers need to evaluate before adopting an AI-CDSS. This translates to the following research questions:

1) What guidelines and tools are available to healthcare providers to evaluate an AI system before adoption?
2) What considerations do healthcare providers need to evaluate before adopting an AI-based clinical decision support system?

In this work, AI adoption encompasses both the decision to procure or not procure an AI-CDSS and its implementation. The following chapter provides the necessary background for this work and presents key concepts of AI, AI-CDSS as a technology in healthcare and its enhancement through AI, and preidentified categories of AI-CDSS adoption considerations. The third chapter presents the methodology of this work. In the following fourth chapter, the results of the literature search, the expert interviews, and the final checklist of AI-CDSS adoption considerations are presented. The fifth chapter discusses the results and relates them to current research and efforts by stakeholders outside of academia, reflects on the methods in light of the results, and draws implications for adopters of AI-driven CDSS and future research. The concluding sixth chapter summarises the main results with regards to the research questions, along with key implications for adoption of AI-CDSS and research.

2. Background

This chapter provides the necessary background for this work and presents key concepts of AI, background information on CDSS with an expanded justification for why these specific tools

require further scientific contributions, and preidentified categories of AI-CDSS adoption considerations.

2.1. Key concepts of AI

AI is an area of computer science focused on developing and researching systems and machines that can mimic human intelligence and functions, including learning, problem-solving, logical thinking, and decision-making. AI-CDSS are 'weak AI' or 'narrow AI', which are systems that perform a specific task or a limited set of tasks. In contrast, 'strong AI' or 'artificial general intelligence' is able to comprehend, learn, and apply knowledge in various tasks at a human-like or higher cognitive level, although this type of AI remains theoretical at this point. A subfield of AI is machine learning (ML), which enables computers to learn from data, recognise patterns, and make decisions or predictions without explicit programming (Stryker & Kavlakoglu, 2024). In ML, algorithms are procedures that process data, identify patterns, and create models representing the output of an algorithm executed on data (Brownlee, 2020). A commonly used class of ML algorithms are neural networks, which are based on the structure and functioning of the human brain and made up of layers of interlinked artificial neurones that process and analyse large quantities of complex data to identify patterns and relationships. Deep learning (DL), a subfield of ML, uses neural networks with many layers (known as 'deep neural networks') for a closer simulation of the human brain's intricate decision-making ability. Having an input layer, typically hundreds of hidden layers (though at least three), and an output layer as opposed to neural networks in classic ML models with typically one or two hidden layers, deep neural networks can automatically extract features from large, unannotated, and unstructured datasets and predict or infer what the data means without human intervention. This facilitates large-scale ML, which is why DL is often used in applications such as natural language processing (NLP) or computer vision that require a fast and accurate recognition of complex patterns and connections in large datasets (Stryker & Kavlakoglu, 2024). NLP is a field of AI focused on enabling "computers and digital devices to recognize, understand and generate text and speech", with chatbots and clinical text mining being examples of NLP applications (Stryker & Holdsworth, 2024). Computer vision, another area of AI, deals with machines and systems enabled to process and understand visual information, such as images and videos (IBM, 2021). The AI fields described are visualised in Figure 1.

The concept of 'big data' is relevant to all these AI technologies. Big data is characterised by the V's, including the large *volume* of data, the high *velocity* of data generation and flow, the *variety* of formats of big data (i.e., structured, semi-structured, and unstructured data), the challenge to ensure data accuracy and reliability (i.e., *veracity*) with potential noise or errors in

large datasets, and the real-world *value* that can be gained from big data. The large quantities of high-quality data and diverse data types are needed to train and refine AI models that can leverage data efficiently to make decisions or predictions (Badman & Kosinski, 2024).

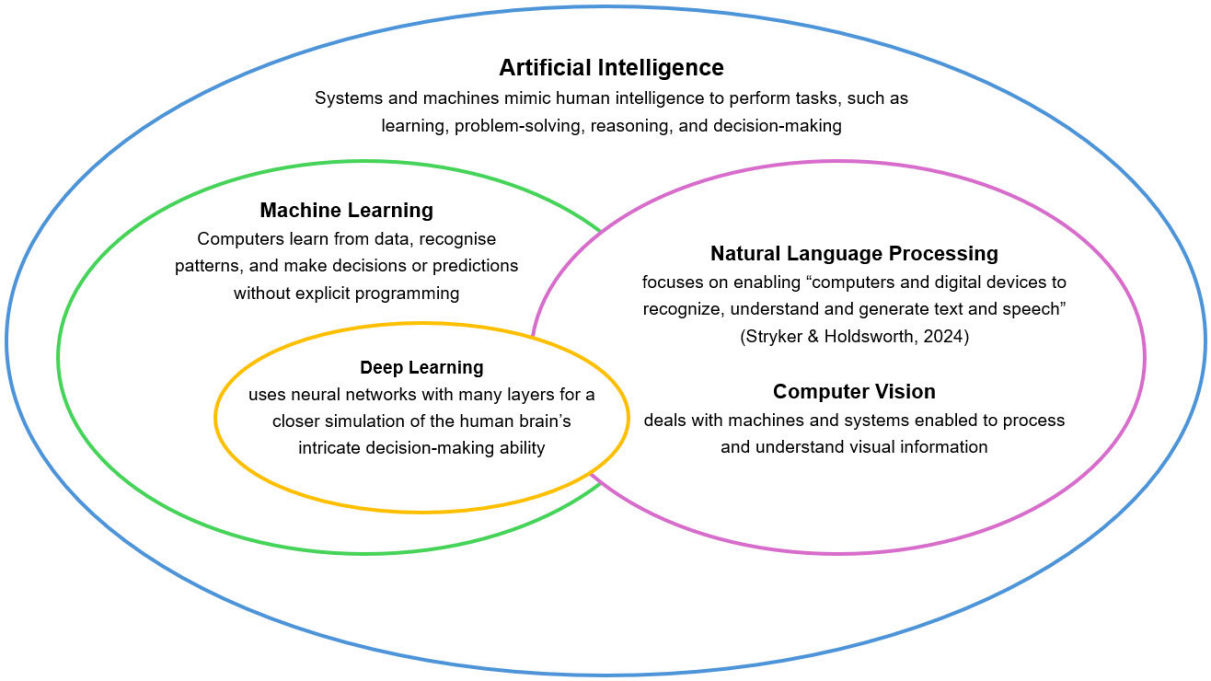

Figure 1: Key AI fields (own illustration based on information from IBM, 2021; Stryker & Kavlakoglu, 2024; Stryker & Holdsworth, 2024; and the illustration of Halejak, 2023)

To understand AI-CDSS adoption, it is essential to grasp the processes in the AI life cycle. De Silva and Alahakoon (2022) proposed an AI life cycle consisting of 19 stages across the phases of design, development, and deployment. The first stage of the life cycle focuses on the identification and formulation of the problem in a way that AI can address. This is followed by a review of the required datasets and AI ethics, as well as a review of technical literature on AI algorithms, applications in similar settings, and pre-trained models. Further stages of the design phase are data preparation (i.e., organising and structuring identified data sources while considering issues such as data ethics and security risks), data exploration (incl. comparison with benchmarks from industry and algorithmic baselines in the literature, alignment of data granularity, relationship checks for attributes and data points, outlier handling, data quality checks, and data splitting), and external data acquisition if building AI models is not feasible due to limitations in the available data. The development phase begins with data pre-processing, where it is ensured that the acquired data can be fed into the algorithm with minimal loss of information and data quality. The subsequent phases consist of building the initial AI model (i.e., choosing a suitable algorithm to develop the model and evaluating the model) and conducting data augmentation to address dataset limitations that

impact the model's output. Furthermore, the development stage involves determining a performance evaluation benchmark, building multiple AI models based on the initial model and the performance benchmark, evaluating primary metrics (e.g., accuracy, precision, F1 score) that build on the previously determined performance benchmark, and using explainable AI methods to improve the understanding of how input data leads to outputs of the AI model. The evaluation of secondary metrics is described as the first stage of the deployment phase. In this stage, the model's "intelligence" to perform the task, its computational effectiveness (e.g., computational and memory performance), factors relevant to the AI's trustworthiness (incl. privacy, robustness and cybersecurity, fairness, explainability, and interpretability), and usability should be evaluated (De Silva & Alahakoon, 2022). Prior to market placement in the European Union (EU), an AI-based CDSS would need to undergo a conformity assessment (Kleine et al., 2025), which is described in Chapter 2.2. This is followed by the deployment of the AI model for operational use and a risk assessment, covering technical risks (due to the model's integration with the technical infrastructure and workflows), AI ethics, governance, and regulatory considerations (De Silva & Alahakoon, 2022). A post-deployment technical and ethical review of the AI project, conducted by a panel of experts, steering committee, or regulatory agency, may also be required depending on its scope and the industry. The post-deployment review stage also includes post-implementation documentation and service level agreements for the use and maintenance of the AI model. The deployment phase further consists of AI service operationalisation to scale up its deployment, process and system hyperautomation where beneficial and appropriate from a regulatory and ethical perspective, and monitoring and continuous evaluation of the model's performance and outcomes (ibid.). It is important to note that this is a proposed life cycle, and the actual life cycle of AI-CDSS may differ (e.g., a less comprehensive coverage of secondary evaluation metrics if there are no regulatory requirements), highlighting the need for healthcare providers to be equipped with questions to assess the suitability and trustworthiness of an AI-based CDSS for their facility.

2.2. Fundamentals of CDSS and their enhancement through AI

CDSS are computer systems that support healthcare professionals, such as clinicians, in making clinical decisions at the point-of-care (Elhaddad & Hamam, 2024). One class of CDSS is 'knowledge-based' systems, which rely on rules derived from knowledge sources (incl. medical literature, patient-centred protocols, clinical guidelines, and expert knowledge) to generate decision support (Gholamzadeh et al., 2023). Using medical knowledge that corresponds to the state of the art and data collected from patients (e.g., medical history, test results), these systems assist healthcare professionals with diagnosis, treatment, and patient management by providing them with timely suggestions and individualised patient information

(Elhaddad & Hamam, 2024). Thus, CDSS aim to improve the quality and safety of patient care (ibid.). Specifically, they can be used to: reduce the frequency of medication errors (e.g., in drug-drug interactions) and adverse medical events (e.g., reminder system for blood glucose measurement to reduce hypoglycemia incidence), improve adherence to medical guidelines, as well as to treatment and follow-up reminders, reduce healthcare costs (e.g., minimise test duplication, suggest cheaper medication options), improve or automate administrative functions (e.g., diagnostic code selection, clinical documentation), support diagnostics (e.g., select the most appropriate test in imaging) and increase diagnostic accuracy (e.g., for peripheral neuropathy), provide direct decision support to the patient using CDS-enhanced personal health records, and enhance a workflow in an electronic health record (EHR) by improving data retrieval and presentation. CDSS frequently utilise web-applications or are integrated into EHRs and computerised provider order entry systems (i.e., healthcare provider can electronically enter and manage medical orders, such as medications and laboratory tests). Various devices, including desktops, tablets, smartphones, biometric monitors, and wearable devices can be used to deliver CDSS and display outputs (Sutton et al., 2020). These systems can be applied in a wide range of healthcare areas, such as a primary care setting and a tertiary care hospital (Chen et al., 2023).

The integration of AI in CDSS has enhanced their capabilities, facilitating the processing and interpretation of large volumes of data “with unprecedented speed and accuracy” (Elhaddad & Hamam, 2024, p. 1). The key AI technologies in CDSS are ML, DL, and NLP. CDSS that use these technologies have different use cases in healthcare. One application area is diagnostic support, which includes the use of deep learning models in medical image analysis (e.g., X-rays, histopathological images) to accurately diagnose diseases, such as skin cancer. Diagnostics can also be enhanced through NLP algorithms, which can provide deeper insights into patient conditions by extracting information from unstructured clinical notes and converting it into structured data for integration into EHRs. AI in CDSS can also analyse patient data (e.g., genetic profile, medical and treatment history) to gain a comprehensive view of patient characteristics and offer personalised treatment suggestions (Elhaddad & Hamam, 2024). For instance, a reinforcement learning agent (i.e., a type of ML where agents learn decision-making through rewards in a dynamic environment) that provides individualised treatment strategies for sepsis in intensive care in real-time, learning from patient data and treatment decisions, was found to select treatments with higher value (i.e., lower mortality) than those chosen by human clinicians (Komorowski et al., 2018). Related to personalised healthcare is the use of ML-based CDSS to identify individuals at high risk of developing a particular health condition (e.g., DL-based cardiovascular complication prediction in diabetes patients), which allows for early and individualised interventions. Moreover, AI-CDSS could assist in predicting a specific

health condition or identifying it early by detecting subtle patterns that suggest underlying pathological processes (Elhaddad & Hamam, 2024). CDSS that use AI or statistical pattern recognition to provide decision support based on patterns identified in a dataset are known as 'non-knowledge-based' systems (Gholamzadeh et al., 2023). Apart from the 'base' (i.e., either AI-determined rules based on patterns in the data or rules derived from knowledge sources) and patient data, CDSS consist of an inference engine and a communication interface (Sutton et al., 2020). The components and interactions in both knowledge-based and non-knowledge based CDSS are presented in Figure 2.

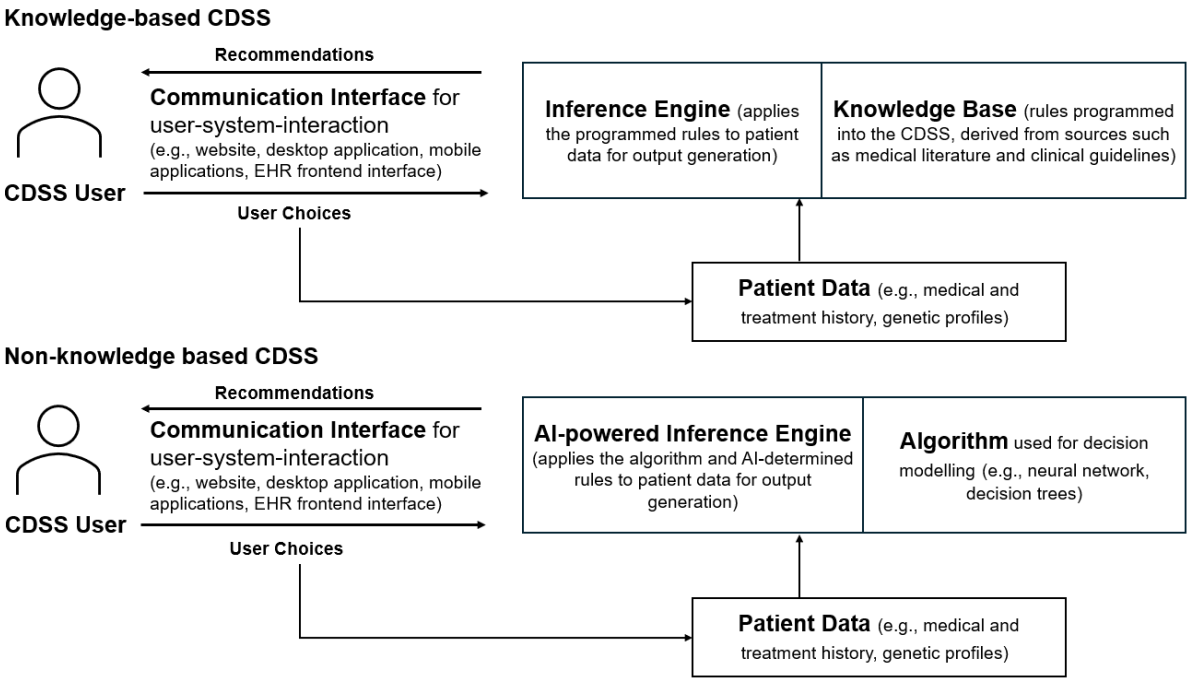

Figure 2: Components and interactions in knowledge-based and non-knowledge based CDSS (illustration based on Sutton et al., 2020 and supplemented with information from the same source)

2.3. Regulatory requirements for AI-CDSS in the European Union

As AI-CDSS can be used for the "diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease" (Art 2 Par 1, Medical Device Regulation, 2017/745), they are classified as medical devices in accordance with EU law. Compliance with the Medical Device Regulation (MDR) is required to receive the Conformité Européene (CE) mark that allows the sale of a medical device in the European Economic Area. The four classes of a device's potential risk to patients and users are Class I (low risk), Class IIa (moderate risk), Class IIb (medium risk), and Class III, which represents high risk (Kleine et al., 2025). CDSS are generally classified as Class IIa devices. If they could seriously harm an individual's health,

they are classified as Class IIb (i.e., “a serious deterioration of a person’s state of health or a surgical intervention”) or Class III (i.e., “death or an irreversible deterioration of a person’s state of health”) devices (Annex VIII, Rule 11, MDR, 2017/745). Class II medical devices must undergo a conformity assessment by a notified body (i.e., an organisation designated by an EU country to conduct conformity assessments of certain products before their market entry) to ensure their safety, efficacy and quality through different assessment activities, including clinical evaluation (Kleine et al., 2025).

For AI-CDSS used in the EU, the MDR, which does not explicitly cover AI applications, is complemented by the EU AI Act (Busch et al., 2024). This law aims to promote the development and use of trustworthy AI and safeguard the health, safety, and fundamental rights of EU citizens (Art 1 Par 1 AI Act, 2024). Thus, the regulation not only applies to manufacturers but also to deployers of AI (Art 2 Par 1 AI Act, 2024), which includes healthcare providers. The AI Act uses four risk categories: minimal risk (i.e., unregulated AI applications, such as spam filters), limited risk (i.e., AI applications that could manipulate or deceive users, who have to be informed that they interact with an AI system, such as a chatbot; Art 50 AI Act, 2024), high risk (i.e., the most regulated AI applications in the AI Act; Chapter III AI Act, 2024), and unacceptable risk (i.e., prohibited AI applications, such as social scoring systems; Chapter II AI Act, 2024) (Future of Life Institute, 2025). One condition for being classified as a high-risk AI system is the obligation to undergo a conformity assessment by a third-party (Art 6 Par 1b AI Act, 2024). Since this is required for CDSS according to the MDR, as described above, they would typically be classified as high-risk AI systems under the EU AI Act. Requirements for high-risk AI systems include: a risk management system (Art 9 AI Act, 2024), data governance (Art 10 AI Act, 2024), transparency and provision of information to deployers (Art 13 AI Act, 2024), human oversight (Art 14 AI Act, 2024), accuracy, robustness and cybersecurity (Art 15 AI Act, 2024), a quality management system (Art 17 AI Act, 2024), a fundamental rights impact assessment (Art 27 AI Act, 2024), post-market monitoring (Art 72 AI Act, 2024), and reporting of serious incidents (Art 73 AI Act, 2024). This list is not exhaustive (Busch et al., 2024).

In addition to medical device and AI-specific regulations, the use of AI-CDSS also requires compliance with data protection requirements. Data protection in the EU is regulated through the General Data Protection Regulation (GDPR). It aims to protect the individual's fundamental right to privacy and control over personal data.

2.4. Preidentified categories of AI-based CDSS adoption considerations

The following categories of aspects that need to be considered before adopting an AI-CDSS were defined based on a review article focused on the potential of AI-driven CDSS (Elhaddad & Hamam, 2024), the categories formulated in the first expert interview by a bioethicist who works as a researcher in the areas of eHealth and AI in healthcare, and knowledge of AI evaluation from the author's work on a scoping review of tools applicable for trustworthy AI evaluation in healthcare, as well as a sub-review on trustworthy AI checklists for clinicians (which is also work in progress):

- Clinical effectiveness
- AI trustworthiness
- Cost-effectiveness and other economic considerations
- Usability of the system
- Workflow integration and interoperability
- AI vendor reliability and support

'Preidentified' means that these categories were identified prior to the literature search and were subject to refinement throughout the research process based on findings from expert interviews and the review of AI evaluation resources. The final categories are presented in the Results section, specifically, in the long version of the AI-CDSS adoption considerations checklist.

A fundamental category of AI-CDSS adoption is its clinical effectiveness or utility. Healthcare providers should be confident that the system performs well in clinical practice and improves the outcome(-s) they are interested in.

Another essential consideration for healthcare providers planning to adopt an AI system is its trustworthiness. In their 'Ethics guidelines for trustworthy AI', the European Commission's High-Level Expert Group on AI (AI HLEG) outlined lawful AI, ethical AI, and robust AI as the three components of trustworthy AI (European Commission, 2019). Moreover, the guideline includes seven requirements of trustworthy AI, which were operationalised in the 'Assessment List for Trustworthy AI' (ALTAI; European Commission, 2020) and are presented in Table 1. While there is no universal definition of trustworthy AI and interpretations of its requirements vary, the understanding of AI trustworthiness in this work is based on the AI HLEG's seven requirements due to their strong overlap with existing trustworthy AI governance frameworks (McCormack & Bendechache, 2024) and their use as a lens for trustworthy AI in previous works

(Hohma & Lütge, 2023; Vetter et al., 2023; McCormack & Bendechache, 2024). Trustworthy AI should be distinguished from responsible AI, which is often used synonymously. While responsible AI focuses on the ethical aspects of AI and its auditability, as well as its accountability and liability, trustworthy AI also encompasses requirements such as robustness, which are not considered responsibility principles (unlike fairness and accountability, for example) but help build trust in a system (Díaz-Rodríguez et al., 2023). Since this thesis aims to identify the questions that need to be asked before and while adopting an AI-based CDSS and, in doing so, goes beyond responsible AI issues, the broader concept of trustworthy AI was chosen as one of the categories where considerations must be identified.

Table 1: Trustworthy AI Requirements in the ALTAI (European Commission, 2020)

Trustworthy AI Requirement	**Categories**
Human Agency and Oversight	Human Agency and Autonomy; Human Oversight
Technical Robustness and Safety	Resilience to Attack and Security; General Safety; Accuracy; Reliability, Fall-back plans and Reproducibility
Privacy and Data Governance	Privacy; Data Governance
Transparency	Traceability; Explainability; Communication
Diversity, Non-discrimination and Fairness	Avoidance of Unfair Bias; Accessibility and Universal Design; Stakeholder Participation
Environmental and Societal Well-being	Environmental Well-being; Impact on Work and Skills; Impact on Society at large or Democracy
Accountability	Auditability; Risk Management

Another category of AI-based CDSS adoption considerations is their cost-effectiveness. Even if the deciding authority conducted an economic evaluation of a system and reimbursement for its use is available, healthcare providers must understand the financial implications of adopting an AI-driven CDSS, including expenses for AI system implementation, maintenance and monitoring, as well as staff training. Differences in perspectives on cost-effectiveness justify

the need for healthcare providers to perform an economic evaluation of AI prior to its adoption. Cost-effectiveness analyses by health authorities can focus on maximising population health or social welfare (Reckers-Droog et al., 2024), while healthcare providers are primarily concerned with the specific health benefits to their patients and the financial benefits (including improved efficiency), as well as the costs for their facility. Healthcare providers may also want to conduct such analyses when there are other comparable technologies available.

A crucial adoption consideration is also the usability of an AI-driven CDSS. System design should be sensitive to the end-user and consider their work processes, cognitive load, and their preferences in interacting with an AI system (Elhaddad & Hamam, 2024). While being a key part of usability, workflow integration extends to other considerations, such as training of healthcare professionals, their autonomy, and data interoperability to ensure that an AI-CDSS brings improvement rather than disruption to existing workflows (Elhaddad & Hamam, 2024). An informed AI adoption decision should also include answered questions about AI vendor reliability and support (e.g., response and resolution times).

These categories are connected and overlap. Apart from the overlap between usability and workflow integration, the ALTAI covering ‘Impact on Work and Skills’ as a trustworthy AI aspect (European Commission, 2020), which is also an issue of workflow alignment, is another example. However, the aim is not to define mutually exclusive categories of AI-based CDSS adoption considerations, but rather to use these categories to structure the identification and synthesis of such considerations in a way that is actionable for healthcare providers planning to adopt an AI-CDSS. The following chapter describes the methodology to achieve this.

3. Methods

This chapter presents the literature search and selection of AI evaluation tools and guidelines for analysis, the primary data collection through expert interviews, and the analysis of the literature sources and primary data collected to develop a checklist for AI-CDSS adoption.

3.1. Literature search and selection

Different searches were conducted for trustworthy AI evaluation tools and resources related to the other preidentified categories of AI-CDSS adoption considerations.

3.1.1. Trustworthy AI evaluation tools

Evaluation resources in the category of trustworthy AI were identified in a scoping review on 'procedural' tools applicable to the evaluation of AI trustworthiness in healthcare. This review, which is a work in progress, has been conducted by the author of this thesis (SK), María Villalobos-Quesada (MV), and Lisa Soleymani Lehmann (LL). Conducting both a scientific and grey literature search, the authors included 95 tools in their analysis. The literature search and pre-selection of records by scanning their titles and abstracts (scientific literature) or their description and aims (grey literature) were performed by SK. Specifically, SK used keywords such as 'artificial intelligence', 'machine learning', 'evaluation framework', and 'trustworthy' to search the following scientific databases for any literature published between 2017 and November 2024: PubMed, Google Scholar, Scopus, PhilPapers, IEEE Xplore and arxiv.org. The grey literature search consisted of a search for publications from 88 organisations working in the area of trustworthy AI (the list of organisation was based on lists from Goirand et al., 2021, and AIethicist.org, as well as organisations identified during the literature search), an open google search that combined 'AI' with keywords related to trustworthiness (e.g., 'ethical AI') and evaluation instruments (e.g., 'checklist'), a search of grey literature databases (Nesta's 'AI Governance Database', OECD's 'Catalogue of Tools & Metrics for Trustworthy AI' filtered for procedural tools, and AlgorithmWatch's 'AI Ethics Guidelines Global Inventory'), and a screening of other literature source collections related to trustworthy AI (Jobin et al., 2019, 'Responsible AI Knowledge-base' via https://github.com/alexandrainst/responsible-ai, and the overviews of published AI principles, frameworks and tools, and research on fairness and explainability available at AIethicist.org). The temporal scope of this search was chosen because of the dynamic nature of technological advances in the AI field and efforts to ensure AI trustworthiness (i.e., regulations and soft law). It covers the EU Medical Device Regulation 2017/745 and the EU AI Act coming into force.

For the scoping review, two independent reviewers (SK and MV) decided on the inclusion or exclusion of evaluation tools based on the full information available. Doubt or disagreement was resolved by including a third reviewer (LL). For an evaluation tool to be included, it had to offer structured guidance that is not merely descriptive (i.e., going beyond the presentation of relevant principles) or exploratory (i.e., it helps to evaluate compliance with trustworthy AI requirements rather than exploring which problems related to AI trustworthiness could possibly occur) by including an evaluation instrument, such as a checklist, for quantitative or qualitative assessment of trustworthy AI adherence. Technical or mathematical tools, workshop material (e.g., design cards, role plays), privacy impact assessments, and evaluation tools for the managerial and organisational structures rather than for trustworthiness evaluation at the AI

system-level, were excluded. The ALTAI is an example of a procedural (i.e., non-technical) evaluation tool for trustworthy AI (European Commission, 2020). Tools were included in the scoping review if they addressed at least one of the seven trustworthy AI requirements (ibid.) with at least one item. Moreover, they had to be designed specifically for the health sector (i.e., clinical or wellbeing and lifestyle applications) or as 'non-industry-specific' tools (such as the ALTAI) with applicability to AI in the health sector. Tools were also excluded if they were not available via open access or institutional log-in, required personal information for access (e.g., name, e-mail address), or were published by authors in a personal capacity (e.g., LinkedIn posts). Only tools that were published in English were included.

Since AI trustworthiness is one of multiple AI-CDSS adoption categories and including all 95 tools identified and analysed in the scoping review was not feasible for this thesis, inclusion and exclusion criteria were defined to select a subset of these tools for the present research. These tools have different target users (i.e., policy/oversight, developers/researchers, organisations in general, management/business personnel, auditors/compliance, users as operators/service providers such as healthcare providers, non-expert end-users such as citizens, or unclear target group). Only those for which 'users as operators/service providers' and 'management/business' were identified as (one of) the tool's target users were included in this thesis. This was done to reflect that both healthcare professionals and management personnel play a role in evaluating an AI system before adoption for clinical practice. While non-industry-specific tools were included, evaluation tools designed with governmental use as the primary area of application were excluded even if they may have been applicable to healthcare. Tools being adapted versions of the ALTAI (i.e., specific to a certain country, Government of Serbia, 2023; or an abbreviated version of the ALTAI, Dignum et al., 2021) were also excluded to avoid duplication, since the trustworthy AI considerations are virtually identical.

3.1.2. Other categories of AI-CDSS adoption considerations

To identify AI guidance and evaluation resources related to the other pre-identified categories of AI-CDSS adoption considerations, searches were conducted using Google, PubMed, and Scopus. The Google search was conducted first and was deemed a suitable method for finding relevant resources, based on the experience from the scoping review, where such a search led to the inclusion of tools published by public, commercial, nonprofit, and academic sectors. Specifically, seven searches were conducted to search separately for literature on: (1) economic aspects of AI in healthcare, (2) usability of AI and CDSS, (3) usability of CDSS only (since the search string including AI and CDSS yielded only a modest number of relevant

CDSS-specific records), (4) user experience, (5) workflow integration, (6) AI readiness and maturity, and (7) AI procurement and deployment. Records found through the Google searches were screened until their relevance declined significantly. Indicators of this were a high number of duplicates of results from previous pages and an increasing number of records that could be excluded without screening their full text. A preliminary search using simple keyword combinations (e.g., CDSS usability evaluation) was conducted for each Google search to identify relevant records. These records served as quality indicators for each search string, as it was tested whether the actual search strings could retrieve them. This was done in addition to a general preliminary check of the relevance of the search results. The Google searches were performed between 5 April and 13 April 2025. The records used as references for relevance in each search, along with the exact search dates, are provided in Supplementary Material 1. No temporal filter was applied to the Google search. This is because not all Google search results display a publication date, and applying a time filter could have led to the exclusion of relevant records. For the search in PubMed, the records were retrieved on April 21, and for Scopus, on April 27, 2025. The search results in the academic databases were filtered for English and German articles published between January 2017 and April 2025. The reasoning for choosing this period is the same as that described for the scoping review. Moreover, the 'Humans' filter was applied to the PubMed search, as the search without this filter would have been both infeasible and likely inefficient, given that the 'Other Animals' filter would have retrieved a large number of articles not relevant to this thesis. Since articles that are not indexed do not have the 'Humans' label, potentially resulting in the exclusion of relevant records, a separate search was conducted for non-indexed publications. Both searches were conducted with the same search string. The records in both academic databases were screened for reference articles identified in the Google searches to verify the suitability of the search strings before proceeding with the actual search and screening. The search strings consisted of: 1) technology-related keywords, such as 'artificial intelligence' and 'clinical decision support system*', 2) keywords related to adoption considerations, such as 'usability' and 'deploy*', and 3) keywords related to the type of resource, such as 'guideline' and 'checklist'. Supplementary Material 1 includes the search strings for the databases and each Google search. All fields were searched in PubMed, whereas in Scopus, the keywords were searched only within article titles to keep the search manageable. The results retrieved from Scopus and PubMed were exported to Rayyan.AI, a tool that facilitates the screening of literature (e.g., labels to filter for keywords in titles and abstracts) and documentation of inclusion and exclusion decisions (e.g., automatic count of records with a specific exclusion label), thereby enhancing the efficiency of the screening process (Ouzzani et al., 2016).

Both Scopus and PubMed were only searched after the Google searches and an initial round of inclusion and exclusion of records had been performed. This was done due to the exploratory nature of this review and thesis. Specifically, the Google searches helped clarify the types of records that should be included and informed the identification of relevant keywords for the academic database searches to reflect aspects of AI-CDSS adoption that had not been considered in the preidentified categories of adoption considerations (e.g., AI readiness/maturity as a whole category). To mitigate the risk that this improved knowledge of available AI evaluation resources and the refinement of the search strategy might lead to inconsistencies in applying exclusion criteria, records from the Google searches that raised doubts were marked for re-review to be conducted during the screening of records from the academic databases. To be included in the review for this thesis, a record needs to: 1) provide practical procedural guidance or an evaluation tool with specific considerations for AI adopters, 2) be applicable to healthcare delivery, and 3) be specific to AI, CDSS, or AI-based CDSS. Further, the following exclusion criteria were used for all searches:

- Duplicate
- No access
- Not in English or German
- Not applicable to healthcare delivery
- Not specific to AI or CDSS
- Published in a personal capacity
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters
- Actionable procedural guidance preceding a more practical tool

As CDSS are, by definition, tools used in clinical settings, only AI adoption resources applicable to the actual delivery of healthcare to patients were included. This excludes literature on AI in public health, administrative tasks in healthcare, non-patient-facing areas within healthcare facility (e.g., Human Resources), the manufacturing of healthcare goods (e.g., pharmaceuticals), and fields such as Health Economics and Outcomes Research, where healthcare is studied but not directly delivered. ‘Applicable’ to healthcare delivery means that, as in the scoping review, evaluation tools and guidelines that are non-industry-specific were considered for inclusion. Resources intended for any other specific industry or sector (e.g., education, finance) were excluded. Moreover, only records that were specific to AI or CDSS were included. This decision was made because searching for and screening literature on digital health—which may or may not cover AI or CDSS—would have made the scope of this thesis infeasible. Records on knowledge-based CDSS were considered for inclusion. This was

because knowledge-based CDSS have been the subject of research for a considerably longer period, and some best practices or considerations (e.g., usability, workflow integration) are also relevant for AI-CDSS. Regarding the types of records included in the analysis, they had to be either a 'procedural evaluation tool' (as described for the scoping review) or 'a practical procedural guidance'. The latter was defined as any resource that is not necessarily an evaluation tool but presents AI or CDSS adoption considerations in a way that is easy to identify (e.g., by providing a table or outlining clear considerations followed by paragraphs of explanations) and therefore practical to use. For example, a scientific article in which the full text must be read to extract AI adoption issues was not considered 'practical' guidance. Moreover, to be included, any resource had to explicitly mention a group that would fall under AI or CDSS adopters (e.g., healthcare professionals, management personnel, procurement teams, deployers) as the target users of their guidance or tool.

Adopters of AI-based CDSS and healthcare providers are used interchangeably in this thesis. In the context of this work, 'healthcare providers' refers to individuals delivering healthcare, such as clinicians, or those involved in adopting an AI system at the organisational health facility level. Clinicians have been defined as professionals qualified to practice medicine in a clinical setting (Centers for Medicare & Medicaid Services, 2024), which may include physicians and physician assistants, nurse practitioners, psychiatrists, and dentists (The George Washington University, n.d.). The term 'healthcare professionals' is broader in scope and extends to anyone qualified to provide healthcare services, including, but not limited to, registered nurses, physical and occupational therapists, dietitians, paramedics, pharmacists, and medical technologists (Kavey et al., 2009).

3.2. Primary data collection

In addition to the search for literature with AI-CDSS adoption considerations, four expert interviews were conducted. Three experts were contacts of the author and the fourth interviewee was recruited through the author's supervisor. They are from the following countries: United States, England (working in the U.S.), Costa Rica (working in the Netherlands), and Germany. Their positions and academic backgrounds are presented in Table 2.

Table 2: Positions and academic backgrounds of the experts

Position	**Academic Background**
Postdoctoral Researcher working in eHealth and medical AI (Expert 1)	Bioethics and law, biotechnology, and biomedical sciences

Postdoctoral Researcher studying the use of health data, including AI and its impacts (Expert 2)	Social science of the Internet (ethical and social implications of health technology use), PhD thesis on clinical decision support software
Medical Director of Research at a health technology company, Associate Professor of Medicine, and Physician (Expert 3)	Medicine, philosophy (experience in AI ethics)
CEO and Co-Founder of a company offering telemedicine solutions, formerly a senior physician in intensive care (Expert 4)	Medicine, background in research as a Professor

A template for an interview consent form (The University of Edinburgh, 2013) was adapted (Supplementary Material 2) and sent to the experts prior to the interviews. Their transcription and analysis were conducted in accordance with this consent form. Since this research was planned, conducted, and completed within eight weeks, the interviews informed the work on this thesis in different phases. The first interview, conducted in the second week (8 April 2025), helped clarify the purpose of the planned AI-CDSS adoption checklist. Moreover, along with the second interview (15 April 2025), it informed the at that time ongoing screening of literature sources and planning of their analysis. The third (2 May 2025) and fourth interview (7 May 2025) were conducted in the second half of the research process, and validated the approach of using categories to identify recurrent AI-CDSS adoption issues from the literature.

Using insights from experts to refine the methods of this research was a secondary aim of the interviews. The main purpose was to learn how each expert categorises questions of AI-CDSS adoption, which considerations need to be addressed in each category, how AI-CDSS differ from AI in healthcare in general, and how they would approach the development and use of a checklist for AI-CDSS adoption. The interview guide is provided in Supplementary Material 3. A semi-structured approach to the interviews was chosen, as it allowed both the exploration of these pre-identified questions and response to questions and ideas raised in the interviews that had not been considered in the planning of this research. In fact, the interviews were initially meant to focus exclusively on the identification of AI-CDSS adoption issues, and not on the specifics of the checklist's development. Only after the first interview, which included a discussion about the development and use an AI-CDSS adoption checklist, a question about this, along with sub-questions, was added to the interview guide. The questions provided in the interview guide were asked to all four experts. To reduce the risk of misinterpretation in the analysis, the categories of AI-CDSS adoption considerations described by each experts were summarised in the interviews before the follow-up question was asked.

The first interview lasted one hour and 20 minutes, as the discussion about the purpose of the checklist was extensive. The other three interviews lasted between 30 to 40 minutes. Three interviews were conducted via Microsoft Teams, and the fourth was conducted via Google Meet. The final transcription of the interviews was based on an auto-generated transcription through Microsoft Teams, and the fourth was generated by the interviewee using the AI tool Gemini.

3.3. Analysis of literature sources and primary data for checklist development

To prepare the analysis, literature sources were assigned to different categories (e.g., usability, trustworthy AI, etc.). A separate document was created for each category, into which all relevant items or text passages from the sources were copied. Using the comment function in Microsoft Word, each item or passage was then assigned to a category (e.g., 'reporting of vulnerabilities, risks, and adverse impacts' as a trustworthy AI issue), in order to summarise similar considerations across sources and link related questions. These categories were developed inductively based on the items retrieved from the included records. As soon as a category was defined, the question that led to this category was quoted, or a new question was formulated by the author and used to develop a list of relevant AI-CDSS adoption considerations (herein referred to as *'the long checklist'*). Some questions in this list and therefore also the categories were refined during the analysis, for example, if another literature source included a new aspect. For example, the category 'reporting of vulnerabilities, risks, and adverse impacts' initially included biases instead of adverse impacts. However, after analysing other resources, 'adverse impacts' was adopted in the question and category to better reflect the broad range of potential negative effects associated with AI-CDSS. All resources and their assigned categories is provided as Supplementary Material 4.

The interviews were analysed using thematic analysis. This qualitative analysis approach used to identify themes or patterns in the material was deemed suitable, as the goal was to gain a deeper understanding of how experts conceptualise AI-CDSS adoption considerations and the development of a checklist. The familiarisation with the data, outlined as the first step of thematic analysis by Braun and Clarke (2006), was achieved through the check and correction of the auto-generated interview transcripts. This was followed by the generation of codes for text passages. The themes were primarily defined through deduction, as they reflect the questions from the interview guide. In the review of the initial themes, one inductively derived theme was added ('Discussing unsolved questions about AI in healthcare'), as it was not included in the interview guide but relevant across the interviews. Therefore, the thematic analysis used a hybrid approach of deduction and induction to derive insights on pre-identified

questions and allow flexibility in the identification of issues relevant to AI-CDSS adoption. The interview transcripts, the transcripts with codes, and a table with all codes linked to the themes can be found in Supplementary Material 5 – 7.

The final checklist consisting of 20 questions was developed based on the considerations identified in the literature (as presented in the long checklist) and the findings from the expert interviews. A draft of this checklist was sent to all experts for feedback five days (three business days) before the thesis submission. The feedback request included the following questions:

1. What do you think about this checklist?
2. Are there any considerations missing in the checklist? If yes, which ones?
3. Do you think any question/consideration should be removed? For instance, because you think that another consideration that is currently missing is more important.
4. Which questions would you revise? For example, because you believe additional aspects of a consideration should be covered in the question. Or you think that the phrasing of a question could be improved (to make it more accurate/practical)?
5. Do you agree with who is responsible for answering the checklist questions?

The checklist draft and feedback questions are provided in Supplementary Material 8. Feedback was received from two interviewees via E-Mail and one expert provided feedback in an additional online meeting. This led to a revision of the checklist, and the final version is presented in Chapter 4.4.

4. Results

This chapter presents the AI evaluation literature sources identified, findings from the expert interviews, the synthesis of AI-CDSS adoption considerations, and a 20-question checklist to evaluate these considerations.

4.1. Tools and guidelines for AI evaluation

Of the 95 tools identified as applicable for evaluating AI trustworthiness in healthcare in the scoping review in progress by this author and colleagues, 30 were included in this thesis. Out of the 2,092 records screened after retrieval from PubMed and Scopus, 14 were included. Furthermore, 32 records were included from the 766 search results retrieved via Google. As a result, a total of 76 literature resources with guidance or an evaluation tool were included. The most common reason for exclusion was that records were not considered to be practical

procedural guidance or an evaluation tool with specific considerations for AI adopters. The PRISMA flow diagram is presented in Figure 3. Additionally, a breakdown of the search results for each database and Google search is available in Supplementary Material 9.

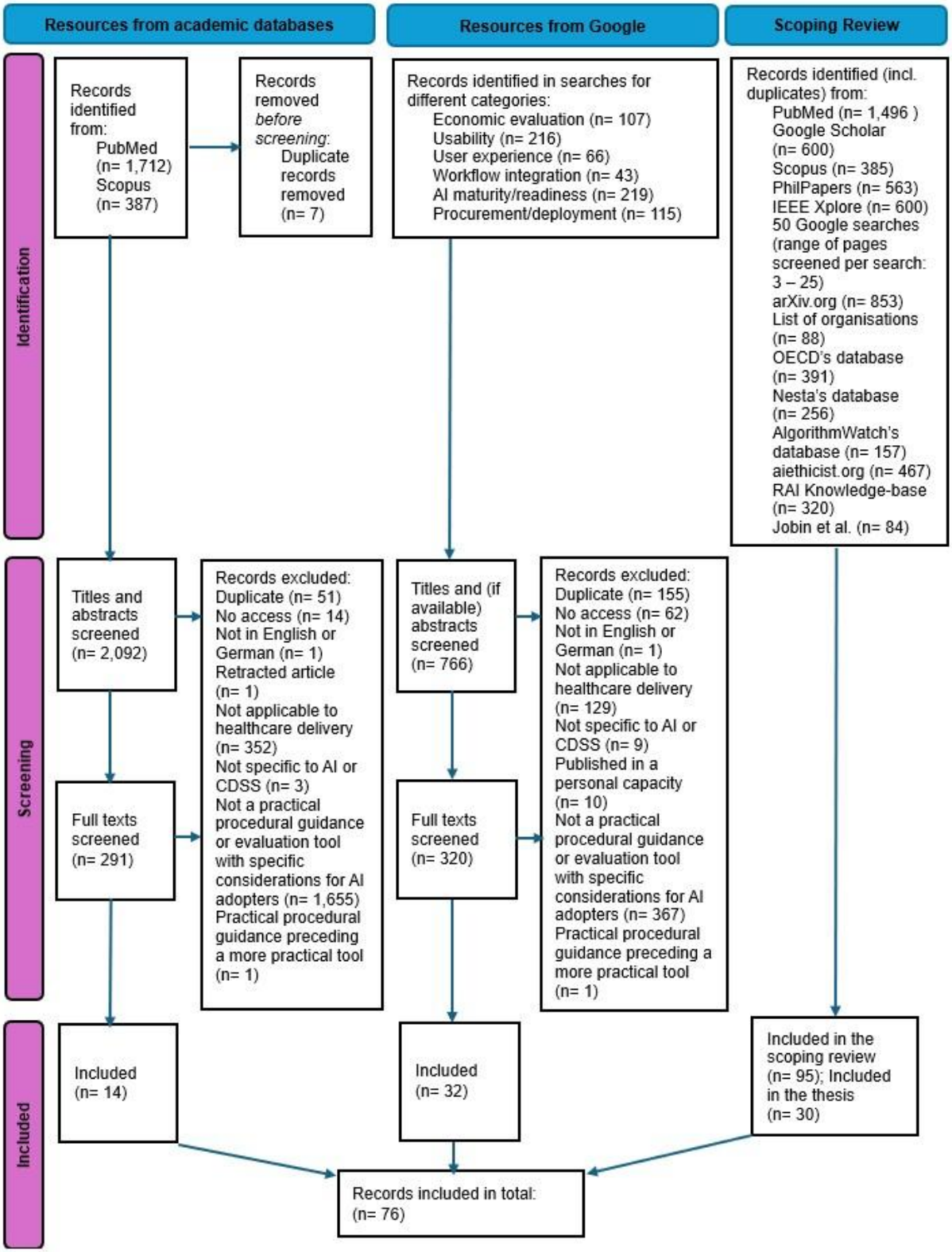

Figure 3: PRISMA flow diagram for the search results

About half of the included resources are peer-reviewed literature, while the other half comprises grey literature. While the search was open to literature published from 2017 onwards, the majority (47 out of 70) of sources with an identifiable publication year were published between 2022 and 2025 (Figure 4). Two records (Fossum et al., 2011; Horsky et al., 2012), published in 2011 and 2012 and identified via Google (where no temporal filter was applied), were included because their usability considerations for knowledge-based CDSS were deemed applicable to AI-CDSS. These records helped compensate for the relative lack of usability-related resources among the included literature. The publication year could not be identified for seven grey literature records.

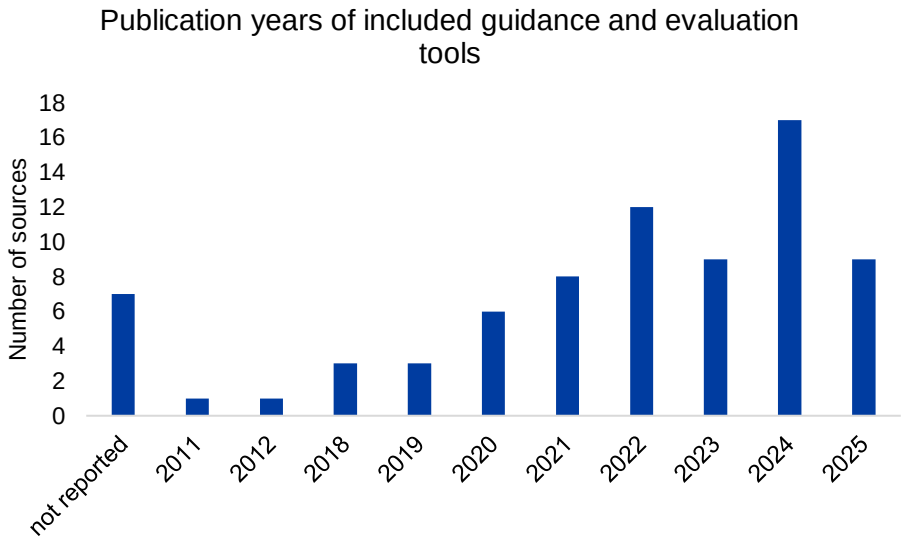

Figure 4: Publication years of included guidance and evaluation tools

Around one third (26) of the guidance or evaluation tools were developed by authors affiliated with institutions in the academic sector. The commercial sector also produced nearly one third (20) of the included resources, while the nonprofit and public sector each contributed eight sources. Every sixth guidance or evaluation tool was developed through a multisectoral collaboration; that is, involving at least two organisations or at least two authors affiliated with organisations from different sectors. The sectoral distribution of the sources is presented in Figure 5.

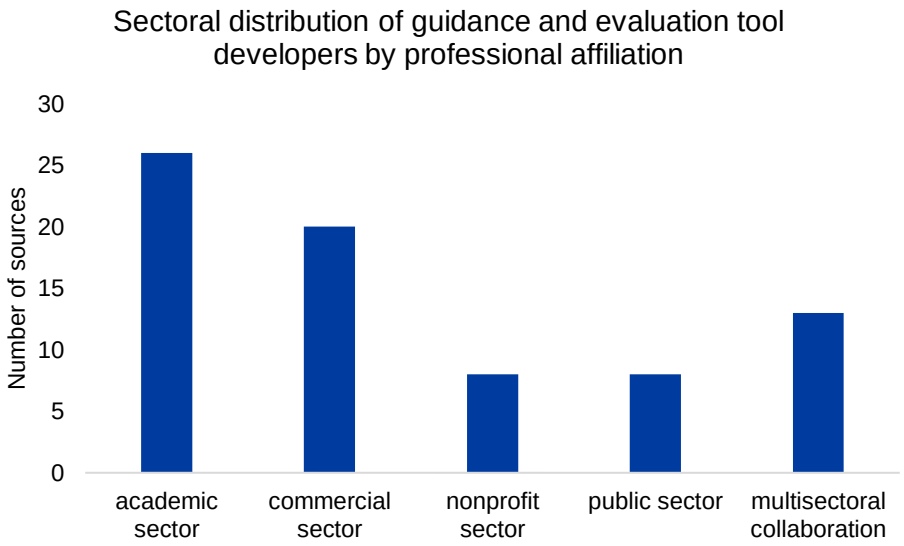

Figure 5: Sectoral distribution of guidance and evaluation tool developers by professional affiliation

Generally, the majority of the sources originated from North America (29 resources) and Europe (16), although organisations and authors from Australia (5), Asia (e.g., Singapore, Saudi Arabia), South America (Brazil), and Africa (South Africa) also contributed to the literature sources (Figure 6). Organisations headquartered in the U.S. or authors affiliated with U.S.-based institutions accounted for the highest proportion (26) of the included sources. They were followed by other countries with English as (one of) their official language(-s), namely the United Kingdom (7 records), Australia, and Canada (5). Around every fifth (15) guidance or evaluation tool was developed through a collaboration involving at least two organizations or at least two authors affiliated with organizations based in different countries.

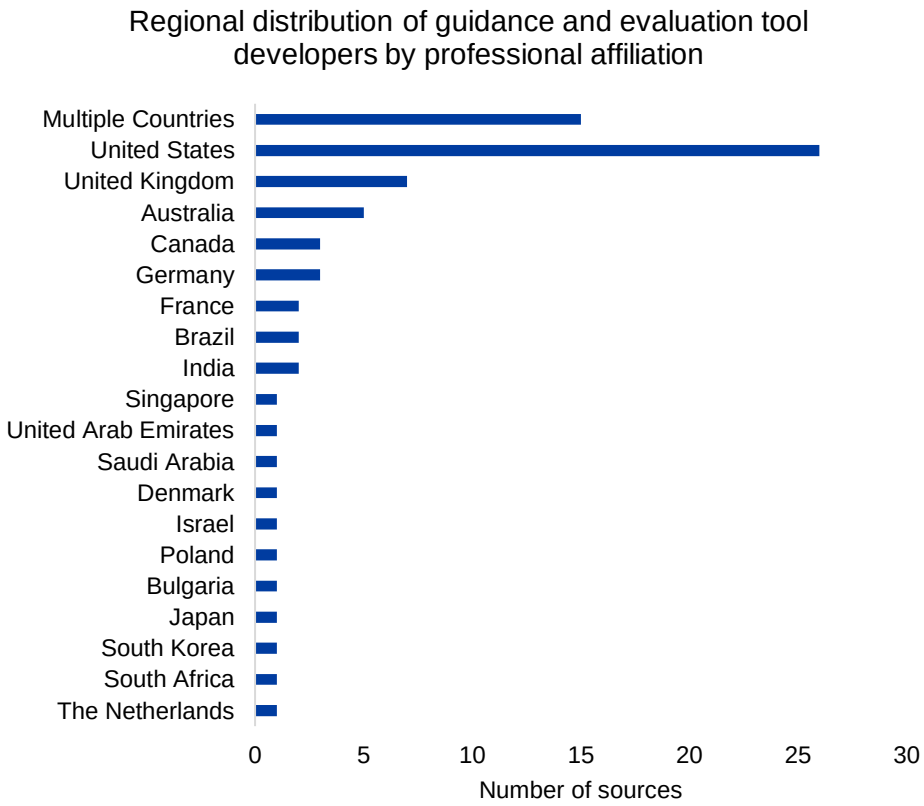

Figure 6: Regional distribution of guidance and evaluation tool developers by professional affiliation

Around one half (39 sources) of the included guidance and evaluation tools are specific to healthcare, while the other half is non-industry-specific (37). Of the healthcare-specific sources, roughly half address healthcare in general, and the other half focus on a specific healthcare domain. 'Specific healthcare domain' includes various ways of differentiating healthcare contexts. The focus of the reviewed sources ranged from hospitals and medical diagnostics to specialties such as radiology and cardiology.

With 30 evaluation tools included, the highest proportion of the literature sources falls under the trustworthy AI category (Figure 7). It was followed by AI maturity (15 records) and vendor evaluation sources (9), which were primarily retrieved from the grey literature via the Google searches. The fewest guidance and evaluation tools were found for usability (5), economic aspects (3), and workflow integration (1). Roughly one in six literature sources was identified as 'cross-categorical' when it could not be clearly assigned to any of the specified categories.

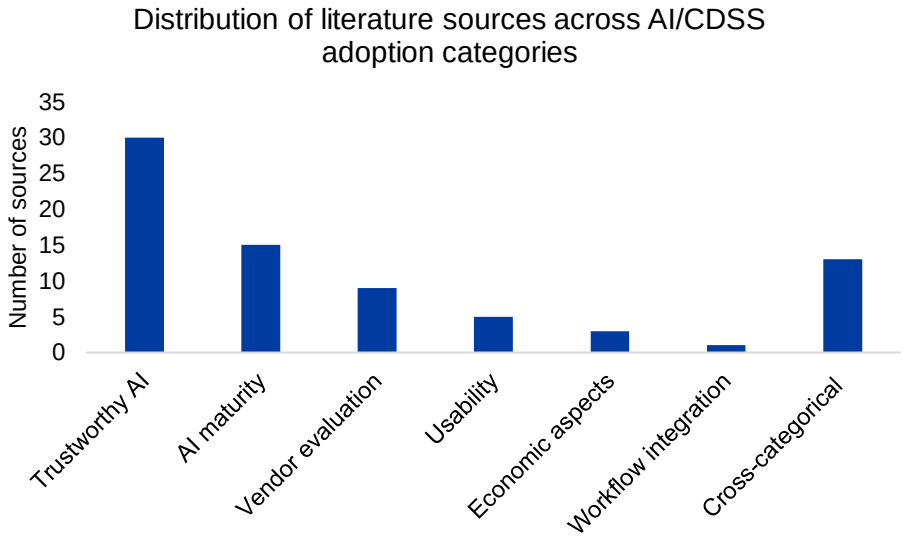

Figure 7: Distribution of literature sources across AI/CDSS adoption categories

All resources can be found in separate tables for each category in Supplementary Material 10 with information on their purposes (e.g., self-assessment of trustworthy AI), the type of literature they were retrieved from (i.e., peer-reviewed literature vs. grey literature), and the sector of the authors' professional affiliations (i.e., academic, commercial, public, nonprofit sector, or multisectoral collaboration). For the purposes of a guidance or evaluation tool, direct quotes were provided using information from the resources to avoid an inaccurate presentation through subjective interpretation. Additionally, a table showing the region of the authors' professional affiliations for each tool can be accessed in Supplementary Material 11.

4.2. Expert interviews

This chapter presents the findings from the expert interviews, organised into five themes.

Structuring the landscape of AI-based CDSS adoption considerations

One recurrent category of AI-CDSS adoption considerations put forth by the experts is legal and regulatory compliance. For AI-CDSS, compliance is required with medical device law, data protection regulations, consumer protection law, and non-discrimination law. Specific regulations named were the EU AI Act, GDPR, and the European Medical Device Regulation. One interviewee pointed out that AI adopters in healthcare have to consider compliance with both supranational and national laws. Two interviewees also stated that legal compliance should be addressed as the first AI-CDSS adoption consideration, as exemplified by the following statement:

> “Legal compliance, this is what I sometimes say, legal compliance is the floor, not the ceiling. So that in in any ways, maybe it helps if you actually do legal compliance first. So your first question is: Does this actually meet the law? If it doesn't, you almost don't need to consider any of the other three categories.” (Expert 2)

It was noted that legal and regulatory compliance is a minimum requirement, as the laws “are very thin” (Expert 2) and do not cover most of AI. One expert distinguished legal compliance from trustworthiness, stating that, while there is a small overlap between them, trust in the AI system’s safety and effectiveness pertains to anything that falls outside of legal guarantees. The interviewee provided the evidence base for an AI-CDSS as an example for this difference.

> “So a really good one would be evidence, right. So medical device law currently, is fairly limited. It doesn't apply to a lot of AI tools and it tends to de-risk them. So it tends to class them as like risk level 1 or risk level 2 rather than like 3 for example. As a consequence, that tends to mean that the actual evidence barrier is quite low. So the amount of evidence that is required to generate, that is required to prove that the thing works is often minimal and it can just come from the software developer. Uhm, trustworthiness would require from a clinician's perspective normally a higher standard of evidence. Uhm, so normally clinicians would want to know that it's been used by somebody else, that it has all of these types of things. Uhm, so that would fall into social acceptability.” (Expert 2)

The evidence that an AI-CDSS provides a healthcare benefit was discussed in every interview. It was emphasised that the AI system should address a need identified in clinical practice, ideally one recognised by the healthcare staff themselves. Both the healthcare need and the evidence for the AI system’s effectiveness in improving patient outcomes and/or reducing workloads are therefore closely linked to its acceptance by those involved in clinical care. The following statements serve as examples:

> “Uhm, and it's also important about is it fulfilling an actual need rather than a want? So clinicians and patients will quite willingly accept the use of new technologies if it's clearly solving a problem or serving a purpose. If it is just the case of somebody in a managerial position saying we really want to buy this software and you better use it, that's not going to really work. So you want to also, ideally you would like the idea to come from the clinicians.” (Expert 2)

> “I mean, it's, it's going to be a successful project if the, finally, if the physician or the nurse really think, uhm, is, is thinks that that it, it will create a less workload and a benefit for his or her patient. I think this is the, the most crucial point. If there's no benefit, there's no, no reason of implementing AI.” (Expert 4)

Another factor influencing healthcare professionals’ and patients’ acceptance of an AI-CDSS is its safety, which every expert mentioned as a relevant consideration. To prevent the AI-CDSS from causing harm to patients (‘non-maleficence’), its accuracy needs to be validated. I It was also noted that specific accuracy requirements (e.g., sensitivity and specificity) will depend on the risks and potential impacts of errors in a given use case. Related to safety are the concepts of trustworthy AI and ethical justifiability, which were mentioned across interviews. As described by the experts, they cover the generalisability of the AI validation and the representativeness of training data for the implementation setting, avoidance of unfair bias and discrimination, privacy and cybersecurity, healthcare professionals’ autonomy and de-skilling, patient autonomy and consent process, the AI system’s impact on the clinician-patient relationship, and transparency. Multiple experts linked transparency to clinicians’ and patients’ trusts into the AI-CDSS. They mentioned the importance of explainability—so that healthcare professionals understand what their decisions are based on—as well as communication to clinicians about how the AI system is used (e.g., how to handle disagreements with the output), and the disclosure of AI use and related information to patients. Transparency and trust were further discussed in relation to the potential impact of an AI system on the clinician–patient relationship. It was emphasised that transparent communication about the system’s use fosters not only trust in the AI itself, but also trust in the clinician using it:

> “I think coming up with some language for how clinicians can explain the use of AI in the particular context to patients in a way that is going to reinforce their relationship and not, uhm, not make the clinician feel, you know, that, uh, that the patient may not trust their own decision, the clinician's decision making or thinking or expertise in a way that doesn't undermine the clinician's expertise, uh, is I think one of the challenges. Like some clinicians may feel that patients, uhm, you know may not have the same respect maybe for the, the clinician if they're use, if they're, if they have to rely on the AI. Uhm, and I think the, I think that giving clinicians the, the language to explain that this is a new tool, it is, uhm, going to you know be for their benefit.” (Expert 3)

Further questions raised include whether the use of the AI-CDSS will overly mediate the clinician–patient relationship and disrupt its empathetic nature. One expert also argued that

patients have an ethical personal right to have AI used in their treatment and that it would be unethical not to use an AI system that is well-trained for a specific question.

Another category of AI-CDSS adoption considerations put forward is the technical feasibility, which includes the availability of the necessary technical infrastructure for use of the system and awareness of differences between the technology used by the AI manufacturer and the technology in the adopter's healthcare practice. It needs be assessed how such differences can impact the AI-CDSS's performance in the implementation setting.

The usability of the system was described as part of technical feasibility, workflow integration, and as a separate AI-CDSS adoption category across the interviews. Considerations mentioned related to the usability are the integration of the AI system with the healthcare provider's electronic health record, the understandability of the output to clinicians and patients, the clinical relevance and actionability of information, a built-in feedback mechanism for clinicians, and explanations of how the AI generated an output. The following statement highlights the importance of the system's usability:

> "So I think kind of thinking about the entire, uhm, life cycle, uh, from a usability perspective is, is really, uh, is, is very important. It's not just about, uhm, it's not just about developing the technology and giving it to the clinician. It's about how do you distil that information for the clinician to make sense of it. How do you then take that information and distil it in a meaningful way for the patient so that it can actually influence clinical care?" (Expert 3)

As described above, the system's usability is related to its implementation and integration into workflows. Further questions that one interviewee assigned to this category are user training for AI use, tailored training and implementation material, the vendor's support channels (e.g., a designated contact, hotline, or chat), contingency plans for system failure put in place by the vendor. Another expert also mentioned change management processes, including the mobilisation of people to adopt AI and the response to their concerns, the clarification of who should use the AI-CDSS, how it should be used, and what the user's responsibilities are, as well as the development of an iterative process for evaluating the AI adoption and collecting feedback. The following statement provides more detail to how the interviewee described the clarification of AI use as a workflow integration consideration:

> "I think the next step is to, uhm, make sure that the technology is easily integrated into the existing workflows. Uhm, to clarify, well who is supposed to be using it, how do they

use it? Uhm, who's accountable if something goes wrong? That's another piece of this that clinicians are going to wanna know about. Uhm, what do they do if they have a problem or they disagree? Uhm, what do they do, what are they supposed to do with the output of it? How are they supposed to explain this to patients?" (Expert 3)

Another AI-CDSS adoption question discussed in two interviews is the post-deployment monitoring of the system. One expert stated that monitoring should be conducted by both the vendor and the deployer, as the vendor's ways of determining success do not always align with what healthcare providers consider best. The interviewee further explained that healthcare providers should take on responsibility for the pre- and post-deployment measurement of outcomes:

> "I don't think that it is necessarily the responsibility of the manufacturer, uhm, to make the pre- and the post-study of, of the, of the system. I honestly, see that more as a responsibility of the, uhm, of the healthcare provider in this case. The reason for that is that you want it to be independent of your manufacturer." (Expert 1)

This pre- and post-deployment measurement of outcomes independent from the vendor requires the deployer to clearly define the outcome measures, which can be healthcare quality outcomes, staff and patient satisfaction, or economic value. It also requires clarity on who will have access to the data being recorded and used by the AI-CDSS, and if this access is sufficient to conduct the planned measurements.

Further adoption questions raised by the experts are the pricing of the system, its cost-effectiveness and return on investment, the reimbursement for AI use, a check for comparable products, liability for errors, and the environmental impact of the AI system. A summarised overview of the AI-CDSS adoption consideration categories is provided in Table 3. The full list of themes and codes is provided in supplementary material 7.

Table 3: AI-CDSS adoption considerations mentioned by the experts

AI-CDSS adoption category	**Considerations**
Legal and regulatory compliance	Medical device law
	Data protection law
	Consumer protection law
	Non-discrimination law
Technical feasibility	Deployer's technical infrastructure

	Awareness of differences between healthcare provider‘s technical infrastructure and technology used for AI development
	Usability
Social acceptability	Fulfilment of a need
	Effectiveness
	Safety
Trustworthy AI/ethical justifiability	Quality of the AI-CDSS (e.g., accuracy)
	Generalisability of AI validation
	Bias and discrimination
	Clinician and patient autonomy (incl. consent)
	Privacy and cybersecurity
	Transparency
	Impact on clinician-patient relationship
	Patient’s ethical right for AI use in a treatment
Usability/user experience	Integration with an EHR
	Understandability of AI outputs to clinicians and patients
	Feedback feature in the system
	Explanation of how the AI generated an output
Implementation and workflow integration	Vendor’s side: training and information for AI users, support channels, contingency plans
	Deployer’s side: change management, clarity about AI use and responsibilities, patient consent, deployment evaluation and feedback
Post-deployment monitoring	Responsibility for both vendor and deployer
	Pre- and post-deployment measurements
Other considerations	Price, cost-effectiveness, ROI, reimbursement, comparable products, liability, environmental impact

Identifying differences between AI-CDSS and AI in healthcare in general

While the interviewees stated that the described adoption considerations already differentiate AI-CDSS from other types of AI in healthcare, they were asked to share further differences. Across interviews, it was emphasised that AI-CDSS are intended to assist with decision-making, not to replace or make clinical decisions independently. However, since they influence care decisions, they are generally considered higher risk than AI used for administrative

functions. Multiple experts also noted that there are risk differences within the category of AI-CDSS—for example, an AI system supporting decisions in preventive care would be lower risk than an AI-CDSS used in an intensive care unit. One interviewee pointed out that the impact on the clinician's decision-making is particularly important for AI-CDSS, compared to other types of AI in healthcare. Thus, the interpretability and usability of the system, as well as its impact on clinicians' work, warrant special attention. The expert further described that while most considerations between AI-CDSS and other types of AI in healthcare are the same, the burden of proof is higher for AI-CDSS given their risk. The following answer was given in response to the question how this higher burden of proof can be reflected in the checklist:

> "Uhm, in terms of, uh, things like: How do you reflect the burden of proof? Uhm, well, that's in the wording of your question. So if it is, uhm, you know, if I was asking something like of an AI scribe, I would just say: Have you got evidence of a usability test? For example. If I was asking something like about clinical decision support, I would want to know: Do you have evidence of the impact on clinical outcomes? So it's the same high level consideration, but it's a different question that reflects the different level of risk." (Expert 2)

Discussing unsolved questions about AI in healthcare

One theme that emerged from the interviews without having a related question in the interview guide is the discussion of unsolved questions about AI in healthcare. One expert described that the liability law regarding errors of an AI-CDSS is currently unclear. Another consideration that the law has not yet addressed is that the evaluation of ML-based CDSS cannot be a one-off exercise, as both the machine learning model and the patient population may change over time (i.e., AI drift). It was also mentioned that there is lack of clarity regarding how to generate evidence for an AI system's effectiveness, as randomised controlled trials— while considered the gold standard of evidence-based medicine—are not particularly suitable for AI-CDSS. As Expert 2 noted, they are "expensive, they take a long time to run, [and] they're not really adaptive". The question of choosing an appropriate comparator for effectiveness (AI-CDSS vs. a different AI-CDSS, AI vs. human, or AI vs. human + AI) is also unsolved. Two experts further pointed out that the emergence of generative AI has raised numerous additional questions that need to be addressed. One interviewee argued for a standardised procedure to consider questions of AI adoption, even if there is not yet a consensus on how to address them:

> "The second problem is that because there is no standardised procedure, they just do it differently every single time. Uhm, so the benefits of having something like a checklist, even if there is not agreement on how you generate evidence, at least you know to

always ask: Uhm, what evidence is there available to, to, to show that it works? What support mechanisms are there in place if something goes wrong?" (Expert 2)

Making the checklist practical and deciding on a practical range for the number of items

Since the four interviews were conducted over five weeks to gather insights for different phases of this research (i.e., planning of the checklist, actual development, and revision of initially identified considerations), the questions about checklist development led to a discussion about the purpose of the checklist in the first interview. It was then decided that the checklist —and the preceding identification of relevant considerations (the long version of the checklist is described in Chapter 4.3)—would not focus solely on the 'to buy or not to buy' question, but rather serve as a synthesis of considerations to guide both this decision and the implementation of an AI-CDSS.

The other three interviews then helped to develop an understanding of practicality for the development of the checklist. The experts emphasised that it should be as simple as possible and not longer than necessary. Two interviewees proposed a range of ten to twenty questions, and another expert also suggested that around twenty items would be appropriate:

> "And with a checklist, you're always balancing. Uhm, you want it to be useful and you also want it to be detailed. So you try and you want to try and ask questions in that checklist that encompass a lot of things, uhm, without having to have like 200 different items on this checklist. Because nobody is gonna use a checklist that's that long. But something that's like 20 items where each of those things represents many things inside them, then people might, might use it." (Expert 2)

It was also pointed out across interviews that the number of questions should also depend on how long it would take to answer them. In addition to an appropriate number of items and the consideration of how time-consuming it is to find answers, one expert also suggested that an electronic version of the checklist with outputs for further guidance based on the user's responses would further enhance its practicality.

Assigning the responsibility for answering the checklist questions

A key question in the interviews was who should answer the questions in the checklist. All experts stated that a checklist with AI-CDSS adoption considerations would require both the vendor and the healthcare provider to answer questions. Specifically, information about the AI system and its quality (e.g., training data, performance testing, biases) needs to be provided by the vendor, while the healthcare provider should answer questions related to the clinical

application of the system (e.g., healthcare need, acceptability of users, workflow integration). Further, the deployer should involve the in-house IT team and the person responsible for data protection issues in answering technology-related questions (e.g., the availability of the necessary technical infrastructure for AI adoption), highlighting that the use of the checklist by a healthcare provider is not a one-person effort. It was thus noted that the checklist should be used at the health system level rather than by individual healthcare professionals. However, as pointed out in the interviews, it is still necessary for healthcare professionals with the relevant domain expertise to be part of the team assessing the AI-CDSS.

4.3. Considerations for the adoption of AI-based CDSS

This chapter presents the synthesis of all identified considerations relevant to the adoption of AI-CDSS, which was primarily created using the literature sources and enhanced with insights from the experts. These considerations are presented as questions, allowing the synthesis to serve as an extended version of or further guidance for the checklist presented in Chapter 4.4. The synthesis includes both questions that healthcare providers should answer and questions that they should ask the vendor (e.g., regarding technical robustness). The defined categories of AI-CDSS adoption considerations and number of questions for each category in the synthesis are provided in Table 4. The synthesis itself with all 227 questions and references can be found in Supplementary Material 12.

Table 4: Categories of AI-CDSS adoption considerations and the amount of questions in the list for each category

Category	**Number of Questions**
Regulatory and Legal Compliance	7
Utility	14
Trustworthy AI: Human Agency and Oversight	10
Trustworthy AI: Technical Robustness and Safety	33
Trustworthy AI: Privacy and Data Governance	12
Trustworthy AI: Transparency	12
Trustworthy AI: Diversity, Fairness, and Non-discrimination	10
Trustworthy AI: Societal and Environmental Well-being	3
Trustworthy AI: Accountability	16
Economic Aspects	10
Usability	12
Workflow Integration	28

AI Maturity	33
Vendor Reliability, Support, and Agreements	27

The first category includes questions about the intended use of the product, compliance with all applicable laws and regulations (incl. AI-specific regulations such as the EU AI Act, medical device regulations, such as the MDR 2017/745, data protection regulation such as the GDPR, and non-discrimination law), and monitoring of this compliance. Moreover, the system's adaptability to new legislation (Fasterholdt et al., 2022), the impact of system modifications on regulatory compliance requirements, and liability were identified as relevant considerations. Another question asks whether the AI-CDSS can account for differences in legal and regulatory requirements across regions when providing decision support (Bottacin et al., 2025).

The 'Utility' category covers clarity about the problem to be solved and the baseline that should be improved, the rationale for using an AI-CDSS and the comparison to other solutions (NHSX, 2020), the measurement of outcomes (Cresswell et al., 2019) and the evidence base for the system's clinical effectiveness (NHSX, 2020), the justification of increased statistical performance in the context of trade-offs (Vollmer et al., 2020), the system's impact on healthcare professionals' work and efficiency (Ghorayeb et al., 2022), the risk of not implementing the AI-CDSS (Awad et al., 2024), and the planning of a process to increase its utility post-deployment (e.g., through user feedback).

'Trustworthy AI' as a category follows the seven requirements presented in the ALTAI (European Commission, 2020). Human agency and oversight as the first requirement, involves questions about the risk of the user's over-reliance on or overconfidence in the system, impacts on the healthcare professionals' decision-making and autonomy, the consideration of human psychology in the system's design (e.g., avoidance of confusion and cognitive biases), and human oversight and control, including the model of oversight (e.g., human-in-the-loop) and specific measures (e.g., detection and response mechanisms for unintended effects; European Commission, 2020).

The technical robustness and safety of an AI system encompasses its resilience to attacks and security (e.g., exposure to cyber-attacks, compliance with cyber-security standards, emergency protocols), safety risks of the system (e.g., technical defects, incorrect medication suggestions) and safety protection (e.g., safety monitoring, contact persons, fault-tolerance), the AI-CDSS's accuracy and performance (incl. internal and external validation, data quality, accuracy and performance monitoring) and implications of low accuracy, its reliability (e.g., robustness against variations or unexpected data inputs) and reproducibility, as well as

contingency and fall-back plans to address unexpected situations or errors of the AI-CDSS (European Commission, 2020).

The subsequent trustworthy AI requirement addresses privacy protection and data governance issues. This includes measures such as conducting a Data Protection Impact Assessment and collecting a minimum quantity of personal information (i.e., data minimisation; (European Commission, 2020). Other considerations in this category relate to procedures in the case of product decommissioning (e.g., access to data; NHSX, 2020).

Transparency as a principle covers the traceability of the AI-CDSS (e.g., logging practices), including the vendor's documentation of the systems' design, development, testing, and validation. Other transparency considerations are explainability (i.e., explanations of how an AI-CDSS arrived at a specific output or recommendation), the vendor's Terms of Use, information that needs to be communicated to different stakeholders (e.g., purpose and limitations of the system, updates, adverse events), the vendor's and healthcare provider's policies on transparency, and the supervision of communication and feedback channels.

'Diversity, Fairness, and Non-discrimination' as a requirement covers measures, such as using training data that is representative of the target population and testing the AI-CDSS for biases to mitigate the risk of unfair bias that systematically and unjustifiably favours or discriminates against particular individuals or groups based on characteristics such as gender and race (European Commission, 2020). A related key consideration is whether the use of the system can disproportionately benefit or disadvantage certain patient groups and contribute to unequal access to healthcare. Further questions for adherence to this requirement are related to the equal access to the system for its intended users and involvement of different stakeholders (e.g., healthcare professionals, patients) in the design, development, deployment, and use of the AI-CDSS.

The subsequent 'Societal and Environmental Well-Being' requirement addresses the risk of fundamental human rights violations and the system's environmental impact (e.g., energy consumption, carbon emissions), as well as measures to reduce this environmental impact (European Commission, 2020).

Accountability is another trustworthy AI requirement and covers the auditability of the system, third-party audits, the compliance with certain standards (e.g., ISO, IEEE) or best practices, the establishment of a review board for AI ethics and monitoring of adherence to the other requirements, processes for the reporting of potential vulnerabilities and adverse impacts of

the system, risk management and redress mechanisms (European Commission, 2020), roles and responsibilities in the use and governance of the AI-CDSS, the development of an AI usage and governance policy, and a plan for the de-implementation of the system if necessary (CHAI, 2024).

A key economic aspect of AI-CDSS adoption is the cost for procuring, deploying, and using the system. This includes costs for the rights of use, system running, system maintenance and updates, evaluation activities, hardware, infrastructure, IT integration, legal and compliance matters, personnel and change management, staff training, and opportunity costs. Evidence for economic evaluation and the quantification of outcomes in monetary terms are also relevant issues. Further questions include familiarity with the reimbursement process, the budget for AI-CDSS deployment and use, the adopter's financial strategy, and insurance coverage for human errors related to the use of the AI system.

Usability as an AI-CDSS adoption category covers the ease of using the system (e.g., navigating through the system with its user interfaces and functions and learning how to use it; Fossum et al., 2011; Ghorayeb et al. 2022), as well as the completeness, understandability, and actionability of information provided by the AI-CDSS (Ghorayeb et al. 2022). It should also be ensured that the system provides all necessary information and interventions without overwhelming the user and causing alert fatigue. Further usability questions ask about the completeness of the functionality of the system (Fossum et al., 2011), its response time (Tegenaw et al., 2023), a built-in feedback mechanism, and access to actions performed by the AI-CDSS and users (Horsky et al., 2012).

The adoption of an AI-CDSS also requires asking questions about how the system integrates into workflows. This involves understanding the impacts on human wok and the specific tasks or actions that healthcare professionals are expected to carry out when using the system (e.g., entering patient information, reviewing recommendations; Bottacin et al., 2025). Questions are also included about the system's alignment with the flow of tasks (e.g., potential redirection of the user from another task), people, information, and the flow of other technologies and tools used in clinical practice (e.g., another CDSS). It is also asked how the AI-CDSS accounts for changes in the workflow over time (Salwei et al., 2021). The availability of the system when needed with the necessary tools and technologies (e.g., HER) and patient data easily accessible, its fit in the workflow of patient visits and interactions between clinicians and the patient, and its fit in the workflow of individual healthcare professionals, the care team, and the broader healthcare organisation are also relevant considerations (Pumplun et al., 2021; Salwei et al., 2021). The redistribution of professional role responsibilities (Horsky et al., 2012) and its

impact (e.g., de-skilling of the staff; European Commission, 2020), the feasibility of major workflow changes if necessary and short-term disruptions through the implementation of the system (NHSX, 2020), the appropriateness of the clinical decision support delivery (e.g., alerts, recommendations, clinical pathways) for the intended healthcare environment and tasks (Horsky et al., 2012), the applicability of advice provided by the system to the implementation setting, as well as barriers and facilitators that may influence healthcare professionals' adherence to the decision support (Van de Velde et al., 2018) are further aspects that need to be considered when acquiring and implementing an AI-CDSS. The integration of the AI-CDSS and the vendor's technology stack with the healthcare provider's IT infrastructure (IT Convergence, 2023), interoperability (incl. integration with an EHR; Aidoc, n.d.; Cresswell et al., 2019), the integration of third-party additions with the system (e.g., external software; Horsky et al., 2012), as well as the possibility to modify it and integrate healthcare professionals' preferences are also asked about.

Key considerations for 'AI Maturity' are healthcare professionals' and patients' acceptance of the AI-CDSS, as well as the support for AI adoption from the leadership team (Pumplun et al., 2021), board, and middle-management (Cisco, n.d.). The synthesis also includes questions about potential conflicts of interest within the adopter's organisation and with other stakeholders (e.g., the vendor; Scott et al., 2021), an AI strategy (Cisco, n.d.), the development of a deployment plan (Intel, 2018), the assessment of differences between the AI development and implementation environment (CHAI, 2024), plans for a pilot project (NHSX, 2020), the establishment of a change management plan (Cisco, n.d.) and an AI committee (Apfelbacher et al., 2024), as well as about the involvement of staff (CluedIn, 2024) in the deployment process and their skills and knowledge to use the AI-CDSS and oversee its operation (Cisco n.d.; Infosys, n.d.; Microsoft, n.d.). It should also be ensured that the healthcare provider has the required personnel and resources to ensure continuous oversight and operation of the AI-CDSS (CNIL, 2022), a process in place for collaboration and open communication between different staff involved in the system's deployment and use (e.g., healthcare professionals, IT department, legal department), sufficient data for AI use (Future Processing, 2024), the necessary technical infrastructure (Pumplun et al., 2021), a data governance framework (Virginia Office of Data Governance and Analytics, 2024; Passerelle, 2024), as well as quality and knowledge management structures (Nortje & Grobbelaar, 2020). One relevant is question is also whether there is a strategy for scaling up the use of the AI-CDSS (e.g., higher patient volume, use across multiple departments of your facility, adaptability to ongoing developments in AI and healthcare)?

‘Vendor Reliability, Support, and Agreements’ covers the management of the relationship with the vendor (Cresswell et al., 2019), the vendor’s support during the deployment (e.g., workflow integration, staff training, system validation in the deployment setting) and use of the AI-CDSS (e.g., system maintenance and updates), their experience and expertise relevant to the system’s use case (e.g., understanding of patient care workflows), communication and feedback channels (e.g., responsiveness and the process for submitting support requests; IT Convergence, 2023; Norton, 2024), the vendor’s incident detection and response procedures, as well as the product’s history (i.e., time on the market, any known past incidents, other customers, and their satisfaction and experiences; Fasterholdt et al., 2022) and the product roadmap (Norton, 2024). It is also important that the vendor can easily explain the AI-CDSS with its technical foundations and purpose (Hosch & Morris, 2024). Furthermore, AI-CDSS adopters should conduct a thorough review of the vendor’s Terms of Service, warranties, and any other contractual service level agreements. Relevant agreements include performance guarantees (Hosch & Morris, 2024), as well as the delineation of accountability and liability for errors or negative outcomes related to the use of the system. This includes, among others, a clear definition of product failure vs. human error (NHSX, 2020), indemnity clauses, and the consideration if the vendor has insurance coverage to support those liabilities (NHSX, 2020; Hosch & Morris, 2024). A key consideration is also whether the ownership (incl. intellectual property rights), usage rights, and licensing terms are clearly defined for both the AI-CDSS and the data (Stout, 2025). Provisions for contract termination and transfer to a different vendor (NHSX, 2020), as well as the healthcare provider’s ability to publish contractual details are further considerations.

Since the aim of this synthesis was to identify all considerations relevant to AI-CDSS adoption, the length of questions was not limited. Table 5 exemplifies that some questions are short, while others include listings of relevant points or sub-questions.

Table 5: Example of questions in the list of AI-CDSS adoption considerations

Usability	Is the AI-CDSS easy to use? This includes easily • navigating through the system with its user interfaces and functions, • becoming productive using its main menu, • seeing all information on the screen (incl. eye-catching display with an appropriate font style and size) and finding specific information, • entering data and completing the CDSS fields, • correcting a data entry error,	Fossum et al. (2011); Silveira et al. (2019); Ghorayeb et al. (2022); Tegenaw (2023)

	• learning how to use the system, • and remembering how to use it.	
	Are nomenclatures and terminologies used in the AI-CDSS clear and consistent?	Horsky et al. (2012), Silveira et al. (2019)

4.4. Checklist for AI-CDSS adoption

This chapter presents the checklist developed for healthcare providers to evaluate the adoption of AI-based CDSS.

Who should use the checklist

The purpose of this checklist is to provide healthcare providers with the 'right' questions that need to be considered in both the decision about the adoption of an AI-CDSS and its implementation in clinical practice. Therefore, the checklist is meant to be used by teams at the health system level who are either evaluating the procurement of an AI-CDSS or preparing for its deployment. The questions were formulated in a way that was generally considered applicable to different types of AI, healthcare areas, provider settings (e.g., large hospitals, 'small' clinics), regions, and agreements with the vendor (e.g., whether ongoing system maintenance and post-deployment support are included). Therefore, individual healthcare professionals in smaller healthcare practices can also use this checklist.

What is covered by the checklist

This checklist reflects the synthesis of AI-CDSS adoption considerations presented in the previous chapter. It includes questions about legal compliance, evidence of the system's effectiveness, the quality of the AI-CDSS, the feasibility and acceptance of its deployment and use, and considerations for planning its implementation and using it in clinical practice.

How to use the checklist

To answer the questions in this checklist, the team responsible for the adoption of the system should request evidence from the vendor and reach out to different departments in the health facility (e.g., healthcare professionals, IT team, legal department). A positive answer to each checklist question does not guarantee a 'right' adoption decision and sufficiently well-carried out implementation practices. Also, the checklist should not imply that adoption of an AI-CDSS is not appropriate if one question is answered with 'no'. Rather, it is important to 1) ask all questions in the checklist, 2) consider the measures that can be taken to address a

consideration if the answer to one question is ‘no’, and 3) consider the specific context where the AI-CDSS should be deployed. The checklist is presented in Table X.

Table 6: AI-CDSS adoption checklist

Number	Item	Who should answer
1	Has compliance with all applicable laws and regulations been ensured (incl. AI-specific laws like the EU AI Act, data protection laws like the GDPR, as well as safety, medical device, and non-discrimination regulations)?	Vendor, healthcare facility
2	What is the evidence that the AI-CDSS improves the clinical outcome you are interested in and how was this evidence generated?	Vendor
3	Were the system validation conditions (e.g., setting, sample, data) representative of your clinical practice?	Vendor
4	Is the deployment and use of the AI-CDSS technically feasible (e.g., having the technical infrastructure, data interoperability)?	Healthcare facility, vendor
5	Is the deployment and use of the AI-CDSS economically feasible (i.e., reimbursement, awareness of costs, economic evaluations)?	Healthcare facility, vendor
6	Is the accuracy of the system high enough for its use case?	Healthcare facility
7	Can the AI-CDSS integrate well into your existing workflows?	Healthcare facility
8	Do healthcare professionals find the AI-CDSS acceptable for use in their healthcare facility?	Healthcare facility
9	Do patients find the AI-CDSS acceptable?	Healthcare facility
10	How is it ensured that your staff will have the knowledge and skills to use (healthcare professionals) and manage (e.g., managers, IT team, legal team) the AI-CDSS?	Healthcare facility, vendor
11	Can it be explained how the AI-CDSS arrived at a specific output or recommendation, and do users understand these explanations?	Vendor
12	Have you considered which information about the system needs to be communicated to your staff and patients?	Vendor, healthcare facility

13	Can the vendor provide evidence of how unfair bias that unjustifiably favours or discriminates against individuals or groups based on characteristics such as gender and race has been mitigated in the system?	Vendor
14	Have you considered all other ways how the deployment and use of the AI-CDSS can harm your patients (e.g., privacy violations), staff (e.g., de-skilling), and the relationship between your healthcare staff and patients? Have you considered measures to reduce these risks?	Vendor, healthcare facility
15	Are there planned measures to respond to adverse impacts of the AI-CDSS (e.g., system failure, workflow disruptions)?	Vendor, healthcare facility
16	Have you considered the liability you carry for claims related to the AI-CDSS?	Healthcare facility
17	Is it clear how the vendor will support you during and after the deployment of the system?	Vendor
18	Do you have a change management plan in place to prepare for the adoption of the AI-CDSS?	Healthcare facility
19	Is there a clear plan for how your vendor and you will conduct post-deployment monitoring of the AI-CDSS?	Vendor, healthcare facility
20	Have you considered how users of the AI-CDSS and patients can provide feedback and raise concerns related to the system?	Healthcare facility, vendor

5. Discussion

The aim of this thesis was to identify guidelines and tools available to healthcare providers to evaluate an AI system before adoption and create a checklist for AI-CDSS adoption based on considerations identified in the literature and expert interviews. Ultimately, 76 literature sources and four expert interviews served as a basis to develop a list of over 230 questions related to AI-CDSS adoption covering legal compliance, trustworthy AI, utility, economic aspects, usability, workflow integration, AI maturity, and the reliability and support of the vendor. Based on this list, a more practical checklist with 20 questions was developed.

While comprehensiveness in a checklist for AI-CDSS adoption is important, it is also crucial to balance this with practicality, as an overly detailed checklist may become impractical to use. The importance of making the checklist as simple as long as necessary and as simple as

possible was emphasised across the interviews and the 20 questions were created based on the experts' suggestions for the number of items. Previously, Vollmer et al. (2020) proposed 20 questions on transparency, replicability, ethical aspects, and effectiveness of AI and ML in healthcare. The lead authors from the National Institute for Health and Care Excellence and the Alan Turing Institute invited a wider group of stakeholders from UK-based and international organisations to engage in discussions about guidance on AI and ML in healthcare over the period of one year. This led to the identification of relevant themes (e.g., transparency, ethics) and the development of a number of questions (Vollmer et al., 2020). While the comparison with their work does not in itself validate the practicality of the checklist developed in this thesis, the fact that a separate group of experts—independent from those involved in this project—also arrived at 20 questions for guiding AI in healthcare suggests that the number of questions chosen here is indeed practical. Another measure to ensure practicality of a checklist, which was described by one interviewee, is having an electronic version that can generate outputs (e.g., recommendations) based on the AI-CDSS adopter's responses to the questions. In fact, the interactivity degree of an evaluation tool was considered a relevant question in the work-in-progress scoping review on trustworthy AI evaluation tools by this author and collaborators. One example of an interactive tool is the AI ethics self-assessment questionnaire by GSMA, which provides recommendations for further action, a summary of the answers, and an ethical score (GSMA, 2022). To enhance the practicality of the checklist proposed in this thesis, the development and delivery of an electronic version with answer-based outputs will be considered.

There may be differences in adoption considerations and the use of the checklist depending on the AI technology and technique(-s) used in a CDSS. For instance, DL models are particularly difficult to interpret (i.e., understanding the decision-making processes of a model), which affects the transparency of AI systems. Differences can also exist based on the application area (e.g., diagnostics vs. clinical documentation), level of healthcare (i.e., primary, secondary, tertiary and quaternary care), healthcare domain (e.g., radiology vs. oncology), size of the healthcare facility, and its region. As noted across the interviews, the risk level of the AI-CDSS is central to the consideration of adoption questions and the burden of evidence required. The degree of advancement of a healthcare area in integrating AI (e.g., radiology is considered to be particularly advanced; Amisha et al., 2019) also influences the requirements for appropriate AI-CDSS adoption, including the evidence that can and should be expected. However, given the absence of a work that synthesises AI adoption considerations into questions that need to be evaluated before adopting an AI-based CDSS, this thesis aimed to provide an evaluation checklist that is applicable to different types of AI-CDSS. Future efforts can adapt this evaluation checklist to make it specific to a specific type of AI-driven CDSS.

All experts mentioned that the responsibility for answering the questions in the checklists should be shared between the healthcare provider and the vendor. For some questions, the vendor therefore needs to provide relevant information to the adopter. Making such information publicly available, would enhance the transparency and verifiability of AI-CDSS. One expert mentioned the publication of a peer-reviewed study as a measure. However, a systematic review, to which the author of this thesis contributed, found a lack of publicly available evidence for ML algorithms in primary care (most of which fell under CDSS for diagnosis or treatment) based on a search of peer-reviewed literature as well as the registration databases of the U.S. Food and Drug Administration and Conformité Européene (Rakers et al., 2024). Therefore, there is indeed a need to publish peer-reviewed studies on different phases of the AI life cycle (incl. development, internal and external validation, economic evaluation) and adhere to the DECIDE-AI reporting guideline for clinical evaluations of AI-based CDSS (Vasey et al., 2022), CHEERS-AI for economic evaluation (Elvidge et al., 2024), or another applicable reporting checklist.

The search results demonstrate that there is a variety of unsolved questions. For example, only three resources on economic aspects were found, and to the best of the author's knowledge, there is yet no guideline or evaluation tool available to support economic evaluations of AI-based CDSS. In fact, it was pointed out across interviews that there are currently numerous unsolved questions, which highlights the need to not only synthesise relevant considerations from different adoption categories in one list but also conduct further research in each area pertinent to AI-CDSS adoption. At the same time, standardised procedures are needed for asking different questions to prevent premature AI adoption even if no consensus exists on issues such as evidence generation, as one of the experts noted. The checklist developed in this thesis serves as such a standardised procedure.

The search also identified records that are thematically relevant but did not meet the definition of practical procedural guidance or an evaluation (e.g., Elhaddad & Hamam, 2024). Therefore, such articles were excluded, even if they discussed relevant AI-CDSS adoption considerations. Additionally, some records described evaluation frameworks or questionnaires that could not be located (e.g., Ji et al., 2021).

Strengths and limitations

The main strengths of this thesis lie in its combination of a literature review with expert interviews and its dual contribution: (1) capturing and presenting relevant AI-CDSS adoption considerations in a comprehensive manner, and (2) providing a checklist with 20 questions as a practical tool to support more informed AI-CDSS adoption. While this was not a systematic

review, both the search strategy from the scoping review and the additional search conducted specifically for this thesis are strengths, as they resulted in the retrieval of 76 relevant resources from peer-reviewed and grey literature across different sectors. For instance, the Google searches conducted for this thesis proved especially effective, identifying more than twice as many included articles as the scientific databases. Records related to AI maturity and vendor evaluation were exclusively identified through the Google searches, and therefore the majority of considerations in these categories would have been missed without them.

This research is not without limitations, some of which stem from the requirements and constraints of this thesis. Although the author received valuable guidance from his supervisor and experts with different academic backgrounds and professional experiences, it is unrealistic for an undergraduate student in health sciences to grasp the different areas relevant to AI-CDSS adoption (e.g., economics, computer science, usability) in their depth and nuance. Moreover, the challenge of planning, conducting, and completing the research within an eight-week period limited both the number of databases searched and the number of interviews conducted. This may have led to some AI-CDSS adoption considerations remaining unidentified. However, while detailed considerations in specific areas (e.g., privacy protection) may be missing, it is reasonable to assume that the risk of not identifying multiple highly relevant AI-CDSS adoption considerations is low, as it is unlikely that 76 resources and four expert interviews would not have included such considerations. It should be noted, though, that there was a trade-off between covering as many relevant resources as possible and ensuring the rigour of their analysis. Although the analysis of literature sources was conducted as planned, the time pressure to categorise items from 76 sources within a few weeks may have led to an inaccurate categorisation of some items (e.g., an item may have been assigned to a previously developed category, even though it represents a distinct consideration). A second round of analysis to review all assigned categories would have reduced this risk. That said, it needs to be mentioned that the majority of considerations appeared in at least two different sources, which effectively led to a re-review of the initially assigned categories during the first, and only, round of analysis. In addition to the time constraint, conducting this research independently meant that no second reviewer was involved, making the selection of literature sources and analysis of data material more prone to bias. For example, this may have led to an exclusion of records that another reviewer might have included. Clear and well-justified inclusion and exclusion criteria facilitated reasonable inclusion and exclusion decisions. These criteria were informed by the scoping review on trustworthy AI evaluation tools, which involved various discussions of conceptual questions to define terms such as ‘procedural’, ‘practical’, and ‘non-industry-specific’. Applying these previously agreed-upon definitions in this thesis reduced the subjectivity of inclusion and exclusion decisions.

Specifically, the findings on the origins of the identified AI adoption resources (i.e., sector, country, and type of literature) may not be representative, as this was not a systematic review aiming to capture all available resources. However, given the strength of the search strategy and the substantial number of resources included, these findings likely offer a reasonably accurate reflection of where guidance and evaluation tools in this area typically originate. Literature published in languages other than English or German was not considered in this work. The possibility that relevant records were missed due to this cannot be excluded. This may also have led to a biased representation of the extent to which different countries contribute to knowledge on AI adoption in healthcare. Despite a potential misrepresentation of individual countries' contributions, the overall finding that most AI adoption-related resources are produced by authors and organisations from developed countries is likely accurate, considering that these countries are at the forefront of AI-related efforts (Maslej et al., 2024). It should therefore be noted that the development of this checklist was shaped primarily by perspectives from developed countries, also including insights from experts based in those countries and the author, who has gained his research experience in AI in healthcare within developed country contexts. This warrants further research to explore the applicability of the checklist to low- and middle-income countries. Moreover, while the experts' conceptualisations of AI-CDSS adoption considerations covered a range of categories, it needs to be mentioned that their perceived importance of considerations may have been shaped by their academic and professional backgrounds (e.g., bioethics). Also, not every adoption category was represented by a dedicated expert. For example, the inclusion of a usability researcher and an economist could have enhanced the list of adoption considerations. A Delphi study involving a larger and more diverse group of experts, representing the full range of relevant disciplines, could further improve both the long checklist of considerations and the final checklist by refining existing items and identifying overlooked ones. It would also allow for more thorough planning of data collection. For example, one limitation of this thesis is that experts had to come up with AI-CDSS adoption considerations spontaneously, which could have resulted in missed considerations, though the experts' feedback for the first checklist draft likely reduced this risk.

Future directions

To summarise the implications derived from this discussion, the following future directions are recommended:

- Clarify further which specific department or person should answer which question in the checklist,

- conduct a Delphi study with a larger number of experts in different areas pertinent to AI-CDSS adoption (e.g., usability and economics) to refine both the long checklist and the 20-question checklist,
- make an electronic version of the checklist with answer-based outputs available,
- test the checklist and its practicality in the real-world,
- research its applicability to low- and middle income countries,
- update the checklist regularly as the AI field and the application of AI in healthcare evolve,
- and transparently report evidence about the AI-CDSS in peer-reviewed studies.

6. Conclusions

Motivated by the absence of an evaluation tool to support the adoption of AI-based CDSS, this thesis aimed to identify relevant guidelines and evaluation tools applicable to the adoption of CDSS and AI in healthcare, and to synthesise AI-CDSS adoption considerations for healthcare providers. Based on 76 literature sources and four expert interviews, a list of over 220 questions and a 20-question checklist were developed to address the need for both a comprehensive synthesis of AI-CDSS adoption considerations and a practical evaluation tool. These considerations were grouped into different categories, namely regulatory and legal compliance, utility, trustworthy AI, economic aspects, usability, workflow integration, AI maturity, and vendor reliability, support, and agreements. The 20-question checklist is an attempt to support both the decision whether an AI-CDSS should be adopted and the deployment of such a system. However, given that this research was conducted as an undergraduate thesis in health sciences, a Delphi study involving experts in each area relevant to AI-CDSS adoption (e.g., usability, economics) would improve both the full list of considerations and the checklist. Additional recommendations, some of which the author of this thesis may pursue, include enhancing the checklist's practicality (e.g., providing an electronic version with answer-based output), testing it in real-world settings, and updating it in response to the evolving application of AI in healthcare.

References

Aidoc. (n.d.). *Your Guide to Navigating the Health System-AI Vendor Relationship*. https://www.aidoc.com/learn/infographic/your-guide-to-navigating-the-health-system-ai-vendor-relationship/ [last access: 24 May 2025]

Al-Zaiti, S. S., Alghwiri, A. A., Hu, X., Clermont, G., Peace, A., Macfarlane, P. & Bond, R. (2022). A clinician's guide to understanding and critically appraising machine learning studies: a checklist for Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML). *European Heart Journal - Digital Health*, *3*(2), 125–140. https://doi.org/10.1093/ehjdh/ztac016

Amisha, Malik, P., Pathania, M. & Rathaur, V. K. (2019). Overview of artificial intelligence in medicine. *Journal Of Family Medicine And Primary Care*, *8*(7), 2328. https://doi.org/10.4103/jfmpc.jfmpc_440_19

Apfelbacher, T., Koçman, S. E., Prokosch, H. & Christoph, J. (2024). A Governance Framework for the Implementation and Operation of AI Applications in a University Hospital. *Studies in Health Technology And Informatics*. https://doi.org/10.3233/shti240527

Awad, S., Loveday, T., Lau, R. & Baysari, M. T. (2024). Development of a Human Factors Based Guideline to Support the Design, Evaluation and Continuous Improvement of Clinical Decision Support. *Mayo Clinic Proceedings Digital Health*, *3*(1), 100182. https://doi.org/10.1016/j.mcpdig.2024.11.003

Ayling, J. & Chapman, A. (2021). Putting AI ethics to work: are the tools fit for purpose? *AI And Ethics*, *2*(3), 405–429. https://doi.org/10.1007/s43681-021-00084-x

Bacchi, S., Kovoor, J., Gupta, A. & Chan, W. (2024). Should this artificial intelligence algorithm be used in my practice now? A checklist approach. *Clinical And Experimental Ophthalmology*, *52*(1), 123–125. https://doi.org/10.1111/ceo.14307

Badman, A. & Kosinski, M. (2024). *What is big data?*. https://www.ibm.com/think/topics/big-data [last access: 24 May 2025]

Bottacin, W. E., De Souza, T. T., Melchiors, A. C. & Reis, W. C. T. (2025). Explanation and elaboration of MedinAI: guidelines for reporting artificial intelligence studies in medicines,

pharmacotherapy, and pharmaceutical services. *International Journal Of Clinical Pharmacy*. https://doi.org/10.1007/s11096-025-01906-2

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brownlee, J. (2020). *Difference Between Algorithm and Model in Machine Learning*. https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/ [last access: 24 May 2025]

Busch, F., Kather, J. N., Johner, C., Moser, M., Truhn, D., Adams, L. C. & Bressem, K. K. (2024). Navigating the European Union Artificial Intelligence Act for Healthcare. *Npj Digital Medicine*, *7*(1). https://doi.org/10.1038/s41746-024-01213-6

California Telehealth Resource Center. (2024). *Healthcare AI Vendor Evaluation Checklist*. https://caltrc.org/artificial-intelligence/aivendorchecklist/ [last access: 24 May 2025]

Centers for Medicare & Medicaid Services. (2024). *Quality Measures & You: Clinicians*. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/QMY-Clinicians [last access: 24 May 2025]

Chae, A., Yao, M. S., Sagreiya, H., Goldberg, A. D., Chatterjee, N., MacLean, M. T., Duda, J., Elahi, A., Borthakur, A., Ritchie, M. D., Rader, D., Kahn, C. E., Witschey, W. R. & Gee, J. C. (2024). Strategies for Implementing Machine Learning Algorithms in the Clinical Practice of Radiology. *Radiology*, *310*(1). https://doi.org/10.1148/radiol.223170

Chen, Z., Liang, N., Zhang, H., Li, H., Yang, Y., Zong, X., Chen, Y., Wang, Y. & Shi, N. (2023). Harnessing the power of clinical decision support systems: challenges and opportunities. *Open Heart*, *10*(2), e002432. https://doi.org/10.1136/openhrt-2023-002432

Cigref & Syntec Numérique. (2018). *Digital Ethics: A Guide for Professionals of The Digital Age*. https://www.cigref.fr/wp/wp-content/uploads/2019/02/Cigref-Syntec-Digital-Ethics-Guide-for-Professionals-of-Digital-Age-2018-October-EN.pdf [last access: 24 May 2025]

Cisco. (n.d.). *Cisco AI Readiness Assessment*. https://www.cisco.com/c/m/en_us/solutions/ai/readiness-index/assessment-tool.html [last access: 24 May 2025]

CluedIn. (2024). *AI Readiness Checklist for Data Leaders and Practitioners*. https://www.cluedin.com/resources/articles/ai-readiness-checklist-for-data-leaders-and-practitioners [last access: 24 May 2025]

Coalition for Health AI. (2024). *Responsible AI Guidance*. https://chai.org/workgroup/responsible-ai/responsible-ai-checklists-raic [last access: 24 May 2025]

Commission nationale de l'informatique et des libertés (CNIL). (2022). *Self-assessment guide for artificial intelligence (AI) systems*. https://www.cnil.fr/en/self-assessment-guide-artificial-intelligence-ai-systems [last access: 24 May 2025]

Correlation One. (2023). *The 6 Essential Building Blocks of AI Readiness*. https://www.correlation-one.com/blog/6-essential-building-blocks-of-ai-readiness [last access: 24 May 2025]

Cresswell, K., Callaghan, M., Mozaffar, H., Sheikh, A., & Digital Health & Care Institute (DHI) (2019). Creating and applying an evaluation framework for the National Decision Support Programme in Scotland. Glasgow: Digital Health & Care Institute. https://doi.org/10.17868/67863

Crigger, E., Reinbold, K., Hanson, C., Kao, A., Blake, K. & Irons, M. (2022). Trustworthy augmented intelligence in health care. *Journal Of Medical Systems*, *46*(2). https://doi.org/10.1007/s10916-021-01790-z

Davenport, T. & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Deloitte. (2024). *AI Readiness & Management Framework (aiRMF)*. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/ai-readiness-and-management-framework.pdf [last access: 24 May 2025]

De Silva, D. & Alahakoon, D. (2022). An artificial intelligence life cycle: From conception to production. *Patterns*, *3*(6), 100489. https://doi.org/10.1016/j.patter.2022.100489

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., De Prado, M. L., Herrera-Viedma, E. & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

Digital Dubai. (2022). *AI System Ethics Self-Assessment Tool*. https://www.digitaldubai.ae/self-assessment [last access: 24 May 2025]

Dignum, V., Nieves, J.C., Theodorou, A. & Aler Tubella, A. (2021). *An abbreviated assessment list to support the Responsible Development and Use of AI*. https://webapps.cs.umu.se/uminf/reports/2021/003/part1.pdf [last access: 24 May 2025]

Elhaddad, M. & Hamam, S. (2024). AI-Driven Clinical Decision Support Systems: An Ongoing Pursuit of Potential. *Cureus*. https://doi.org/10.7759/cureus.57728

Elvidge, J., Hawksworth, C., Avşar, T. S., Zemplenyi, A., Chalkidou, A., Petrou, S., Petykó, Z., Srivastava, D., Chandra, G., Delaye, J., Denniston, A., Gomes, M., Knies, S., Nousios, P., Siirtola, P., Wang, J. & Dawoud, D. (2024). Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence (CHEERS-AI). *Value in Health*, *27*(9), 1196–1205. https://doi.org/10.1016/j.jval.2024.05.006

Emani, S., Swaminathan, A., Grobman, B., Duvall, J. B., Lopez, I., Arnaout, O. & Huang, K. T. (2023). Critically reading machine learning literature in neurosurgery: a reader's guide and checklist for appraising prediction models. *Neurosurgical FOCUS*, *54*(6), E3. https://doi.org/10.3171/2023.3.focus2352

European Commission. (2019). Ethical guidelines for trustworthy AI. High Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai [last access: 24 May 2025]

European Commission. (2020). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment. Publications Office. https://data.europa.eu/doi/10.2759/791819 [last access: 24 May 2025]

European Parliament. (2017). *Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices*. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32017R0745

European Parliament. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

Fasterholdt, I., Kjølhede, T., Naghavi-Behzad, M., Schmidt, T., Rautalammi, Q. T., Hildebrandt, M. G., Gerdes, A., Barkler, A., Kidholm, K., Rac, V. E. & Rasmussen, B. S. (2022). Model for ASsessing the value of Artificial Intelligence in medical imaging (MAS-AI). *International Journal Of Technology Assessment in Health Care*, *38*(1). https://doi.org/10.1017/s0266462322000551

Fossum, M., Ehnfors, M., Fruhling, A., & Ehrenberg, A. (2011). An Evaluation of the Usability of a Computerized Decision Support System for Nursing Homes. *Applied Clinical Informatics*, *02*(04), 420–436. https://doi.org/10.4338/aci-2011-07-ra-0043

Future of Life Institute. (2024). *High-level summary of the AI Act*. https://artificialintelligenceact.eu/de/high-level-summary/ [last access: 24 May 2025]

Future Processing. (2024). *AI readiness assessment: are you prepared for AI integration?*. https://www.future-processing.com/blog/ai-readiness-assessment-checklist/#ai-readiness-checklist-questions-that-you-need-to-answer [last access: 24 May 2025]

Gholamzadeh, M., Abtahi, H. & Safdari, R. (2023). The Application of Knowledge-Based Clinical Decision Support Systems to Enhance Adherence to Evidence-Based Medicine in Chronic Disease. *Journal Of Healthcare Engineering, 2023*(1). https://doi.org/10.1155/2023/8550905

Ghorayeb, A., Darbyshire, J. L., Wronikowska, M. W. & Watkinson, P. J. (2023). Design and validation of a new Healthcare Systems Usability Scale (HSUS) for clinical decision support systems: a mixed-methods approach. *BMJ Open*, *13*(1), e065323. https://doi.org/10.1136/bmjopen-2022-065323

Goirand, M., Austin, E. & Clay-Williams, R. (2021). Implementing Ethics in Healthcare AI-Based Applications: A scoping review. *Science And Engineering Ethics*, *27*(5). https://doi.org/10.1007/s11948-021-00336-3

Government of Serbia. (2023). *Ethical Guidelines for Development, Implementation and Use of Robust and Accountable Artificial Intelligence*.

https://www.ai.gov.rs/extfile/en/471/Ethical%20guidelines%20for%20development%20implementation%20and%20use%20of%20robust%20and%20accountable%20AI.pdf [last access: 24 May 2025]

GSMA. (2022). *Self-assessment questionnaire*. https://www.gsma.com/aiethics-saq/ [last access: 24 May 2025]

Halejak, E. (2023). *ML, NLP, LLM, and Deep Learning Explained – Exploring the Business Potential of AI*. https://inwedo.com/blog/business-potential-of-ai-solutions/ [last access: 24 May 2025]

Han, S.-H., & Choi, H.-J. (2022). Checklist for validating trustworthy AI. *Proceedings of the 2022 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 391–394). https://doi.org/10.1109/BigComp54360.2022.00088

Hehner, S., Biesdorf, S. & Möller, M. (2018). *Digitizing healthcare – opportunities for Germany*. Dusseldorf, Munich, Frankfurt: Digital McKinsey. https://www.mckinsey.com/~/media/mckinsey/industries/healthcare%20systems%20and%20services/our%20insights/digitizing%20healthcare%20opportunities%20for%20germany/digitizing-healthcare-opportunities-for-germany.pdf [last access: 24 May 2025]

Hendriks, M. P., Jager, A., Ebben, K. C., Van Til, J. A. & Siesling, S. (2024). Clinical decision support systems for multidisciplinary team decision-making in patients with solid cancer: Composition of an implementation model based on a scoping review. *Critical Reviews in Oncology/Hematology*, *195*, 104267. https://doi.org/10.1016/j.critrevonc.2024.104267

Higgins, O. & Wilson, R. L. (2025). Integrating Artificial Intelligence (AI) With Workforce Solutions for Sustainable Care: A Follow Up to Artificial Intelligence and Machine Learning (ML) Based Decision Support Systems in Mental Health. *International Journal Of Mental Health Nursing*, *34*(2). https://doi.org/10.1111/inm.70019

Hohma, E. & Lütge, C. (2023). From Trustworthy Principles to a Trustworthy Development Process: The Need and Elements of Trusted Development of AI Systems. *AI*, *4*(4), 904–926. https://doi.org/10.3390/ai4040046

Horsky, J., Schiff, G. D., Johnston, D., Mercincavage, L., Bell, D. & Middleton, B. (2012). Interface design principles for usable decision support: A targeted review of best practices for

clinical prescribing interventions. *Journal Of Biomedical Informatics*, *45*(6), 1202–1216. https://doi.org/10.1016/j.jbi.2012.09.002

Hosch & Morris, PLLC. (2024). *Questions to Ask Your AI Vendor: A Checklist*. https://www.hoschmorris.com/privacy-plus-news/questions-to-ask-your-ai-vendor-a-checklist [last access: 24 May 2025]

IBM. (2021). *What is computer vision?*. https://www.ibm.com/think/topics/computer-vision [last access: 24 May 2025]

Infosys. (n.d.). *What is enterprise AI readiness?*. https://www.infosys.com/iki/topics/enterprise-ai-readiness.html [last access: 24 May 2025]

Intel. (2018). *The AI Readiness Model*. https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/ai-readiness-model-whitepaper.pdf [last access: 24 May 2025]

IT Convergence. (2023). *How to Assess Vendor Capabilities While Selecting AI/ML Solutions*. https://www.itconvergence.com/blog/how-to-assess-vendor-capabilities-while-selecting-ai-ml-solutions/ [last access: 24 May 2025]

ITechLaw. (2021). *Responsible AI Impact Assessment Tool (RAIIA)*. https://www.itechlaw.org/publication/responsible-ai-impact-assessment/ [last access: 24 May 2025]

Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Joshi, I., Cushnan, D. (2020). *A Buyer's Guide to AI in Health and Care. 10 questions for making well-informed procurement decisions about products that use AI*. https://transform.england.nhs.uk/ai-lab/explore-all-resources/adopt-ai/a-buyers-guide-to-ai-in-health-and-care/ [last access: 24 May 2025]

Kavey, R-E. (2009). Healthcare Professionals. In Olsen, L., Goolsby, W.A. M.J. McGinnis: *Leadership Commitments to Improve Value in Health Care: Finding Common Ground: Workshop Summary (2009)* (pp. 117-148).

Kleine, A., Kokje, E., Hummelsberger, P., Lermer, E., Schaffernak, I. & Gaube, S. (2024). AI-enabled clinical decision support tools for mental healthcare: A product review. *Artificial Intelligence in Medicine*, *160*, 103052. https://doi.org/10.1016/j.artmed.2024.103052

Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C. & Faisal, A. A. (2018). The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*, *24*(11), 1716–1720. https://doi.org/10.1038/s41591-018-0213-5

Lifshitz, L.R. & McMaster, C. (2020). *Legal and Ethics Checklist for AI Systems*. https://www.torkin.com/docs/default-source/default-document-library/legal-and-ethics-checklist---lifshitz-mcmaster---nov-3-2020.pdf?sfvrsn=fe4b5ad5_0 [last access: 24 May 2025]

Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J.C., Shoham, Y., Wald, R. & Clark, J. (2024). *The AI Index 2024 Annual Report*. Stanford: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu/report/ [last access: 24 May 2025]

McCormack, L. & Bendechache, M. (2024). A comprehensive survey and classification of evaluation criteria for trustworthy artificial intelligence. *AI And Ethics*. https://doi.org/10.1007/s43681-024-00590-8

Microsoft. (n.d.). *AI Readiness Wizard*. https://adoption.microsoft.com/en-us/ai-readiness-wizard/ [last access: 24 May 2025]

Mimecast. (n.d.). *Questions to ask your AI vendor*. https://assets.mimecast.com/api/public/content/questions-to-ask-your-ai-vendor?v=b35560f3&download=false [last access: 24 May 2025]

Ministry of Economy, Trade and Industry. (2022). *Governance Guidelines for Implementation of AI Principles*. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf [last access: 24 May 2025]

Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Navarro, C. A., Dhiman, P., Beam, A. L., Van Calster, B., Celi, L. A., Denaxas, S., Denniston, A. K., Ghassemi, M., Heinze, G., Kengne, A. P., Maier-Hein, L., Liu, X., Logullo, P., McCradden, M. D., Liu, N., Oakden-Rayner, L., Singh,

K., Ting, D. S., Wynants, L., Yang, B., Reitsma, J. B., Riley, R. D., Collins, G. S., Van Smeden, M. (2025). PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ, e082505. https://doi.org/10.1136/bmj-2024-082505

Morley, J., Floridi, L., Kinsey, L. & Elhalal, A. (2019). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science And Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Mylrea, M. & Robinson, N. (2023). Artificial Intelligence (AI) Trust Framework and Maturity Model: Applying an Entropy Lens to Improve Security, Privacy, and Ethical AI. *Entropy*, *25*(10), 1429. https://doi.org/10.3390/e25101429

National Institution for Transforming India (NITI Aayog). (2021). Self-Assessment Guide for AI Usage. In: *Responsible AI #AIFORALL: Approach Document for India Part 1 – Principles for Responsible AI* (pp. 44 – 49). https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf [last access: 24 May 2025]

Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., Hicklen, R. S., Moukheiber, L., Moukheiber, D., Ma, H. & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*, *2*(6), e0000278. https://doi.org/10.1371/journal.pdig.0000278

NHSX. (2020). AI Buyer's Guide assessment template. [the website for this resource was not accessible anymore by the completion of this thesis].

Nortje, M.A., Grobbelaar, S.S. (2020). A Framework for the Implementation of Artificial Intelligence in Business Enterprises: A Readiness Model. *2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. https://doi.org/10.1109/ICE/ITMC49519.2020.9198436

Norton, B. (2024). *AI Vendor Evaluation: The Ultimate Checklist*. https://amplience.com/blog/ai-vendor-evaluation-checklist/ [last access: 24 May 2025]

Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Open Roboethics Institute. (2019). *Foresight into AI Ethics (FAIE): A toolkit for creating an ethics roadmap for your AI project*. https://openroboethics.org/wp-content/uploads/2021/07/ORI-Foresight-into-Artificial-Intelligence-Ethics-Launch-V1.pdf [last access: 24 May 2025]

Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1). https://doi.org/10.1186/s13643-016-0384-4

Owoyemi, A., Osuchukwu, J., Salwei, M. E. & Boyd, A. (2025). Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study. *JMIRx Med*, *6*, e65565. https://doi.org/10.2196/65565

Passerelle. (2024). *Checklist for AI and ML Readiness*. https://passerelledata.com/checklist-for-ai-and-ml-readiness/ [last access: 24 May 2025]

Personal Data Protection Commission Singapore (PDPC), Infocomm Media Development Authority of Singapore & World Economic Forum. (2020a). *Companion to the Model AI Governance Framework – Implementation and Self-Assessment Guide for Organizations*. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGIsago.pdf [last access: 24 May 2025]

Pumplun, L., Fecho, M., Wahl, N., Peters, F. & Buxmann, P. (2021). Adoption of Machine Learning Systems for Medical Diagnostics in Clinics: Qualitative Interview Study. *Journal Of Medical Internet Research*, *23*(10), e29301. https://doi.org/10.2196/29301

Radić, M., Vienken, C., Nikschat, L., Dietrich, T., Koenig, H., Laderick, L. & Radić, D. (2022). AI-Based Business Models in Healthcare: An Empirical Study of Clinical Decision Support Systems. In *Lecture notes in business information processing* (S. 70–79). https://doi.org/10.1007/978-3-031-17037-9_5

Rakers, M. M., Van Buchem, M. M., Kucenko, S., De Hond, A., Kant, I., Van Smeden, M., Moons, K. G. M., Leeuwenberg, A. M., Chavannes, N., Villalobos-Quesada, M. & Van Os, H.

J. A. (2024). Availability of Evidence for Predictive Machine Learning Algorithms in Primary Care. *JAMA Network Open*, *7*(9), e2432990. https://doi.org/10.1001/jamanetworkopen.2024.32990

Reckers-Droog, V., Enzing, J. & Brouwer, W. (2024). The role of budget impact and its relationship with cost-effectiveness in reimbursement decisions on health technologies in the Netherlands. *The European Journal Of Health Economics*, *25*(8), 1449–1459. https://doi.org/10.1007/s10198-024-01673-3

Reddy, S., Rogers, W., Makinen, V., Coiera, E., Brown, P., Wenzel, M., Weicken, E., Ansari, S., Mathur, P., Casey, A. & Kelly, B. (2021). Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ Health & Care Informatics*, *28*(1), e100444. https://doi.org/10.1136/bmjhci-2021-100444

Riester, M. R. & Zullo, A. R. (2022). Prediction tool Development and Implementation in pharmacy praCTice (PreDICT) proposed guidance. *American Journal Of Health-System Pharmacy*, *80*(3), 111–123. https://doi.org/10.1093/ajhp/zxac298

Rohn, S. (2025). *AI Readiness Framework: How to Prepare Your Organization*. https://whatfix.com/blog/ai-readiness/ [last access: 24 May 2025]

Roll's Royce. (2021). *The Aletheia Framework 2.0*. https://www.rolls-royce.com/~/media/Files/R/Rolls-Royce/documents/stand-alone-pages/aletheia-framework-booklet-2021.pdf [last access: 24 May 2025]

Rusanov, B., Ebert, M. A., Sabet, M., Rowshanfarzad, P., Barry, N., Kendrick, J., Alkhatib, Z., Gill, S., Dass, J., Bucknell, N., Croker, J., Tang, C., White, R., Bydder, S., Taylor, M., Slama, L. & Mukwada, G. (2025). Guidance on selecting and evaluating AI auto-segmentation systems in clinical radiotherapy: insights from a six-vendor analysis. *Physical And Engineering Sciences in Medicine*. https://doi.org/10.1007/s13246-024-01513-x

Salwei, M. E., Carayon, P., Hoonakker, P. L., Hundt, A. S., Wiegmann, D., Pulia, M. & Patterson, B. W. (2021). Workflow integration analysis of a human factors-based clinical decision support in the emergency department. *Applied Ergonomics, 97*, 103498. https://doi.org/10.1016/j.apergo.2021.103498

Saudi Data & AI Authority. (2023). AI Ethics Checklist. In: *AI Ethics Principles* (pp. 41 – 46). https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf [last access: 24 May 2025]

Scalefocus. (n.d.). *Ready. Set. AI. Scalefocus AI Readiness Assessment Survey*. https://www.scalefocus.com/ai-readiness-assessment-survey [last access: 24 May 2025]

Schwab Foundation for Social Entrepreneurship & World Economic Forum (WEF). (2024). *Costs and metrics*. In: *AI for Impact: The PRISM Framework for Responsible AI in Social Innovation* (pp. 25 – 27). https://www.weforum.org/publications/ai-for-impact-the-prism-framework-for-responsible-ai-in-social-innovation/ [last access: 24 May 2025]

Scott, I., Carter, S. & Coiera, E. (2021). Clinician checklist for assessing suitability of machine learning applications in healthcare. *BMJ Health & Care Informatics, 28*(1), e100251. https://doi.org/10.1136/bmjhci-2020-100251

Silveira, D. V., Marcolino, M. S., Machado, E. L., Ferreira, C. G., Alkmim, M. B. M., Resende, E. S., Carvalho, B. C., Antunes, A. P. & Ribeiro, A. L. P. (2018). Development and Evaluation of a Mobile Decision Support System for Hypertension Management in the Primary Care Setting in Brazil: Mixed-Methods Field Study on Usability, Feasibility, and Utility. *JMIR Mhealth And Uhealth*, *7*(3), e9869. https://doi.org/10.2196/mhealth.9869

Stade, E. C., Eichstaedt, J. C., Kim, J. P. & Stirman, S. W. (2025). Readiness evaluation for artificial intelligence-mental health deployment and implementation (READI): A review and proposed framework. *Technology Mind And Behavior*, *6*(2). https://doi.org/10.1037/tmb0000163

Stout, D.W. (2025). *Ultimate Guide to AI Vendor Risk Management*. https://magai.co/ultimate-guide-to-ai-vendor-risk-management/ [last access: 24 May 2025]

Stryker, C. & Holdsworth, J. (2024). *What is NLP (natural language processing)?*. https://www.ibm.com/think/topics/natural-language-processing#:~:text=NLP%20enables%20computers%20and%20digital%20devices%20to%20recognize%2C,with%20statistical%20modeling%2C%20machine%20learning%20and%20deep%20learning [last access: 24 May 2025]

Stryker, C. & Kavlakoglu, E. (2024). *What is artificial intelligence (AI)?*. https://www.ibm.com/think/topics/artificial-intelligence [last access: 24 May 2025]

Sutten, M. (2025). *Are Healthcare Organizations Evaluating AI Vendors Correctly Before Signing?*. https://www.pharmacyquality.com/blogs/are-healthcare-organizations-evaluating-ai-vendors-correctly-before-signing [last access: 24 May 2025]

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N. & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *Npj Digital Medicine*, *3*(1). https://doi.org/10.1038/s41746-020-0221-y

Szabo, L., Raisi-Estabragh, Z., Salih, A., McCracken, C., Pujadas, E. R., Gkontra, P., Kiss, M., Maurovich-Horvath, P., Vago, H., Merkely, B., Lee, A. M., Lekadir, K. & Petersen, S. E. (2022). Clinician's guide to trustworthy and responsible artificial intelligence in cardiovascular imaging. *Frontiers in Cardiovascular Medicine*, *9*. https://doi.org/10.3389/fcvm.2022.1016032

Tegenaw, G. S., Amenu, D., Ketema, G., Verbeke, F., Cornelis, J. & Jansen, B. (2023). Evaluating a clinical decision support point of care instrument in low resource setting. *BMC Medical Informatics And Decision Making*, *23*(1). https://doi.org/10.1186/s12911-023-02144-0

The George Washington University. (n.d.). *What Is a Clinician and How Do I Become One?*. https://healthcaremba.gwu.edu/blog/what-is-a-clinician [last access: 24 May 2025]

The University of Edinburgh. (2013). *Interview Consent Form*. https://edwebcontent.ed.ac.uk/sites/default/files/imports/fileManager/Interview_Consent_Form.pdf [last access: 24 May 2025]

Torres, C. S. R., Sánchez, C. B. & Narita, T. (2021). Ethical Assessment of AI for Actors within the Entrepreneurial Ecosystem: Application Guide. In: *Inter-American Development Bank eBooks*. https://doi.org/10.18235/0003269

TrustArc Canada Inc.. (2024). *Responsible Artificial Intelligence Checklist*. https://trustarc.com/wp-content/uploads/2024/05/Responsible-AI-Checklist-.pdf [last access: 24 May 2025]

Van De Velde, S., Kunnamo, I., Roshanov, P., Kortteisto, T., Aertgeerts, B., Vandvik, P. O. & Flottorp, S. (2018). The GUIDES checklist: development of a tool to improve the successful use of guideline-based computerised clinical decision support. *Implementation Science*, *13*(1). https://doi.org/10.1186/s13012-018-0772-3

Van Smeden, M., Heinze, G., Van Calster, B., Asselbergs, F. W., Vardas, P. E., Bruining, N., De Jaegere, P., Moore, J. H., Denaxas, S., Boulesteix, A. L. & Moons, K. G. M. (2022). Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. European Heart Journal, 43(31), 2921–2930. https://doi.org/10.1093/eurheartj/ehac238

Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., Watkinson, P., Weber, W., Wheatstone, P. & McCulloch, P. (2022). Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*, e070904. https://doi.org/10.1136/bmj-2022-070904

VDE Verband der Elektrotechnik, Elektronik und Informationstechnik e.V.. (2022). *VCIO based description of systems for AI trustworthiness characterisation VDE SPEC 90012 V1.0*. https://www.vde.com/resource/blob/2242194/a24b13db01773747e6b7bba4ce20ea60/vcio-based-description-of-systems-for-ai-trustworthiness-characterisationvde-spec-90012-v1-0--en--data.pdf [last access: 24 May 2025]

Vetter, D., Amann, J., Bruneault, F., Coffee, M., Düdder, B., Gallucci, A., Gilbert, T. K., Hagendorff, T., Van Halem, I., Hickman, E., Hildt, E., Holm, S., Kararigas, G., Kringen, P., Madai, V. I., Mathez, E. W., Tithi, J. J., Westerlund, M., Wurth, R. & Zicari, R. V. (2023). Lessons Learned from Assessing Trustworthy AI in Practice. *Deleted Journal*, *2*(3). https://doi.org/10.1007/s44206-023-00063-1

Virginia Office of Data Governance and Analytics. (2024). *AI Data Readiness Checklist*. https://digitalgovernmenthub.org/examples/ai-data-readiness-checklist/ [last access: 24 May 2025]

Wang, H. E., Landers, M., Adams, R., Subbaswamy, A., Kharrazi, H., Gaskin, D. J. & Saria, S. (2022). A bias evaluation checklist for predictive models and its pilot application for 30-day hospital readmission models. *Journal Of The American Medical Informatics Association*, *29*(8), 1323–1333. https://doi.org/10.1093/jamia/ocac065

White, N. M., Carter, H. E., Kularatna, S., Borg, D. N., Brain, D. C., Tariq, A., Abell, B., Blythe, R. & McPhail, S. M. (2023). Evaluating the costs and consequences of computerized clinical decision support systems in hospitals: a scoping review and recommendations for future practice. *Journal Of The American Medical Informatics Association*, *30*(6), 1205–1218. https://doi.org/10.1093/jamia/ocad040

Wilson, P. M., Ramar, P., Philpot, L. M., Soleimani, J., Ebbert, J. O., Storlie, C. B., Morgan, A. A., Schaeferle, G. M., Asai, S. W., Herasevich, V., Pickering, B. W., Tiong, I. C., Olson, E. A., Karow, J. C., Pinevich, Y. & Strand, J. (2023). Effect of an Artificial Intelligence Decision Support Tool on Palliatie Care Referral in Hospitalized Patients: A Randomized Clinical Trial. *Journal Of Pain And Symptom Management*, *66*(1), 24–32. https://doi.org/10.1016/j.jpainsymman.2023.02.317

Writer. (n.d.). *AI readiness: The essential quiz and framework for modern enterprises*. https://writer.com/blog/ai-readiness-framework/ [last access: 24 May 2025]

You, J. G., Hernandez-Boussard, T., Pfeffer, M. A., Landman, A. & Mishuris, R. G. (2025). Clinical trials informed framework for real world clinical implementation and deployment of artificial intelligence applications. *Npj Digital Medicine*, *8*(1). https://doi.org/10.1038/s41746-025-01506-4

Use of AI tools for the thesis

Tool	Reason for use	Where it was used	Link
ChatGPT	Improve the flow of sentences and correct grammatical mistakes	Introduction, Background, Methods, Results, Discussion, Conclusions	https://chatgpt.com/
DeepL	Improve the flow of sentences and correct grammatical mistakes	Introduction, Background, Methods, Results, Discussion, Conclusions	https://www.deepl.com/translator
QuillBot Grammar Checker	Check text passages for grammatical mistakes	Introduction, Background, Methods, Results, Discussion, Conclusions	https://quillbot.com/grammar-check

Eidesstattliche Erklärung

„Ich versichere, dass ich in dieser schriftlichen Studienarbeit alle von anderen Autor*innen wörtlich übernommenen Stellen wie auch die sich an die Gedankengänge anderer Autoren*innen eng anlehnenden Ausführungen meiner Arbeit besonders gekennzeichnet und die entsprechenden Quellen angegeben habe. Zusätzlich versichere ich, dass ich beim Einsatz von KI-gestützten Schreibwerkzeugen diese Werkzeuge in der Rubrik „Übersicht verwendeter Hilfsmittel“ mit ihrem Produktnamen, meiner Bezugsquelle (z.B. URL) und Angaben zu genutzten Funktionen der Software sowie zum Nutzungsumfang vollständig aufgeführt habe. Davon ausgenommen sind diejenigen KI-gestützten Schreibwerkzeuge, die von meinem zuständigen Prüfungsamt bis zum Zeitpunkt der Abgabe meiner Studienarbeit als nicht anzeigepflichtig eingestuft wurden („Whitelist“). Bei der Erstellung dieser Studienarbeit habe ich durchgehend eigenständig und beim Einsatz KI gestützter Schreibwerkzeuge steuernd gearbeitet.“ (Limburg et al.: „Plagiarismus in Zeiten Künstlicher Intelligenz.“ ZFHE 17(3) S. 91–106: 103)

Hamburg, den 26.05.2025
Sergej Kucenko

Supplementary Material 1: Search strategy

Search strings and search dates for each database and Google search

Database	Search String
PubMed April 21, 2025	(“artificial intelligence” OR “machine learning” OR “deep learning” OR “clinical decision support system*”) AND (“usability” OR “user experience” OR “economic*” OR “cost-effectiveness” OR “workflow” OR “interoperability” OR “readiness” OR “maturity” OR “suitability” OR “vendor” OR “adopt*” OR “deploy*” OR “implementation” OR “uptake” OR “purchas*” OR “procur*” OR “buy”) AND (“guideline” OR “evaluation framework” OR “checklist” OR “questionnaire” OR “evaluation model” OR “self-evaluation” OR “self-assessment”)
Scopus April 27, 2025	Searched within article title: (“artificial intelligence” OR “machine learning” OR “deep learning” OR “clinical decision support system*”) AND (“usability” OR “user experience” OR “economic*” OR “cost-effectiveness” OR “workflow” OR “interoperability” OR “readiness” OR “maturity” OR “suitability” OR “vendor” OR “adopt*” OR “deploy*” OR “implement*” OR “integration” OR “uptake” OR “purchas*” OR “procur*” OR “buy”) AND ("guideline" OR "framework" OR "checklist" OR "questionnaire" OR "evaluation model" OR "self-evaluation" OR "self-assessment")
Google: Economic aspects 5 – 6 April, 2025	(“artificial intelligence” OR “machine learning” OR “deep learning” OR “clinical decision support system”) AND (“economic*” OR “cost-effectiveness” OR “cost effectiveness” OR “cost-utility” OR “cost utility” OR “cost-benefit” OR “cost benefit”) AND (“guideline” OR “framework” OR “checklist” OR “questionnaire” OR “evaluation model” OR “self-evaluation” OR “self-assessment”)
Google: Usability 7 May, 2025	(“artificial intelligence” OR “machine learning” OR “deep learning” OR “clinical decision support system”) AND (“usability” OR “user-friendl*” OR “user friendl*” OR “ease of use” OR “user interaction” OR “user interface”) AND (“guideline” OR “framework” OR “checklist” OR “questionnaire” OR “evaluation model” OR “self-evaluation” OR “self-assessment”)
Google: Usability (CDSS focus) 8 – 9 May, 2025	("clinical decision support system") AND (“usability” OR “user-friendl*” OR “user friendl*” OR “ease of use” OR “user interaction” OR “user interface”) AND (“guideline” OR “framework” OR “checklist” OR “questionnaire” OR “evaluation model” OR “self-evaluation” OR “self-assessment”)

Google: UX 9 – 10 May, 2025	(“artificial intelligence” OR “machine learning” OR “deep learning” OR “clinical decision support system”) AND (“user experience” OR “UX”) AND (“guideline” OR “framework” OR “checklist” OR “questionnaire” OR “evaluation model” OR “self-evaluation” OR “self-assessment”)
Google: Workflow integration 11 May, 2025	(“artificial intelligence” OR “machine learning” OR “deep learning” OR “clinical decision support system”) AND (“clinical workflow*” OR “clinical practice” OR “healthcare practice” OR “interoperability”) AND (“guideline” OR “framework” OR “checklist” OR “questionnaire” OR “evaluation model” OR “self-evaluation” OR “self-assessment”)
Google: AI maturity 11 – 12 May, 2025	(“artificial intelligence” OR “machine learning” OR “deep learning” OR “clinical decision support system”) AND (“readiness” OR “maturity”) AND (“guideline” OR “framework” OR “checklist” OR “questionnaire” OR “evaluation model” OR “self-evaluation” OR “self-assessment”)
Google: Vendor evaluation 13 May, 2025	(“artificial intelligence” OR “machine learning” OR “deep learning” OR “clinical decision support system”) AND (“buy*” OR “procur*” OR “purchas*” OR “adopt*” OR “uptake” OR “deploy*” OR “implement*” OR “suitability” OR “vendor”) AND (“guide*” OR “framework” OR “checklist” OR “questionnaire” OR “evaluat*” OR “assess*”)

Reference articles to verify the relevance of each search string (i.e., these articles were taken as an indicator that the search string can retrieve relevant articles)

Database/Search	**Reference articles**
Google search: economic aspects	“Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence (CHEERS-AI)”
	“Machine Learning Methods in Health Economics and Outcomes Research-The PALISADE Checklist: A Good Practices Report of an ISPOR Task Force”
	“Economic evaluation for medical artificial intelligence: accuracy vs. cost-effectiveness in a diabetic retinopathy screening case”
Google search: usability	“A Checklist for the Usability Evaluation of Artificial Intelligence (AI) mHealth Applications Graphical User Interface”
	“Evaluation Framework for Successful Artificial Intelligence–Enabled Clinical

	Decision Support Systems: Mixed Methods Study”
Google search: usability (CDSS focus)	“Assessing the Usability of a Clinical Decision Support System: Heuristic Evaluation”
	“Design and validation of a new Healthcare Systems Usability Scale (HSUS) for clinical decision support systems: a mixed-methods approach”
	“Usability of clinical decision support systems”
Google search: user experience	“Evaluating acceptance and user experience of a guideline-based clinical decision support system execution platform”
Google search: workflow integration	“The Clinical Practice Integration of Artificial Intelligence (CPI-AI) framework”
	“FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare”
Google search: AI maturity	“Self-Assessment Tools Self-assessment tools to determine the AI maturity level of hospitals”
	“AI Readiness Wizard”
	“AI Readiness & Management Framework”
Google search: vendor evaluation/cross-categorical resources	“Clinician checklist for assessing suitability of machine learning applications in healthcare”
	“Healthcare AI Vendor Evaluation Checklist”
	“AI Vendor Checklist: Top Capabilities to Assess”
PubMed	“Design and validation of a new Healthcare Systems Usability Scale (HSUS) for clinical decision support systems: a mixed-methods approach”
	“FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare”

	“Clinician checklist for assessing suitability of machine learning applications in healthcare”
Scopus	“Artificial Intelligence (AI) Trust Framework and Maturity Model: Applying an Entropy Lens to Improve Security, Privacy, and Ethical AI”
	“FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare”
	“Clinician checklist for assessing suitability of machine learning applications in healthcare”

Supplementary Material 2: Interview guide

First of all, thank you so much for agreeing to take part in this research.

I would like to start by explaining a bit about this research. This research is a bachelor's thesis project and aims to identify questions or considerations that healthcare providers need to think about and evaluate before adopting an AI-based clinical decision support system or CDSS. These considerations will be synthesised into a checklist that healthcare providers can use to make informed adoption decisions. In the context of this work, 'healthcare providers' refers to individual healthcare professionals delivering healthcare, such as clinicians, or those involved in adopting an AI system at the organisational health facility level. Therefore, in addition to searching for AI evaluation literature, this research also seeks insights from individuals with knowledge of AI in healthcare and/or the adoption of AI in health facilities. The interview will take about 50 minutes (changed to 40 minutes from the second interview onwards). The main goal of it is to hear your perspective on which considerations need to be evaluated before adopting an AI-based CDSS.

I will record this interview, so that I can transcribe it later. Names of people and organisations will be removed and not mentioned in the transcript. The information collected will be used for research purposes only. Do you consent with the recording?

Do you have any questions about this thesis project or the interview before we start?

1. *If you were to define different categories for all the questions or considerations that healthcare providers need to think about and evaluate before adopting an AI-based clinical decision support system (or CDSS), what would those categories be?*

2. The interviewee will now be asked to go through each category and say which questions need to be considered in these different categories. For example: "Now if you think about trustworthiness: what specific questions or considerations do healthcare providers need to think about and evaluate to ensure that an AI-based CDSS is trustworthy?"

    - Depending on the specificity of the considerations mentioned, follow-up questions will be asked. For instance, AI fairness is a specific consideration of trustworthy AI but within trustworthy AI, it is a broad category of issues. In that case, the interviewee will be asked: "And within the category of AI fairness: what are

questions that healthcare providers need to have answered to ensure that the AI-based CDSS is 'fair'?"

- Depending on how specific these considerations are to AI-based CDSSs (rather than mentioned generally related to AI in healthcare), follow-up questions will be asked. For example: "If you think about the fairness considerations you mentioned: how do considerations about the fairness of AI in healthcare in general and an AI-CDSS in particular differ?
- If not mentioned, the interviewee will be asked about preidentified categories

3. *How would you approach the development and use of a checklist with adoption considerations for AI-based CDSSs?* (added after the first interview because this was asked spontaneously and deemed relevant for further interviews)

    - What would be an appropriate range for the amount of questions?
    - Who should answer the questions?
    - What are other measures to make the checklist as practical as possible? (all three added for the fourth interview because they were relevant in the prior interviews)

4. After each category, the interviewee will be asked how they think a specific consideration should be evaluated. All the considerations mentioned by the interviewee will be summarised for that. For example: "You mentioned fairness, explainability privacy, human oversight, and technical robustness as trustworthy AI issues and you also mentioned specific considerations within these trustworthy AI categories. I would like to go through these considerations with you and ask you what methods or literature sources can a healthcare provider use to ensure that these issues have been properly considered and addressed before adoption of an AI-CDSS. Let us start with fairness. How would you as a healthcare provider ensure that the AI-CDSS you or your organisation adopt is fair?" (this question was not asked anymore after the first interview because, in response to the previous question, every expert stated that the sources or methods to ensure a consideration has been addressed involve requesting information from the vendor and various departments within the healthcare facility)

I have asked all my questions. Do you have any comments or topics that you would like to address that we have not talked about but could be relevant for our study?

Thank you so much for taking the time for this interview.

Supplementary Material 3: Interview consent form

Hamburg University of Applied Sciences (HAW Hamburg)
Faculty of Life Sciences, Department of Health Sciences

Interview Consent Form

Research project title: "Adoption of AI-driven Clinical Decision Support Systems: A Checklist for Healthcare Providers Based on a Narrative Review of AI Evaluation Resources and Expert Interviews"

Research investigator: Sergej Kucenko, student in Health Sciences (B.Sc.)
Research participants name:

Thank you for your interest in participating in this research. This research is a bachelor's thesis project and aims to identify questions or considerations that healthcare providers need to think about and evaluate before adopting an AI-based clinical decision support system. These considerations will be synthesised into a checklist that healthcare providers can use to make informed adoption decisions. In the context of this work, 'healthcare providers' refers to individual healthcare professionals delivering healthcare, such as clinicians, or those involved in adopting an AI system at the organisational health facility level. Therefore, in addition to searching for AI evaluation literature, this research also seeks insights from individuals with knowledge of AI in healthcare and/or the adoption of AI in health facilities.

The interview will take about 50 minutes. Risks associated with your participation are not anticipated by the research investigator. You have the right to stop the interview or withdraw from the research at any time.

This consent form is necessary to ensure that you understand the purpose of your involvement and that you agree to the conditions of your participation. Please read this information sheet and then sign this form to certify that you approve the following:

- the interview will be recorded with Microsoft Teams and a transcript will be produced
- you can request the interview transcript and correct any factual errors
- the transcript of the interview will be analysed by Sergej Kucenko as research investigator
- access to the interview transcript will be limited to Sergej Kucenko, as well as Professor York Zöllner and Professor Walter Leal as the examiners of this bachelor's thesis
- In any summary interview content, or direct quotations from the interview, that are made available through academic publication or other academic outlets, neither you nor your organisation will be named. However, your role will be mentioned (e.g., "bioethicist and post-doctoral researcher in digital health"). It will be ensured that other information in the interview that could identify yourself is not revealed.
- the actual recording will be deleted immediately after the submission of the thesis for examination
- any variation of the conditions above will only occur with your further explicit approval

By signing this form I agree that;

1. I am voluntarily taking part in this project. I understand that I don't have to take part, and I can stop the interview at any time;
2. The transcribed interview or extracts from it may be used as described above;
3. I have read the Information sheet;
4. I don't expect to receive any benefit or payment for my participation;
5. I can request a copy of the transcript of my interview and may make edits I feel necessary to ensure the effectiveness of any agreement made about confidentiality;
6. I have been able to ask any questions I might have, and I understand that I am free to contact the researcher with any questions I may have in the future.

Printed Name

Participants Signature Date

Researchers Signature Date

Contact in case of questions or concerns: Sergej.Kucenko@haw-hamburg.de

Supplementary Material 4: Assigned categories in the evaluation of literature sources

This material is provided as a separate document.

Supplementary Material 5: Interview transcripts

Interview 1 – 4 April 2025– via Microsoft Teams

Expert 1: Postdoctoral Researcher working in eHealth and medical AI

Discussed and agreed that the transcription will be started with the recording via Microsoft Teams *[00:00 - 00:21]*

Expert: *[00:22 – 00:27]* Okay. All yours. I'm about to send you the, the inform consent form.

Interviewee: *[00:29 – 01:57]* Perfect. Thanks so much. Uhm, yeah. So first of all, thank you so much for agreeing to take part in this research. Uhm, yeah. I'd like to start explaining a bit about this research first. Uhm, yeah, research the bachelor's this project and it aims to identify questions or considerations that healthcare providers need to think about and evaluate before adopting an AI-based clinical decision support system or CDSS. These considerations will be synthesised into one checklist that healthcare providers can use to make informed adoption decisions and in the context of this work, healthcare, healthcare providers refers to either individual healthcare professionals delivering healthcare, such as clinicians, or those involved in adopting an AI system at the organisational health facility level. Therefore, in addition to searching for AI evaluation literature, this research also seeks insights from individuals with knowledge of AI in healthcare and, or the adoption of AI in health facilities. So yeah, the interview will take about 50 minutes. The main goal of it is to hear your perspective on which considerations need to be evaluated before adopting an AI-based CDSS, and I will record this interview so that I can transcribe it later. Names of people and organisations will be removed and not mentioned in the transcript. The information collected will be used for this research only. So first of all, do you consent with the recording?

Expert: *[01:58 – 01:58]* Yeah. Thanks.

Interviewer: *[01:59 – 02:07]* Thanks so much. And do you have any questions about this thesis project or the interview before we start?

Expert: *[02:08 – 02:16]* Nope. I think it's okay. Uhm, is, are you planning to publish the results beyond your thesis?

Interviewer: *[02:17 – 02:25]* I would like to. I haven't yet figured out how to, uhm, how to do it yet, but, but it would be ideal to do it.

Expert: *[02:26 – 02:26]* Thanks.

Interviewer: *[02:27 – 02:44]* Okay. So the first question is, if you were to define different categories for all the questions or considerations that healthcare providers need to think about and evaluate before adopting an AI-based CDSS, what would those categories be?

Expert: *[02:45 – 13:35]* Now, that's a difficult question [laughing]. Uhm, I think, uhm, that, I think that the first, the first consideration or the first obligation that you have is to not cause harm. So you need to evaluate the security of your system. And the security I don't mean the data security, but that it is not causing harm to patients. So there you have aspects regarding the accuracy of your system, for example. What type of claims is the system making? And, uhm: Who are the users? Uhm, and if you need to make, regarding to the users, you need to make changes in your workflow and in your in your personnel. Uhm, I probably I should have started with compliance instead of security, so I think that's probably the number one would be compliance. Uhm, so you would have… If you're talking about clinical decision support systems in AI, then you have a whole, a whole battery of, of things that you should comply with, including the GDPR. Or, uhm, for data security and privacy, you have the AI Act. And probably most of your clinical decision support system of this sort would fall under the, the high risk category, uhm, of the AI act, which means that you would have to comply with the highest standards. You would need to be registered. You would need to have a [not understandable] in your system, etcetera. Uhm, and apart from that, you have not only the supernational, but you have the local, the local legislation. And you would have to take a look to see what is your, what is your institutional policy regarding this. Uhm, so yes, so I will start by compliance. Then I would go to the security part, so not causing harm. Non-maleficence if, if you take the, the more ethical wording. Uhm, there is, there is something that is very important for adopting clinical decision support systems in, in practice. Uhm, and this is what is the what is the value of your of your system. Uhm, we are quite I don't want to say stranded, but the system is struggling a lot to determine maybe what is, how to determine value and what is value. Uhm, and this is

important because you will not get things reimbursed unless you are capable of showing what type of value you have. Uhm, so that will depend a little bit on type of clinical decisions support system and the clinical pathway that you are, that you applying the clinical decision support system. But to have some kind of determination of value, for example if there is a cost-effectiveness analysis already being done, that, that is fantastic. But bad news is that most of the, most of the digital health technologies in general do not possess something like this, and it's difficult to perform. So, uhm, and even if a cost-effectiveness analysis has been performed, you need to take a look to see if it is really translatable, uhm, into your particular system. So, uh, both if there is that's great. If there, but if there is you need to check if that is applicable to your, to your population and to your workflow or to your clinical setting. Uhm, now the other the other thing that is, uhm, maybe regarding compliance, sorry, going back to that, compliance is not important just for the sake of compliance itself, but also I think one very important aspect is how do you go about privacy? And even go a little bit beyond, beyond what the regulation says. And that is important because the regulation acts as a, as a minimum requirement in order to, uhm, to protect citizens or to protect, to protect in general. So we can also protect institutions. So there is, for example, regarding, uhm, regarding privacy is, is one aspect. For example, there's also for clinical decision support systems, I think it's very important to think about transparency, and that is linked to, for example, to the right of the information and the right of information is stated in the GDPR, for example. But it is also stated in the case of Holland, for example in legislation that governs the doctor-patient relationship. Uhm, so there's a there's a relationship there between how to, how to, how to comply with, with the national, with the, with the supernational or the EU level legislation. But there's also something that that has to do with, with what are the, what are the best practices as a doctor? So I think that for clinical decision support systems, it's very important to first determine if the clinical decision support system is only to be used by the doctor. Or if it's going to be used by the doctor and the patient together. And depending on who is the, the edge-user to see if all the obligations, doctor to patient, or doctor to yeah, doctor to patient, or manufacturer to doctor, or manufacturer to patient are being, are being successfully complied and addressed. I'm trying to get my daughter making a little bit of a mess. Let me see. So we've said cause harm, compliance, the value, the value of the cost-effectiveness. Uhm, it is important to determine as well if there is some kind of reimbursement option or reimbursement route already planned for the clinical decision support system.
That's important because you need to, at the end of the day have sustainable, a sustainable product. Uhm, we talk about data and, and privacy. There's also an important thing for, uhm, for me, that is also related to the evidence that you have available about the clinical decision support system or in general anything you're going to apply to clinical practice. Uhm, it would be great to have available what kind of evidence, hard evidence is there about all the claims

made by the by the, by the manufacturers. And the problem sometimes is that this is seen as a, as a trademark, uhm, trademark protected or IP or sensitive information. So you don't always have access to all this information. So that's an issue. And also the other issue is to determine or to have very clear who is gonna have access to the data being recorded and used by the by clinical decision support system and if you would have access to that. And that is important for you to be able to make your own evaluations and don't depend only on the evaluations of the manufacturer. We talked already about transparency. Uhm, he other, the other nice thing to know when when you're trying to bring something into clinical practice is what has been done regarding implementation of the system. So I don't see it necessarily as a requirement, 'cause it may be possible that you're looking at a very new product. There are certain steps that should have been taken or explored by the manufacturers anyhow. Uhm, and that is to ensure that you will, that you have from the manufacturer the support and enough information in order to apply this into your workflow. For example, has it, with whom has it been tested? What kind of users is it? For example, has it been tested for accessibility in different groups? Uh, if, for example, patients are supposed to also have access to the information provided for the for the clinical decision support system? What is the workflow that they foresee? I don't. Yeah. Sorry, I'm talking a lot.

Interviewer: *[13:36 – 13:51]* No, no, I'm listening. So you mentioned the support by the vendor. Uhm, can you further elaborate on specific questions or aspects that, uhm, you need to consider, which kind of support should be provided?

Expert: *[13:52 – 18:18]* Yeah, it's difficult because they're very different types of clinical decision support systems. But if you think that you're gonna adopt something into clinical practice, there are different types of clinical decision support systems. They may be, there may be some that are very easy to use, right? For example, if the, if the scores being used by professionals are already known, if the scales are already known. Uhm, if the only thing that is happening is that, the processing of the data is, is different, it may be possible that they don't need much training, for example. But, uhm, I think anyhow you need training, uh, or what we should check. The thing is, like I said, it's a little bit difficult because it's, it's a wide range of, of things, but let's just generalise it, maybe some of these are not applicable to other cases. Uhm, but for example, if training is necessary, so if it's really new for your people, then you need to have the right training. And I think that that should be done by the, by the manufacturer, or at least with the support manufacturer, or being guided by the manufacturer. There should be materials with information about a system available, and that is because, okay, you have a training, but trainings are forgotten or someone goes on holiday and then comes back, and there are new people coming in. Maybe there are people that are just passing by in the, in the

department. So I do think the materials are necessary as well. Depending on the on the level of risk of the clinical decision support system, then you can think about a 24/7 support. Uhm, the chat or, or telephone or something where you can just call the people if it's, if this is not working. It's important for you to know what kind of measurements are being taken if the, if the, if the system stops working. So what kind of contingency plans, uhm, what kind of contingency plans does the, the manufacturer has in place in order to contain the damage of the AI system is not working anymore. Uhm, oh, I was, I was thinking that monitoring is, is another aspect that they should be looking at though. Is, for example, is your AI not changing through time? Uhm, Is it stable as you expect? Uhm, etcetera. So post-implementation, I think one of the requirements should be what is being done regarding monitoring. Uhm, okay, going back to support, hmm, yeah, uhm, then I just think that implementation in the technical, in the technical sense of implementing something is also important. As I mentioned there are things that may be very easy to implement or there already kind of set up for your system, but there maybe others that are a little bit more tricky. Uhm, and it would be nice to know if those implementation and training materials are actually tailored in any, in, in any way. The reason for that is that maybe training people in, in the UK is not gonna be the same as training them here in, in the moment, like or in Sweden or etcetera. So maybe those things.

Interviewer: *[18:19 – 18:52]* So to summarise you you mentioned different, uhm, categories. So the first one is compliance, the second one is cause no harm, the third one is value where you also pointed out cost effectiveness. Uhm, and then the fourth one which relates to the value is the hard evidence about the claims of the manufacturer and the fifth one is what has been done regarding the implementation of the system, which particularly includes the support also by the vendor if I understood that correctly.

Expert: *[18:56 – 18:58]* Yeah, I think, uhm, sorry, once again?

Interviewer: *[18:59 – 19:43]* So the five categories just to summarize for me to correctly understand are compliance for one, which includes specific regulations, and also covers different aspects such as privacy, transparency. Uhm, and the second one is the the system shouldn't cause harm. The third category being what's the value of the system. Uhm, specifically also cost-effectiveness considerations. The fourth category relating to the third one is hard evidence about the claims of the manufacturer and the fifth one is what has been done regarding implementation of the system, so also particularly the support of the vendor. So the first question here, did I understand that correctly?

Expert: *[19:44 – 19:44]* Yeah, I think so. [laughing]

Interviewer: *[19:45 – 19:46]* Great and sorry.

Expert: *[19:47 – 19:51]* And, and. Well, yeah, yeah. No, go ahead. Go ahead.

Interviewer: *[19:52 – 20:23]* Yeah. And then my next question is, you've also mentioned that there is a broad range of clinical decision support systems, but if you would try to differentiate AI systems for clinical decision support from AI and healthcare in general. Uhm, how would you, what would you say, what consideration should be or are specific to AI-based CDSS as opposed to AI in healthcare in general?

Experts: *[20:24 – 21:02]* Sorry, I didn't understand the question. I am though a little bit surprised that I didn't mention ethics in, in the requirements in the requirements [laughing] that I just mentioned. The thing is like regarding ethics and and trustworthiness, you can see that some of this, or all these things that we have been talking about is embedded or is, is, uhm. is considered into this, this ethical transformation aspect. But yeah, sorry about that. [laughing]

Interviewer: *[21:02 – 21:04]* Mm hmm. No, no worries.

Expert: *[21:05 – 21:16]* Now you asked me what considerations of AI clinical support, uhm, of clinical decision support systems that use AI…which considerations are?

Interviewer: *[21:17 – 21:38]* So the question was, uhm, how would you say, or what considerations are specific to AI systems, uhm, which are specifically developed for clinical decision support as opposed to considerations when deploying AI in healthcare in general.

Expert: *[21:39 – 25:15]* Hmm, yeah. So AI in general is really, really broad, so I think, uhm, so I don't really find the way to the to the answer. Uhm, I think maybe what, what is key about the clinical decision support system is that, first of all, it's meant to assist if you are, if you literally take the name. Uhm, so it's, it's not supposed to make a decision for, to assist a healthcare professional to make a decision. So it is, it is, yeah, it is a very small part of all the types of AI in healthcare. What makes it very special is that it sits at a decision point between doctor and, uhm doctor and patient and to make decisions about what is gonna happen. The very important considerations there are related to autonomy, and the autonomy is you're not only of the patient, but also of the doctor because it's going to aid someone take a decision. So it's going to influence a decision at the end of the day. So, uhm, there are, even if you, if you do everything right, but if you do not, uhm, if you do not foster that, that doctors feel safe, for

example, for making decisions to go against the clinical decision support system advice, for example, and how to protect them and how to include those in a healthy working environment, you would probably have issues. So I think, I think that's very special from, from clinical decision support systems, it would depend a lot as well on the risk level of the decision being made. So it's not the same thing, a clinical, clinical decision support system that you're using for prevention, for very early prevention or for well-being, or that you're using to determine the, the risk of a heart attack in an ICU. So, uhm, yeah. So I, I, if I would have to rescue one thing, I would, I would talk about autonomy and deskilling when you said the, the clinical decisions, in comparison to, to the healthcare in general. Uhm. and if you're talking about that, then it's very important how transparent and explainable the AI system is, because otherwise you won't be able to know what you're basing your decisions on. And again, if you are, if the system is meant to bridge doctor and patient and the situation gets a little bit messier because the information needs to be clear not only for the patient, but also for the doctor. Uhm. and there is the questions about how do you inform your patient that you're using this type of system? Uhm, and that would depend on the workload that you have. And how accessible this is for the, for the patient. Uhm, yeah. Does that help a bit?

Interviewer: *[25:16 – 25:42]* It helps, of course. [laughing] Do you? Can you think of any other considerations that are specifically relevant for clinical decision support? Uhm, or if you hear about considerations that are often mentioned, uh, in with regards to AI in healthcare and then if you think are there any specific questions for clinical decision support?

Expert: *[25:43 – 26:01]* Mm hmm. Umm. [with pauses]

Interviewer: *[26:02 – 26:03]* Yes, sorry I phrased it. Phrased it horribly. [laughing]

Expert: *[26:04 – 27:06]* No, no, no, no, no. [laughing] The, The thing is like AI in in healthcare for example, you can also have prognostic, you have diagnostic tools, you have. So it's very close to things that clinical decision super systems can do. So as something that differentiates this apart from the, from the, the information or the decision part, I cannot think of, of anything that seems to me to be very particular ‘cause the other things I can find them in other type of, of AI for healthcare. Hmm, I don't know. I still. I still argue for the ones that I put forward. Uhm.

Interviewer: *[27:07 – 27:40]* That's fine. Yeah. Before we move on to the next question, maybe, so you've mentioned, uhm, some categories of considerations. Uhm, have you? Do you wanna elaborate on a specific, do you wanna elaborate more on a specific category or? Uhm, for example, yeah, looking at compliance, ethics, the support of the vendor, the value of the

system. Are there any specific considerations within these categories that you haven't mentioned yet?

Expert: *[27:41 – 31:42]* Hmm. Maybe, for example, it's important to consider the, the value may not always be economic, and that not all value has been, uhm, that we have been able to calculate in economic terms, what value is. Uhm, but that's something that we need to work on, I think in general. There are a lot of people looking into this. So I hope that we soon enough will get some kind of, yeah, better model for determining value, uhm, for digital health in, in general including clinical decision support system, AI clinical decision support systems. Uhm, yeah. So the, the last of the, of the aspects which you mentioned at the end is the, is the monitoring part, so the post-deployment, the post-deployment monitoring. And I do think it's very important to, to do that. Uhm, monitoring should be done not only by the manufacturer, but also by the, uhm, by the deployer of the systems, in this case the, the healthcare provider. The reason for that is that although manufacturers have their own ways of determining how successful they are, that doesn't always, uhm, match what, what we consider to be the best thing. So maybe regarding their monitoring, uh, needs to be addressed, not only the, and again, this depends on the level of risk of the I system, but if it's a life/death situations then I do think that also the, the deployer should, should do part of this in the monitor. Uhm, regarding AI, like the trustworthiness and ethics aspect, I do think is, is difficult and it hasn't been kind of streamed as other, as other aspects. Uhm, there are questions I, I think that we should, we should, we need to, to answer regarding the inclusiveness, inclusiveness of, of the systems. Uhm, things regarding to bias, discrimination, that we should, you know, like we cannot avoid them. I think that's, that's the nice thing. But that brings me back to the point that this ethics part kind of links to all the other parts together. So for example, you should be able to monitor these aspects throughout the whole, also after deployment, for example. Uhm, and I do think that one way or the other one, you cannot forget that you should be asking the people using the system, so it's not, the top-down approach is not always, is not always the best one. And I think that, for example, if you want to promote trust of your, of your users and your patients in the system, then you should spend some time asking yourself if this type of systems are promoting trust in the healthcare system and in your doctors or is doing exactly the opposite.

Interviewer: *[31:43 – 32:24]* One question regarding post-deployment monitoring. So you mentioned that manufacturers have their own way of monitoring an AI system post-deployment, and how would you say or how does the monitoring differ between the manufacturer and the healthcare providers? So if I'm a healthcare provider and I would need to consider, ensure a trustworthy or a suitable monitoring. What questions should I look for,

uhm, what, so what question should I look for, for my own monitoring? And what should I know that the manufacturer is also monitoring? So again phrased, yeah.

Expert: *[32:25 – 36:16]* Yeah, yeah. So I think that again that, that depends so much depending on the AI system you're looking at. For example, uhm, we had one example about an automated triage system for emergencies. And basically what the provider was, uh, or was the manufacturer was providing the, the institution was, uhm, user, uhm, tevredenheid [dutch word], uhm, how happy were users with the system, for example. Uhm, the amount of things handled and how were they handled. How long did it take and comparisons, for example with how long is it, a telephone, telephone calls that do the same thing triage. So for example, it's it's very hard because, uhm, direct comparisons of these two groups, for example, the people that call and the people that go online, is very different. So actually you cannot compare them because your population is different. However, the manufacturer makes this direct comparison of people that go with people that, that go in, go into the digital system. So this kind of methodological considerations, when you present the data are very important, because that means that you're communicating to the person that buys your system or buys your service. That, uh, this is, this is how you compare it and that this is okay and this is how you should be measuring it. But when you take into account that those two populations, for example, are not comparable, then you cannot do this. Or at least you should be able to say look, this is not, this cannot be a direct comparison but, but this is, this is what the, in this case what the manufacturer was doing. Uhm, the same with the, with the reports about the usability of the system. How happy were the, were the people with the system? Then questions about, for example, who are exactly these users? You know, like how, how was chosen, what were the, what were the, the measurements to, or the, the measures to ensure that your study was not biased. Those kind of things are very, are very tricky to get from your manufacturers in some cases, especially if they're the ones making the studies themselves, themselves. Now the other thing that you have is that you don't have the report. So if there is something, uhm, that is being done by a third party, then they usually don't give you the full report. They just cherry pick the, the questions or the answers that they want to show and that's, that's it. But yeah, again that would depend a lot on on what is important for you and your clinical case. And I think that that's something probably that you need to ask yourself. So, for my clinical decision support system, what are my outcome measures and how do I want to measure them? Yeah, along the way. If you don't have cleared that in the beginning, for example, that’s something that we see a lot, you, your comparison pre- and post-implementation are going to be difficult. Because if you're prepared to do that, then you would say, oh, I'm interested to measure, for example: How often my patients did have to return to tertiary care from, from the GP? And then you would, you would be able to ensure that your data's adequate to make that analysis afterwards. What we

see a lot is that people come and implement the system and don't know how they're gonna measure it if they're successful or not.

Interviewer: *[36:17 – 36:40]* Right. So if I understand you correctly, the healthcare provider needs to have information on how specific outcomes or, uhm, aspects were measured pre-deployment, uhm, to then also use the same measure to, to keep track of it post-deployment. So if we look at patient outcomes and usability. Did I understand you correctly?

Expert: *[36:41 – 38:35]* So what I was trying to say is like, so there are two different things. So one is what the, what the manufacturer has done. And the other thing is what I'm going to do once I implement it. So if, I don't think it's the, I don't think that it is necessarily the responsibility of the manufacturer, uhm, to make the pre- and the post-study of, of the, of the system. I honestly, see that more as a responsibility of the, uhm, of the healthcare provider in this case. The reason for that is that you want it to be independent of your manufacturer. So you cannot ask your manufacturer to do that because then it's gonna be measured by the, using their own. And if you're going to, if you want to determine what is the success of your system, you need to know what is your baseline. So you, you as, as institution need to sit down and think what are my outcome measures? What is what I want to measure? What is, what is important for me? What is important for my hospital? And that can be clinical, that can be from the side of, for example, from the how happy is your, is your personnel? How, uhm, do there's something that is called the quadruple aim, uhm, and for example you can, you can use it as a guide to say, okay, what are my, is my patient happy? Is my people happy? Are my clinical outcomes as I expect and am I economically doing better than before? So that that may be a way to go, but that would, that would depend on what is the important things for you as an institution or as a doctor of the, of the technology. Sorry, did I, did I came…[laughing]

Interviewer: *[38:36 – 38:37]* Oh no, it makes total sense.

Expert: *[38:38 – 38:40]* Did I do it understandable? Okay.

Interviewer: *[38:41 – 39:12]* Yeah, it was understandable. One question regarding usability. So you've mentioned usability now as a pre deployment consideration which, uhm, the manufacturer should ideally measure. If I'm not mistaken, you haven't mentioned usability as a major category before, uhm, so are there, so would you also consider it as one adoption category? And if yes, what are other specific questions or aspects that one should look into before adopting an AI-based clinical decision support system.

Expert: *[39:13 – 39:19]* Yeah, for me usability falls under implementation. So it's one of the aspects of implementation.

Interviewer: *[39:20 – 39:35]* And are there any specific usability questions that I should look into as a healthcare provider?

Expert: *[39:36 – 43:15]* Hmm. Uhm, yes, there are quite a lot of. Usability is quite a well-studied subject. So there are already, quantitative instruments, for example, for checking what is the, the usability of, of your system. Uhm, let me see if I, clinical decision support system for, well, I, there's something that for example, there, there are things that is, is that there, there is something called, for example, the intention of the use. So how prepared would you be to, to use a particular system? So if you're talking a clinical about a clinical decision support system, you could talk not only about using the system because that would be, that would be part of of your practices in the hospital or in your, in, in your institution. But it could also, for example, determine what is your, uhm, your intention to use the information being provided by the system. So how good does it does it feel? Uhm, so there's the intention of use. The other, the other question for example is, is it making my life easier. So it's helping me to, to perform my my job better and faster or not. Because if it's not helping then, then you need a good reason why not. So, it can be, for example it can be that a clinical decision support system comes to identify a gap in care that didn't, wasn't addressed before. So it's not necessarily bad. But if that's the case, then you need to adapt your workflow to this, because you cannot give more work to people that are already working a lot. Uhm, there are things regarding the actionability of the information, for example. Uhm, and the reason for that is that a clinical decision support system can give you a lot of information, but it may not necessarily be the information necessary for your healthcare system. So each healthcare system has its own peculiarities and their decision, although the clinical standards are supposed to be international in general, there are variations between countries, so it is important to know that your system does provide clinically relevant and actionable information. Uhm, depending on the clinical decision support system, if you have an interface, for example, that provides information or provides, uhm, like interactive ways of looking at the at the data or the results, uhm, then I think that you should take a look to see if the system is understandable, if it's, if people can do something with it. Uhm, hmm. Maybe this helps? [laughing]

Interviewer: *[43:16 – 43:31]* Yeah, that helps. Yeah. So before we move on to the next question, is there any category or consideration that you haven't mentioned or that you maybe came up with now, that you haven't thought about at the start of this interview?

Expert: *[43:32 – 43:36]* I get very uncomfortable when you tell me that. I think like, what am I forgetting? [laughing]

Interviewer: *[43:37 – 43:38]* Sorry, I don't want pressure you. [laughing]

Expert: *[43:38 – 44:23]* No, no, no, no, I'm joking. It's okay. It's like when a professor says like: “Are you sure that you are not forgetting anything? Yeah, no, I, I, I think that's, yeah, I guess it depends on how, how broad or, or narrow you interpret some things that we have been talking about, as what happened with implementation and usability. Uhm, but no, I think, hmm, yeah, the only thing that I, that I would add is the, the monitoring.

Interviewer: *[44:24 – 44:46]* Okay. And if you would take all these categories and considerations and synthesise them to one checklist that a healthcare provider can use to make an informed adoption decision and consider the relevant aspects of AI-CDSS, uhm, how would you approach creating such a checklist? So yeah.

Expert: *[44:47 – 45:33]* Hmm. Yeah. It’s tricky. And it's, it's for only use of, so do you foresee it as for example I would submit this checklist to the manufacturer before, and then the evidence would be and information would be seen by the healthcare provider is this? Because there, there is a lot of information that the healthcare provider will not automatically have about a system or it will probably be very hard for him or her to find in the documentation about the system.

Interviewer: *[45:34 – 46:21]* Yeah. So how I think of the checklist for now is the healthcare provider would receive the checklist first and then he would have to look into different sources of information. I think the manufacturers, probably the main one. Uhm, as you mentioned, there is information that the healthcare provider probably doesn't have, but he can also use the checklist to, uhm, use other methods or to discuss and consider some considerations internally, within the healthcare facility, without communicating to the manufacturer. So I would say I see this checklist as one list where you would basically derive information from different sources to assess these considerations.

Expert: *[46:22 – 47:00]* Yeah. Because then I, I think you're you're moving more like, for example to the HTA or the auditability systems. Then, then a checklist by itself, because, for example, you could say, okay, what I want is to check what information, for example, do

healthcare providers need to ask to the manufacturers. Uhm, but I, I do think that that is very time consuming, so I, I would, I would consider a checklist, which would be pre-filled by the, by the manufacturer, which will be checked by the healthcare provider and then the, the questions there, the still remain, and the evaluation of the evidence that they provide would be kind of graded or evaluated by the healthcare provider. Uhm, but it's it, then it goes a little bit more beyond the, the kind of the, the assessment that you would, that you would do. For example if you're only interested in, uhm, for example if you're setting is, you would come in contact with with the system and you're doing a pilot and you want to evaluate the clinical decision support system based on your own experience. So I think those are, those are two options that you would, that you could have like. Uhm, one just like that, the experience of the healthcare provider with the system, and the other one would be, uhm, yeah, more, more on the auditability or the, or the healthcare, or the health technology assessment style, let's say.

Interviewer: *[48:22 – 48:53]* So if I understand correctly, you would, so there is, uhm you could use the checklist to send it to the manufacturer and he would fill out as much information as possible and the remaining, and you would first also consider the information that was filled out, but then also probably there is some information left that isn't for the manufacturer to fill out or where there might might not be any evidence, which should then be considered internally or in communication with other stakeholders.

Expert: *[48:54 – 51:44]* Yeah, it, it can also help you tell the manufacturer what you need. Because maybe they're also, for example, if you say look, I don't see any evidence about implementation of the system for example in primary care. You have implemented it before in in other settings. Uhm, would you invest in, in, in, in an for example implementation study? Would you do that together with us? Uhm, and then you can get an, an arrangement of how to do that. For example, how to do a pilot before you actually buy the system. Uhm, so those kind of things, so it can it can help you find gaps. But it's, it's very different, right. So it wouldn't be a checklist if you do it that way. It would be really like an evaluation instrument by which you request evidence to the to the manufacturer that will be checked by the healthcare provider and then decisions will be made about how, how sufficient is that evidence in how happy you are with that evidence? And the checklist sounds more to me like what's something that you would do once the system, yeah, based on what you know about the system, or, or based on yeah, so there, the burden of, of both of them is very different. Uhm, and the quality of the information that you would get is also very different. Because, for example, if my checklist says has the provider is, is the provider compliant with the GDPR or with the, has a CE mark, this is really easy to do, right? Because it's just a yes or no question. Uhm, so the, you could just write an e-mail and say: “Hey guys, you're CE compliant, right?” Or you could just check in on

the website to see if CE compliant or not. But, uhm, there are other things that are a little bit more difficult to determine. Uhm, for example, if you have to, have to check the recent studies they have done. Or to get to know where has it been implemented before. What kind of data did they use to develop it? What was your training data? What was your, uhm, where did you validate, externally validated your, your system? Yeah, those kind of things… what are the issues that you have had before? Yeah. Yeah, I think both have strengths and limitations.

Interviewer: *[51:46 – 51:47]* Right, got it.

Expert: *[51:48 – 51:48]* Yeah.

Interviewer: *[51:49 – 52:14]* So if you would think about this evaluation instrument, how would you make it as practical as possible? So one question is perhaps also how many questions, it's of course, it probably depends on how broad you, you define these categories, but maybe if you think about practicality, how would you make it as practical as possible?

Expert: *[52:15 – 54:08]* Mm hmm. Yeah. So I think it's a, oh, yeah, it's a, 'cause, I think it's like if you're doing this for an institutional level, then it's maybe not that bad, right? But if you're doing this from a healthcare, so one doctor that ia interested in one tool for one patient, that is, that's quite different. I, I do like the idea of having a comprehensive evaluation which, in which you would request information to the, to the, to the manufacturer. But it's not a checklist. If, if there would be a checklist, like I think I could comfortably go through 30 questions. I think you could, you can maybe stretch it to 50 questions, depending on how difficult the questions are, or because sometimes you have to choose from and then you have 20 options, right? So those type of questions are, are more difficult than if just, just a yes or no or a scale or, or something like that. But yeah, having, having 10 questions for so many aspects, that would mean two questions per aspect. So that is, uhm, that is little. So for example, if you, if you have seven or eight. Yeah, probably like seven or five questions per, per item or something like that. As an average.

Interviewer: *[54:09 – 54:09]* And if you would?

Expert: *[54:10 – 54:10]* So it's 50 questions plus. [laughing]

Interviewer: *[54:12 – 54:39]* All right. 50 Questions plus got it. And are there any other methods or instruments or also maybe literature sources that you could use except or in addition to

request information from the manufacturer to make adoption decisions and assess considerations?

Expert: *[54:40 – 56:55]* Yeah. So you can look online to see if they have published anything, but there's, there's definitely one, one way to go. It requires time and it requires also expertise, right? Because not, not everyone would be able to to find it and understand it. There are some things on the way, like the register of high risk AI systems that, that the AI Act requires. So we just don't know what information is going to come out there and how useful and how detailed the information is gonna be. Uhm, but that may be in the future a good source of information. Uhm, you have now the CE marked products. Uhm, let me think. I think it's like there is no like consensus right now regarding quality of AI systems or, or a scale or something like that. So there's, there's not really much that you can, that you can find, like a standardised something. I'm not aware. I'm not aware of it. It does help a lot, for example when in general tools are listed or are recognised by a particular group, for example the cardiology society or the cancer network of blah blah. And so they, they do make efforts to, but that's more like the tools that are consumer-faced. So it's not, I don't know if they do the same thing for, for other type of, of technologies. So that's, that could also be like a, a more sector-specific, uhm, sector-specific evidence or sector-specific endorsement or something like that. Yeah.

Interviewer: *[56:56 – 57:31]* And is there any other method aside from looking into available evidence that is published, uhm, to assess considerations? For example, if we look at the different categories like, uhm, cost-effectiveness or value, is there anything? So, or, uhm, phrased differently, would you rely as a healthcare provider to 100% on published information or information from the manufacturer? Or do you have any other methods to make evaluations internally?

Expert: *[57:32 – 59:58]* Yeah, so there are, uhm, for example, what I mentioned about the about the cost-effectiveness and about the implementation results and the usability results. The fact that there's evidence there doesn't mean that it's gonna be the same for you. So, uh, so I do think that there's some things that you need to test yourself because they may not be applicable to your, to your case. But as what evidence is already available, uh, yeah, that may be possible that, that your institution has already a particular model for determining the, the value of something, or the, the chance of reimbursement, for example, which is very important. Uhm, yeah, but these instruments, I, I do think that are quite institutionally managed. So I'm not really sure what, what would we be definitely looking into what from the institution. Uhm, what is necessary for the institution? Uhm, is important. But as for general information, you need to generate your own, your own information. And again, this depends on the, on the risk

of the system. The riskier the system, the more you need to invest in that. Uhm, but that, that goes with how the, for example, how the hospitals normally work. You introduce a new, a new, a new part of your of your health path, and then you need to evaluate. So I don't think that is rare to have, to do something like that. Uhm, yeah, yeah, but I don't know, I don't, for example, maybe the things like notified bodies for example and third bodies are only responsible to communicate to the to the manufacturer. So you cannot ask them anything. And for example, they hold a lot of information about it. Uhm, and the rest of it, if it's not publicly available, then it's just in the hands of the, of the manufacturer. And so I, I don't know.

Interviewer: *[59:59 – 01:00:23]* Right. So the next the next question is you mentioned the 50 plus questions before, so if you could maybe, uh, elaborate more on how these questions would look like. So were these questions relating to, uhm, information request, requests from the manufacturer? Or yeah, maybe you can tell me more about the type of questions.

Expert: *[01:00:24 – 01:00:39]* Uhm. Yes, there, there are things that the need to be quantitative, some things that need to be qualitative, I think. So, for example measurements about, uhm, about performance, I think should be, uhm, well, if it's applicable, should be quantitative. So you should know what is your accuracy. You should know what is your, uhm, what are your false positives and false negatives. Uhm, yeah. So you could, you could think about, for example implementation information is going to be a mix about the two. So there are measurements for usability, for acceptability, et cetera, but also, uhm, there are things about workflow analysis, for example, which is usually very hard. Usually won't quantify those. You would have a mixture between those. Uhm, regarding the value and, and reimbursement methods, uhm, you know that, I don't think we can put everything on manufacturers. I think that this this checklist as well should be, you should be able to determine what is, what you have to do as, as, as buyer of the system. Because you cannot, you cannot, I cannot expect a manufacturer to do all these things for you. Sorry, I need water.

Interviewer: *[1:01:57 . 01:01:57]* Sure.

The interviewee went to get a glass of water [01:01:58 – 01:03:51]

Interviewer: *[01:03:52 – 01:04:31]* Yeah, I, I have a follow-up question to your response. So just to make sure I understood you correctly, you mainly request information from the manufacturer, but there is also some, there are also some considerations that are for you, your responsibility as a healthcare provider. So which, for which you wouldn't request information from the manufacturer, but I think just also to clarify, if I understood that correctly, I think post-

deployment monitoring is one of these considerations that would also be on you as a healthcare provider to consider. Uhm, yeah.

Expert: *[1:04:32 – 01:04:59]* Yeah. So at least some aspects of it. But for example, the AI act for high risk AI systems, for example, required post-monitoring. So, uhm, yeah, so in some sense there is some post-deployment monitoring that you would do as the one deploying the system. Uh, but there's also an obligation, also legal, for the manufacturerrs, to do this as well.

Interviewer: *[01:05:00 – 01:05:16]* Right. So yeah, just to finally clarify: the checklist would mostly be, or would be a mix of mostly considerations you would request information for from the manufacturer and then also some aspects that you would need to consider yourself.

Expert: *[01:05:17 – 01:06:11]* Yeah, how I see this. Maybe that based on this information provided by the manufacturer, then you would determine what are the actions that you need to follow. For example, you say, okay, this, this, uhm, this, this AI system is interesting for me. However, if I want to deploy it in my context, I need to do this, this, this and this, or I need to be aware of this, this and this. Or I need to follow up on this and this. Uhm, maybe your, your, your decision at the end is gonna be like I cannot use it yet because I need to get this organisational stuff sorted out before for example. But at least would it would flag what kind of things do you need to do in order to do a responsible adoption of the, of the AI system.

Interviewer: *[01:06:12 – 01:06:50]* Right. And one follow up question again to this. Uhm, so you've mentioned some considerations such as the acceptability or the performance of the system, which can be quantitatively measured. Uhm, how would you, as a healthcare provider, try to assess the compliance with specific requirements which are maybe, which are difficult, more difficult to quantify, or where you might ask: Okay, have you considered this and this? Where you would just rely on the manufacturer saying yes or no. How would you approach this?

Expert: *[01:06:51 – 01:08:04]* Yeah, I think this, I think that's, that's really hard, right? So if you don't have a quantitative measure, or if you have a quantitative measure but you don't have a gold standard, for example, then that that just makes things, things difficult. Uhm, I, I, I think it would depend if I were to, for example, if, if you say, uhm: “Oh, is this, is this design universal, for example? And the manufacturer says: “Yes, of course.” Then I have two options, to say yes in my checklist or to ask him: “Oh, what kind of evidence do you have for, for this?” And I think that step is important to do, because, because the answers of the manufacturers can be overpositive or can be positive in, in the, in the perspective that they have. Maybe not, yeah,

maybe not, the maybe not the, uhm, maybe you would make different considerations. Or for you there are things that are more important than, than others.

Interviewer: *[01:08:05 – 01:08:10]* How would you proceed if the manufacturer would say: ”No, I don't have any evidence for universal design.”

Expert: *[01:08:11 – 01:08:39]* Yeah, then you cannot say this. So if you if you just think that you're design universal but you have no, done nothing to actually test it or do it in that way, then you just have to say that, that you don't have universal design. That you maybe have a very good design, but you don't have universal design.

Interviewer: *[01:08:30 – 01:09:00]* Right. Okay, got it. Okay, so these were also all my questions. Uhm, have I missed any topic or consideration that you think is relevant? So now you can think broader. Maybe not specifically regarding, maybe not specifically regarding particular considerations, but all kinds of questions or topics that I missed with regards to the adoption of the AI-based clinical decision support systems.

Expert: *[01:09:02 – 01:10:10]* Yeah, I was thinking, uhm, like the, like the question of how far would you like to go to, uhm, for example, if, if you’re, okay, wait. If you think that a checklist would help in the adoption of of an AI system in clinical practice, then probably you need to think that this checklist would be part of the procurement process. So I think that checking with people that, uhm, that are a little bit closer to how an AI system gets adopted into, for example, a hospital, uhm, would be important to try to make the checklist, the goal of the checklist to align with the use of the checklist. Because for example, if this is part of the procurement process then you can go all the way and ask all the evidence and etcetera. But if what you want to do is more at the clinical, at the level of the, of the doctors and how happy are they with a particular system or what is their interest, for example, just to kind of move before the procurement process then you're, you're, you're tool would be different. So I think that is interesting to see at the different levels that such a, such a tool could work. And maybe define very well, or maybe at the end have a recommendation of also how to use this tool in which context. So in, in which, in which part of the process of the adoption of the AI system would be important.

Interviewer: *[01:10:38 – 01:11:29]* Mm hmm. Yeah, the way I currently think of the tool I wanna develop is, is basically providing healthcare providers, so as I said, this could be healthcare professionals, but this can also be the people overseeing the deployment of an AI with the questions they should consider, and usually they would, or as you also said, they would mainly

request information from the vendor, but it would basically be a tool which provides questions or the questions they should ask, uhm, to ensure that the AI-based clinical decision support system is suitable for their clinical practice, uhm, before they would actually choose to adopt and buy such a system or rent.

Expert: *[01:11:30 – 01:13:42]* Yeah, because if it's like that, I do think that you're looking at a procurement, being part of the procurement process. But that is a very, uhm, protocolised process. Uhm, so to, to align with, uhm, to align with what, what are the, I don't work in that. [laughing] Uhm, but, yeah, getting to know how, how that that process takes place may give you some relevant information about how this realistically could be applied. Cause doctors are, doctors individually, at least in Holland, are usually not the ones to decide in getting some kind of, of adopting a particular technology or not. It's usually at the level of the institution of the GP practice, for example. Or the hospital, uhm, or if it's included in some kind of guideline, which then the thing would come from top-down and say: “You know, for diabetes, this type of diabetes patients, for example, you would use this, this type of pump or you can, you can use this type of pump and then you can use it.” So the same thing you could think about a clinical decision support system that there would be a guide and this says: “For this and this, then this type of systems works. Or you could apply this type of systems.” Uhm, similarly to what happened, for example with the, with the risk assessment scores, for example. Uhm, yeah. So it's it's just interesting to, to sit, to sit down, to think like from a, if this is, if this is meant to define, buying or not buying a particular technology, it's interesting to think like, how is that process of buying or not buying a technology surrounded by others, by other stuff, by other… Because maybe they're also redundant things are really being established by the procurement process.

Interviewer: *[01:13:43 – 01:15:02]* Yeah, I would. Yeah, I would say it's, I wouldn't see it like as the sole basis of buying or not buying, but as a help to making such decisions. Uhm, yeah, but then again, I suppose it depends on, uhm, on how the specific needs and, uhm, considerations that the healthcare provider gives to an AI system because, yeah, depending on, because, one healthcare provider might want that the healthcare, that the AI system considers all different or is completely trustworthy, if you can say so, and that would and only then he would adopt it, and another might say, well, it's, it performs well, it's cost effective, it's mostly trustworthy, but it has this issue or these kind of issues, but it would still adopt it. So, that's, that's also a bit difficult because it's, I guess, maybe, uhm, not the, I'm looking for the right word. I would say that, uh, the bar for, for making adoption decision might differ, differ between healthcare providers. Uhm, but yeah.

Expert: *[01:15:03 – 01:16:15]* Yeah. For example, in Holland this is defined by how, by the, by the insurance companies. So, they're interested in savings. So if you come and you say this clinical decision support system would save you so much money, they would probably just reimburse it. Uhm, but that's what I mean. Like the, the decisions being made. For example, if you're more interested in, in an implementation checklist than this, I see that as as what you're talking about. So how the clinicians see the application of this system in their clinical practice? Uhm, what are they compliance? Or what, what, what are the requirements that they see necessary for them to be able to do their job? Uhm, so then you're looking at not something that defines the, the binary buying, but something that defines the necessary requirements for the implementation process. This can be taken into account by once the, the, the system has been, or when the system is being bought, or or, yeah, making the decision, but it's not directly linked to to that. But then I would refer to it more as an implementation checklist and.

Interviewer: *[01:16:17 – 01:17:12]* Yeah, yeah. I think the checklist that I had in mind was quite a in between or at the intersection between both because the how clinicians think of the AI system was, is one part of it, because I imagined that I have these pre-identified categories and trustworthiness is, is one aspect. Especially the impact on working skills is I would say heavily on the implementation side, but then on the other hand, buying not buying decisions… so ideally I would have a checklist that would bring together all considerations both especially on, on the deployment side, but which would simultaneously help to make buying decisions. Uhm, I wouldn't say it's, I, I wasn't thinking of the checklist specifically as buy or not buy. So, yeah.

Expert: *[01:17:13 – 01:17:41]* Yeah. So I think that what I'm trying to tell is that I don't know enough of the process of deciding buying or no buying. It is because that happens at an institutional level. So there are things that work, the things that we need to, we know that they work, but they're not implemented in clinical practice and that decision is made high up and I am not, I don't have enough information to, uhm, to tell you how the process takes place as to include a checklist in that decision making process.

Interviewer: *[01:17:42 – 01:18:15]* But if you, if I would say before adopting, isn't, uhm, because that's how I was thinking that this was broad enough to include both the buying process and the implementation process. Because if you think about considerations prior to deployment or adoption, you would consider implementation questions in advance. So that's, that's why I would say I had this overlap between this procurement checklist and an

implementation checklist because both of them would, from my perspective, fall under adoption. Yeah.

Expert: *[01:18:16 – 01:18:35]* Yeah. No, I definitely hear you. And I, I think I think you're right. I just, I just don't know how, if, if your goal is to generate a tool that would help you in making a decision buy or not buy, I think it's important to take into account who will be using it and the decision is not taken by the doctor.

Interviewer: *[01:18:37 – 01:19:00]* Right. And yeah, now I'm thinking if I, if I could still define a checklist where I would say it can be used both for buy or not buy decisions and for implementation questions because that was my goal to bring together all considerations of AI adoption. Uhm, so ideally it would be applicable to both processes.

Expert: *[01:19:01 – 01:20:25]* Yeah, like what, what we were talking before that if you, if you make the checklist and the things that you know you need to do yourself, so that they, that they will help you in the implementation process. That's perfectly valid, you know. And that doesn't necessarily have to be done all by the, by the manufacturer, but it would help you to, you know, to buy the system is not only is the system good or not, is also is the system implementable in my context and it may not be implementable. Not because it's not good, but because I cannot do it. Because I cannot, I don't have the right infrastructure, maybe because it's not interoperable. Uhm, oh, that's an important point. Like the technical aspects of that is the [not understandable] construct the, the technical part and is it interoperable and does it work with the systems that you have currently in place? Uhm, seah. Yeah, so that that can definitely help you. So there, there are two things like is the system good enough? For what, for what the claims are. And the second one, can I implement it? And if you ask those questions, it can help you know if, if, if you can buy it or if you can buy it and implement it, or if you're gonna have a lot of hurdles implementation hurdles you still don't. Yeah, but you still need to work on.

Interviewer: *[01:20:27 – 01:20:41]* Yeah. So I think it would help for me to basically define adoption as one umbrella term for buying or not buying. And on the other side implementation: "How well does it work?"

Expert: *[01:20:42 – 01:21:21]* Yep. Yeah, I just, I just don't know very, I'm not just, yeah, I just don't know about it, about how these decisions are being are being made, who is involved, what kind of evidence 'cause they do have this this thing of like: Is it compliant? Is it CE marked? For example, that's a no brainer. Does it comply with the GDPR? That's also a no

brainer. So they already have these kind of things in place for buying technology. Uhm, I think yeah, from, from the medical devices part, there is a lot of this already established.

Interviewer: *[01:21:20 – 01:21:31]* Yeah. Thanks so much for all the insights. And yeah, I would gladly let you know more about, uhm, this thesis, while it's ongoing.

Expert: *[01:21:32 – 01:21:37]* Yeah. No, definitely. Let me know where does the checklist ends up and then yeah.

Interviewer: *[01:21:38 – 01:21:38]* Sure.

Expert: *[01:21:41 – 01:22:06]* There is some people doing like Delphi studies of this with similar, similar subjects. And then what they do is, they have, after the information that you would have, yeah, they would generate a list of, of items and then check with a lot of people what is important and should be in it and should not be in it. So, so for follow-up ideas.

Interviewer: *[01:22:07 – 01:22:15]* Yeah, sure. Because eight weeks for the thesis isn't that long. So I think there is quite quite a lot of room for following up on this work.

Expert: *[01:22:16 – 01:22:18]* Okay. So shall I end the recording?

Interviewer: *[01:22:19 – 01:22:22]* Great. Sure. Yeah. Thank you.

Interview 2 – 15 April 2025– via Microsoft Teams

Expert 2: Postdoctoral Researcher studying the use of health data, including AI and its impacts

Interviewer: *[00:01 – 00:16]* Uhm yeah. If you were to define different categories for all the questions or considerations that healthcare providers need to think about and evaluate before adopting an AI-based clinical decision support system, what would those categories be?

Interviewee: *[00:18 – 1:39]* Uhm, so the categories would be technical feasibility. Uhm, which is basically everything to do with the hardware that the a CDSS would be implemented in. Uhm, things like your electronic health record etcetera. Uhm, there would be social acceptability, which is essentially: Do patients and clinicians *want* [word emphasised by the interviewee] the use of a CDSS in patient care, uhm and also how do patients feel about their data being used by private AI companies, etcetera. Uhm, you have ethical justifiability, which

is essentially ‘do no harm’. Uhm, so what harmful consequences might arise. And those harmful consequences need to be sort of in the whole chain, everything from the individual level all the way up to this system level. Uhm, and then your last one is, is legal compliance. So there's sometimes confusion, I think that AI sits outside, uhm, a regulatory space, but it doesn't. You have to think through data regulation, medical device regulation, uhm safety regulation, discrimination regulation, lots and lots of different types of risk.

Interviewer: *[01:40 – 02:01]* Right, thanks. So to summarise, uhm, you mentioned technical feasibility, social acceptability, ethical justifiability, and legal compliance, [interviewee saying ‘mhm’ and nodding] and you've also mentioned already some considerations within these categories. And could you dive deeper into each category and perhaps elaborate which specific considerations should be taken into account before adopting an AI?

Interviewee: *[02:02 – 7:35]* Sure. Uhm, so if we do technical feasibility, there are a few different things. The, the most practical and the most basic is just: Do you have the tech in order to run the algorithm? Uhm, so if you are, uhm, in a hospital and you're running all of your computers on Windows 97 and you don't even have an electronic house record etcetera, that's not gonna work. Uhm, you're sort of next level up from that is then: Is it the same tech where the algorithm was trained? Because that can also cause different problems. Like, uhm there are often things, for example with image recognition software where the algorithm was trained using a particular type of MRI machine and then you try and deploy it in a hospital that has a different type of MRI machine and then it won't work. The same logic also applies with, uhm AI-CDSS. So anything that might impact the performance, you need to be aware of. That can be anything from like, uhm some electronic house records record more fields than others, uhm, but it also might be your care pathway is different. And so you therefore want to know, uhm, when does something fire [not understandable] and all of this type of stuff, which then leads to your my last consideration that would fall within technical feasibility, which is about usability. Uhm, so you don't just want something to work, uhm, you also want it to work well. Uhm, if you are requiring clinicians to log into four different systems and they've got to use a smart card and then they have to do 2 factor authentication, that is just never gonna to happen. Uhm, and so it's also got to be usable. So those are your sort of main considerations in in technical feasibility. Uhm, in terms of social acceptability, this tends to be: Is it fulfilling a social purpose? Like is it fulfilling a need that people actually want it to fill? Uhm, is it safe? Uhm, and is it effective? Uhm, and so, you know, if you go to your doctor, you probably don't really care so much if they are using a particular type of blood pressure monitoring cuff because you know that that's you know, it's safe, you know that it's useful for your care. Uhm, if you went to the doctor and they said: “Oh, I'm going to predict what illness you might get in the future.” There's probably a bit

more of a social barrier to wanting to know that that works. And there's also the same for doctors. So doctors, remember, have both an ethical duty they have, they take the Hippocratic Oath, uhm, but they also have a legal duty, which is, which is called when they are, it's called their fiduciary duty, basically to act as learned intermediaries. They have a responsibility to care to their patients, uhm and then that therefore means that they are *rightfully* [word emphasised by the interviewee] reticent to use technologies that they don't know are safe and effective. So therefore anything that impacts safety and efficacy falls within this category. So that's things like: How did you verify the thing? How was it evaluated? Is there evidence that it actually works? Has it been subject to some form of trial? All of those types of things. Uhm, and it's also important about is it fulfilling an actual need rather than a want? So clinicians and patients will quite willingly accept the use of new technologies if it's clearly solving a problem or serving a purpose. If it is just the case of somebody in a managerial position saying we really want to buy this software and you better use it, that's not going to really work. So you want to also, ideally you would like the idea to come from the clinicians. Then in terms of ethical justifiability, that's things like: Is it going to change the relationship between patients and clinicians and is that going to have a negative impact? Uhm, are there issues with things like bias and could that therefore cause discrimination? Uhm, what about stigmatisation? Because remember stigmatisation is a big thing, is a big thing in healthcare. Uhm, there is: Does it have an infringement on patients autonomy or their privacy? So it's all of those types of considerations. And then, uh, legal compliance is really everything I said before. It's, it's: Does it comply with medical device? Does it comply with data protection? Does it comply with consumer protection law and does it comply with discrimination law? Uhm, and within those things, you've obviously got multiple different, multiple different categories. Uhm, but sometimes, like I said, there can be a little bit of sort of AI mythology or exceptionalism and people don't think that they have to subject AI to all of those tests. Uhm, if you were to, uh, basically any form of legal test that you would do to an another medical device, you should probably also do to, uhm, AI-CDSS. Uhm, and so essentially all of these considerations are: Why was the thing bought? Whose idea was it to, to design it? Why is it needed? Where is it being implemented? How does it work? How do we know it's safe and effective? Do people want it used in their care? Those are sort of your main blocks of considerations, uhm, which largely fall into the things that fall under, uhm, health technology assessment.

Interviewer: *[07:36 – 07:56]* Got it. Uhm, and regarding the technical feasibility, so you've mentioned the integration into the technical infrastructure and usability. Uhm, does this just for me, ah to note, does this also extend to the technical robustness of the AI system or is this more on the integration into the clinical practice?

Interviewee: *[07:57 – 08:11]* So integration into clinical practise, uhm the things like the technical robustness of of an algorithm, I would put more in terms of legal compliance because that tends to fall within medical device law.

Interviewer: *[08:12 – 08:39]* Right. And would this also, so would legal compliance also cover some other. So from my perspective, there's an overlap between trustworthy AI and legal compliance. I would, I would say trustworthy AI goes beyond legal compliance. But, uhm, you mentioned technical robustness. Does this also apply for privacy and, uhm, can you? Yeah.

Interviewee: *[08:40 – 11:21]* So there is, there is to an extent a low overlap between trustworthiness and legality, but they are they are different. If you, if you've got something that's legally compliant, you actually don't need trust. Trust is only necessary when you exist in a space where there's no guarantees. The law should be providing you with, with guarantees. So I shouldn't need to trust that my AI system is safe and effective in protecting my, uhm privacy. I should just know that that is happening. What I might need to trust it for is: Is it overly mediating the relationship between my, between me and my clinician? Is it disrupting things like the empathetic caring nature of that relationship? And what impact might that have on my, on my care? Uhm, so anything that you can sort of guarantee legally, I would say falls within the category of legal compliance. Anything that falls outside of that scope is going to be in one of those other three other three categories. So a really good one would be evidence, right. So medical device law currently, is fairly limited. It doesn't apply to a lot of AI tools and it tends to de-risk them. So it tends to class them as like risk level 1 or risk level 2 rather than like 3 for example. As a consequence, that tends to mean that the actual evidence barrier is quite low. So the amount of evidence that is required to generate, that is required to prove that the thing works is often minimal and it can just come from the software developer. Uhm, trustworthiness would require from a clinician's perspective normally a higher standard of evidence. Uhm, so normally clinicians would want to know that it's been used by somebody else, that it has all of these types of things. Uhm, so that would fall into social acceptability. So you do get these kind of crossovers. Uhm, but yeah. The other thing that would fall into in terms of legal would be liability. Uhm, at the moment there is no clarity with regards to liability law. So it's not really clear if an if a CDSS went wrong and misdiagnosed a patient. It's, it's entirely unclear whether that would become the fault of the doctor or the fault of the software provider. So at the moment you are requiring clinicians to trust in the use because they are basically trusting that it's getting it right and they're not gonna get into trouble. Once liability law is in place, you will not need to have trustworthiness incorporate that category.

Interviewer: *[11:22 – 11:42]* Yeah, that breaks it down pretty understandably for me. Uhm, and

so you mentioned the liability. Uhm, re there any other issues or considerations that are pretty much at the beginning and pretty much unsolved and but which you still would have to think about if you would create a checklist for AI adoption?

Interviewee: *[11:43 – 13:58]* Yeah, there are. There are a lot. Especially when it comes to generative AI. So there are many unanswered questions with regards to just machine learning. Uhm, liability is one of them, but there is also things like change management. So all of the ways in which we would currently expect, if I was doing a checklist at the beginning of like I would want to know: Has it, has it been verified, has it been evaluated? Was it in a clinical trial? Uhm, that tends to be a one-off exercise for things like drugs. Uhm, for machine learning, it can't be a one off because the machine learning might update or the population might change. You might get things like population drift or data set drift. All of these things are ongoing considerations and the law and sort of practically doesn't know how to deal with that. Uhm, the same is also true with evidence generation. So you know what I was just saying about evidence. Uhm, one of the reasons why there is fairly poor quality evidence that these things actually work is because nobody really knows how to test them. The gold standard of evidence-based medicine has for a long time been the randomised controlled trial. Randomised controlled trials or RCT's are not really appropriate for, uhm, a CDSS. They're expensive, they take a long time to run, they're not really adaptive. Also, what do you compare it to? Do you do AI versus a different AI? Do you do AI versus human? Do you do AI versus human plus AI? Like just we don't, we just don't know. Uhm, and so those are all unanswered questions already. And then generative AI has come in and completely blown everything out of the water because there are all sorts of new, new considerations. Uhm, so I would say for generative AI, there are many, many unanswered questions, uhm many more unanswered questions, which would still fall in the same category. How do you test that it works? How do you know that it's doing what it says it does? How do you make sure that it's safe? Who do you hold accountable if things go wrong? How do you use it and where do you implement it into the care pathway?

Interviewer: *[13:59 – 14:37]* Yeah, great. And if you, uhm, if you think about the adoption of AI in clinical practice, I'd say, uhm, questions like: Have you, do you have evidence for the AI system that it works in clinical practice? …is pretty straightforward. But when we think about these unanswered questions, it's a difficult question, I suppose, but how would you approach taking them into account, uhm before adopting an AI? Because for liability, you mentioned just as a clinician, you would have to trust the AI system that it works. Uhm, how would this look like for other considerations which are pretty much unanswered right now?

Interviewee: *[14:38 – 16:13]* So a lot of this falls within what I would call procedural regularity.

So basically I think it is okay if there are unanswered questions, uhm as long as you follow a process that is the same every time. So I think part of the problem that we're seeing at the moment with AI is that first of all, AI is not treated in the same way as any other type of medical device or software that might be bought into the hospital. Uhm, and so whilst the hospital or you know, any other health body might know to ask quite awkward questions of somebody who's trying to sell them a new drug, they don't necessarily know to do that for AI. Because these companies come in and they sort of bamboozle them and give them all of these hopes on a silver platter. And they're like: Oh, we're gonna solve all your problems. Uhm, so that's problem number one. The second problem is that because there is no standardised procedure, they just do it differently every single time. Uhm, so the benefits of having something like a checklist, even if there is not agreement on how you generate evidence, at least you know to always ask: Uhm, what evidence is there available to, to, to show that it works? What support mechanisms are there in place if something goes wrong? How are you monitoring the ongoing performance? Those are all things that should have an answer, even if they're not standardised, if that makes sense.

Interviewer: *[16:14 – 16:27]* Right. Yeah. So basically just to summarise, you would have to standardised procedure and the checklist could serve as this procedure to consider it as even though there aren't answers to this, you would at least consider it every time.

Interviewee: *[16:28 – 17:21]* Yeah. And it's not even that there aren't answers, there would still be answers to what evidence is there available. What there won't be an answer to is: Does this classify as gold standard evidence? Because nobody has yet reached an agreement. But that at least you've asked and that is something that, uhm can be decided locally. Same as if, okay, we don't know liability, but there should at least be a process in place for monitoring performance and checking when things go wrong. Uhm, in much the same way, you know, you can think of it like, you know, aeroplanes have that red box thing that records everything that's going on. That's the same logic. It's like, okay, we don't necessarily know what's going to happen if this plane was to crash, but we have a recording of everything that happened and we know that there's a process and a procedure for dealing with it if it goes wrong.

Interviewer: *[17:22 – 18:10]* Right, yeah, that makes sense. And going back to these 4 categories that you mentioned, uhm, and again, this question between trustworthiness and legal compliance. So from my understanding, there is some considerations or issues, uhm, such as discrimination. For example, you mentioned there is discrimination law or transparency might be consideration. Uhm, but in, in the cases where the law doesn't cover all considerations or issues that should be addressed, uhm, and there are some specific, uhm, discrimination or

transparency considerations that go beyond the law that ideally should be considered. Would you then categorise them in a different, uhm, group or how would you approach this?

Interviewee: *[18:13 – 18:15]* So stuff that is above the law.

Interviewer: *[18:16 – 18:19]* Right above the law, but within a category that is covered by the law.

Interviewee: *[18:20 – 19:16]* So, so I, I, I would still say that, that that falls into either social acceptability or ethical justifiability. So the discrimination, for example, would, uhm would fall into ethical justifiability where, where you're saying, if you imagine it, legal compliance, this is what I sometimes say, legal compliance is the floor, not the ceiling. So that in in any ways, maybe it helps if you actually do legal compliance first. So your first question is: Does this actually meet the law? If it doesn't, you almost don't need to consider any of the other three categories. Uhm, so if you were, uh, to sort of order them, it would be legal compliance. That's your minimum test. Then it's technical feasibility, which is okay, it complies with the law and I can, I can practically run it. Then it's social acceptability. And then it's ethical justifiability.

Interviewer: *[19:17 – 19:29]* OK, great. This answers this question because it was motivated by seeing a lot of critics of specific laws that they are sometimes not sometimes they don't cover it enough considerations. And…

Interviewee: *[19:30 – 19:57]* No, they don't. Then the laws are very thin and they're not very broad. So they, they, they don't encompass most of AI. So for example, generative AI falls outside of almost all of the remit. And even when they do consider AI, they tend to consider only this part, not, not the whole. So that's why you need all of those other categories, but the you just have to at least check that it complies with the law before you'd think about anything else.

Interviewer: *[19:58 – 20:10]* Thanks. And now, uhm, when you think about the checklist, how would you approach developing and using such checklist and practice for CDSS-AI adoption considerations?

Interviewee: *[20:11 – 22:04]* Well, I would do what you're doing. Find out, you know, what are the, what are the considerations from different, from different perspectives and then. So the, the main thing that you want, would always want with something like a checklist is to know what is the purpose of asking for that item. So if it's to assess whether or not it works or if it is

it to assess whether or not clinicians want it. So you need to know the purpose of the thing. You then need to know how it is going to be answered and is it possible to be answered. There's no point having an item on a checklist that no one can provide an answer to, uhm or consideration of. And you also want to know who is going to be responsible for answering that part, uhm, on the checklist. So I would think through, think through all of those things. Uhm, I would look at existing checklists and see what's there, uhm, and then see what's missing or see what you can adapt, uhm, and build off that. And then I would look for things in the literature, uhm, and what, what jumps out as those sort of most common things. And with a checklist, you're always balancing. Uhm, you want it to be useful and you also want it to be detailed. So you try and you want to try and ask questions in that checklist that encompass a lot of things, uhm, without having to have like 200 different items on this checklist. Because nobody is gonna use a checklist that's that long. But something that's like 20 items where each of those things represents many things inside them, then people might, might use it.

Interviewer: *[22:05 – 22:31]* Great. Yeah, great that you also mentioned the item amount, because that's the question that I've been thinking about a lot. This trade-off between practicality and granularity. Uhm, and if you would use such checklist as an as a healthcare provider or someone trying to adopt an AI system for clinical practice, what would be your main sources of information to ensure that these different requirements have been considered and met?

Interviewee: *[22:32 – 23:55]* Uhm, so I'd want documentation from the developer. Uhm, I'd wanna to speak to the developer. Uhm, and then some things would require documentation from your actual hospital. Uhm, and, uh, things like acceptability testing, that would be talking to your users. Uhm, so talking to clinicians who are [not understandable] gonna have. Uhm, and I would probably recommend as well that, uhm… so these decisions if they're just if a, if a checklist is being used, it's ultimately being used to make a decision whether or not to adopt the thing into your hospital. So I would want the process to be being led by the whoever is responsible for procuring the software. They're gonna go to the developer for the most specific types of questions about the actual algorithm. They're going to go to their in-house, something like a clinic, a chief clinical information officer for things like data protection. They'll probably go to the IT team to understand who is the technology. But then I would want the decision to not only be made by the procuring team. I would want the decision to be made as a part of it, almost like a multidisciplinary team. So you'd want the procurers, I'd also want the clinicians and I'd want a patient representative group involved in that decision making.

Interviewer: *[23:56 – 24:26]* Right. And you've already mentioned we would go to and for specific questions. So social acceptability is something that you would consider internally. Uhm, then specific questions about the algorithm are questions that you would ask the developer. Uhm, and if could you elaborate more on the… Or if you think about all the other considerations that you've mentioned before, uhm, could you dive deeper into each different consideration and tell me who you would go to, to ensure that it has been addressed. So if we think about discrimination or transparency.

Interviewee: *[24:27 – 25:39]* So those, those would all be things that I would want to come from the developer, right? So if I was thinking about discrimination, there are things I would want to know upfront and things I would want to know post-implementation, right? But we're only focusing on the upfront bit right now because we're only thinking about adoption. Uhm, so I would want to go to the developer and I would want to know: Do you have a record of what data you were you trained this on? Do you know how representative that data was? How does it compare? How does the population of that training data set compare to the population of my hospital? Uhm, have you test tested its performance for different types of populations? Do you know if it, if it, uhm, you know, causes any kind of biases. Those, those types of questions. Same with transparency. A lot of transparency is also gonna come from the developer. Like: Is your code in the open? Is your data, is your data in the open? Are there are there logged decisions about why it was decided to use this model instead of that model? So all of that type of stuff is, is going to come from the developing team. Yeah.

Interviewer: *[25:40 – 26:08]* Right. And if you, uhm, so we discussed a lot of considerations also specific to AI-CDSS, uhm, and how would you say does AICDSS differ from AI in healthcare in general? So what are considerations that you would especially have to consider for clinical decision support systems that are powered by AI?

Interviewee: [26:09 – 28:01] So the most of that falls within risk. So something like AI scribe isn't really having any direct impact on patient care, whereas CDSS, a CDSS is having a direct impact on patient care. It's determining what somebody might be diagnosed with, what type of, uhm, treatment they might be given, all of these types of things. So because it's higher risk, therefore the expectations are higher. Uhm, so whilst the considerations might be the same, the, uhm, burden of proof is basically higher. So my, my questions of something like an AI scribe: Does it work, how is it tested, is it going to work in my hospital? Are basically the same, but the answers that I want are different. Uhm, I think where you might want extra considerations with something like, uhm, a CDSS is in things like interpretability. So I think sometimes we think about, uh, AI-CDSS as just being a standalone thing, but it's course it's

not, it's actually a partnership. So what you really want to know with, with AI-CDSS is: How does it change the decision-making of the clinician? And so there are many considerations like how usable is it? How interpretable is it by the clinician? How does it impact them? When does it fire? Like at what point in the clinical cognitive process does it actually, does it actually alert the clinician to the fact that there's a decision to be made? All of those considerations, uhm, are sort of unique to clinical decision support.

Interviewer: *[28:02 – 28:34]* Right, and you've mentioned that the considerations are mostly generally the same. And now if I think about my thesis, so I'll probably, within these eight weeks, maybe as a follow up work, uhm, will not be able to, uhm, break down for each specific considerations, what should be the burden of evidence. So how, how else could I reflect the difference between, uhm AI-CDSS and AI in healthcare? Because yeah, as you mentioned, the main difference is mostly the burden of evidence or the expectations. And how could I reflect that in the checklist?

Interviewee: *[28:35 – 29:27]* So I think you can reflect the things that I've just said, the stuff about interpretability and usability, etcetera. Uhm, in terms of, uh, things like: How do you reflect the burden of proof? Uhm, well, that's in the wording of your question. So if it is, uhm, you know, if I was asking something like of an AI scribe, I would just say: Have you got evidence of a usability test? For example. If I was asking something like about clinical decision support, I would want to know: Do you have evidence of the impact on clinical outcomes? So it's the same high level consideration, but it's a different question that reflects the different level of risk.

Interviewer: *[29:28 – 29:47]* Yeah, that makes it so much more understandable. Thank you. And do you have any other, uhm, yeah. So I already asked all my questions. Is there anything you have to share about the topic of AI adoption or AI-CDSS adoption that I haven't covered with my questions?

Interviewer: *[29:48 – 30:13]* Nope. I mean, you've read my stuff. My, my PhD thesis exists on the web and answers this question. So, you know, there are many, many, many things, but I think this was a, was a good start. And I would just say don't get overwhelmed. Like I think with, remember that you are writing an undergrad thesis in eight weeks, uhm, and don't try and boil the ocean because you will lose your mind. Try and keep it specific.

Interviewer: *[30:14 – 30:19]* I think that's yeah, that's an important piece of advice that I should take more to heart. Thank you so much for taking the time.

Interviewee: *[30:20 – 30:22]* No problem. Have a great rest of your day.

Interviewee: *[30:23]* Thanks, you too.

Interview 3 – 2 May 2025 – via Google Meet

Expert 3: Medical Director of Research at a health technology company, Associate Professor of Medicine, and Physician

Expert: *[00:00 – 00:08]* So, I think, uh, we started the recording and the transcription

Interviewer: *[00:09 – 01:18]* Perfect. So, thanks so much for taking the time. Uhm, yeah, this interview is regarding my thesis and it aims to identify questions or considerations that healthcare providers need to think about and evaluate before adopting an AI-based, uhm, clinical decision support system. And these considerations I plan to synthesise into one checklist that healthcare providers can use to make informed adoption decisions. And when I say, when I say healthcare providers, so it's either the individual healthcare professional delivering health care, such as clinicians, or uhm those involved in adopting an AI system at the organisational health facility level, so management personnel or business personnel. Uhm, so yeah in addition to doing literature search for AI evaluation resources, I plan to do these interviews and yeah it's my third one today, uhm, and it will take around 50 minutes maybe a bit less but yeah the main goal is to hear your perspective on which considerations need to be evaluated before adopting an AI-based CDSS. And yeah do you have any questions before we start?

Expert: *[01:19 – 01:23]* No, that sounds great. Thank you so much, Sergej, for your work in this space.

Interviewer: *[01:24 – 01:37]* Great. Thanks. Uhm, so if you were to define different categories for all the questions or considerations that healthcare providers need to think about and evaluate before adopting an AI-based CDSS, what would those categories be?

Expert: [01:39 – 03:27] Okay. So, uhm, I don't have all this at my fingertips, Sergej, I'm sorry. [expert laughing; interviewer saying: “No worries”] Actually I should. But uhm, clearly, uhm, there's been a lot of work done in this space as you know and, uh, the question is what, ultimately we're interested in the idea of having trustworthy AI and clinical decision support tools for clinicians that healthcare, that both patients can trust the output, that clinicians can

trust the output, and that health systems are gonna feel confident that they should deploy, uhm within a healthcare system. So I think part of the question is how do we unpack the concept of trustworthy AI and what are the, the key categories within that that we are interested in. Uhm, clearly safety is, uh, a priority, right? That the AI system is not going to, uhm, do anything that is, uhm, going to harm the patient. Uh, so is, and, and that's intertwined with the quality of what the AI is doing, right? Uhm there it in, in some ways it also depends Sergej, I think in terms of what kind of AI you're talking about. Are you talking about, uhm, you know when you talk, I think you mentioned like a clinical decision support. If you're just talking about an AI that is deployed on patients' data that is clinician-facing, I think that's a little bit different than something that is using generative AI that is patient-facing. Uhm, so I don't know when you ask, I guess it would be helpful to clarify the, the question a little bit first.

Interviewer: *[03:28 – 03:31]* So I'm focused on, uhm, clinical decision support. So yeah.

Expert: *[03:32 – 10:01]* Okay. So, in some ways I would say the clinical decision support area is simpler, uhm, and, uhm, I think easier to get a handle on. Safety is one. And safety can be, uhm, understood, it's, it's partially, safety is partially accurate. It, it's intertwined with the quality of the output and the accuracy and, and precision of the output of the AI system of the clinical decision support and how does that compare to, uh, ground, what, what we consider to be ground truth. Uh, so and, and it's kind of an interesting question of should the ground truth be the standard of care even if the standard of care is not so good, uhm, or should it be, you know, like 100, you know, are we striving for 100% accuracy in terms of being able to diagnose, uhm, something? Uhm, so I think safety and quality is, is a key consideration. Uhm, I think the other thing, uhm, another category that's important here is the representativeness of the data and the output. You know, do we have some kind of verification and validation that the AI system or the clinical decision support actually works in the population that we're testing it? Not just that it kind of has worked and you know we have some data suggesting that it, it is, uhm, reliable, uh, and has a high quality output but actually in how representative is that testing of the real-world and the patient population that we are actually going to be deploying the clinical decision support in. So, uhm, that's, you know, the kind of the, it's a question about the generalisability, I guess, of the, uhm, of the system to the population that you're trying to inform. Uhm, I think the other one has to the other category here that's important here is transparency in terms of understanding, and transparency kind of occurs at multiple levels. Not only in terms of like how did the AI system get to the decision that it got to, uhm, which may or may not be always available, uhm, but being able to demonstrate to a healthcare system and to clinicians, uhm, the data that and the evidence that the AI system, that the clinical decisions support does actually produce, uhm, a decision that they can rely on and that is clinically meaningful and

valuable. Uhm, so transparency on that end, transparency about, uhm, the use of the system in a clinical setting like clinicians should know that this is happening. Uhm, they, there should be, uh, some sense of what happens if they disagree with the output and how do they handle that. Uhm, uh, so you know there there's multiple layers I think of transparency that are important here. Uhm, I'm sure there's others that are, are not, uhm, immediately coming to my mind. Uh, you know um those are I think the main ones, you know wanting to see the data to prove that something is effective is, is critical that it… I, I think also there's a lot of nuances with the data, right? When we think about a clinical decision support system, it, it can have high sensitivity or specificity. Uhm, and there are trade-offs between those two. And there are risks to systems that you know are overly conservative and there's risks to systems, uhm, depending, I think that the, the goals of the AI system need to be clear. So that you can set it, the responses in a ways that, in a way that makes sense for the clinical outcome that you're talking about…, uhm, and the, the risk associated with the clinical outcome, uh, for patients, right? If there's a very high risk of getting something wrong, you want a, you want an AI system that is gonna be more conservative in the sense of less likely to, uhm, you know have more, uhm, uh, false negatives, right? So I think that the, there has to be some kind of alignment between the, the, the goals of the AI system, uhm, and, uh, and how it, how effective it is or it's, you know the, the outcome. Uhm, a couple other categories that I think are important here which I alluded to a little bit that has to do with the data is just assessments for bias and fairness. That goes to the representativeness of the data. But that's kind of like a category that falls into that category of bias, uhm, and fairness, uh, and making sure that the, uhm, system works equally well, uhm, not only obviously in the population that you're testing but in all the kind of subsegments of the population, the clinical population that would be, uhm, involved. So, uhm, different sociodemographic area, uh, categories of patients, different racial categories and ethnic categories to understand that it is actually, uhm, equally, uh, effective in those different populations is important. Uhm, I think the other thing that you know just to, to mention obviously is whenever you're talking about a, a technology system in a healthcare environment, uhm, making sure that there's strong privacy and security. Uhm, controls is the other category that there's, uhm, good you know good, uhm, cyber security, it's not, it doesn't have threats of that, uhm, is important. Yeah.

Interviewer: *[10:03 – 10:28]* So you've mentioned trustworthy AI as a category and within trustworthy AI, just to summarise there are safety, quality of the output, representativeness or generalizability, transparency, bias and fairness, and privacy and security. And if you now take a step back and look beyond trustworthy AI, are there any other categories that should be considered, uhm, when adopting an AI?.

Expert: [10:31 – 12:00] Uhm, so I mean I think, uh, I think kind of putting on a healthcare system or a healthcare leadership hat. Uhm you know, uh, I think healthcare systems also need to be thinking about the environmental impact, uhm, of these systems. Especially as we think more about generative AI and large language models, uhm, and the, the, the environmental impact of, uh, the compute associated with doing that, using those kinds of models at scale. So that's another kind of factor. There's always, uhm, you know this isn't I, I mean there, there's always going to be trade-offs in terms of cost and return on, on investment and I think part of the question here is, uh, you know how do we think about that, that issue holistically in terms of the benefits to patients in terms of potentially improving outcomes, the benefits to, uh, health care providers in terms of maybe decreasing cognitive load and helping to deliver high-quality care more efficiently, uhm, uh, even when there's may maybe a high cost to the health care system of both developing evaluating and deploying an AI system. And so coming up with some, uh, approach to, uh, think through those trade-offs I think is also important.

Interviewer: *[12:02 – 12:10]* And if you think about usability and workflow integration, uhm, what considerations would you consider relevant within these categories?

Expert: [12:11 – 13:27] So I think that's essential. Uhm, uh, there, that is something that is frequently I think overlooked by people who are developing the AI systems. Uhm, and that is I think the most complex part of achieving our goals with AI in the healthcare space. Uhm, so and that's going to have to happen in a way that is unique to each healthcare system and each clinical environment. Ideally integrating something into the workflow into an electronic health record so that healthcare providers don't have to go outside of the record to use the AI, uhm, patient, uh, decision support tool, uh, is the best. But even doing that, sort of figuring out how do you mobilise and, and just inspire people to adopt it, to use it, to understand the value of it, to not be afraid of it, to not think that it's going to, you know, have a negative impact on their future or, uhm, the care that they deliver to their patients, uh, is, is, is a very important dimension of our thinking about, uh, AI and healthcare.

Interviewer: *[13:29 – 13:57]* And you mentioned the uniqueness of the specific use cases or healthcare facilities. Uhm, but if you would think about some aspects or considerations of usability and workflow integration that have to be considered in any case and you would think about them in a way, maybe in a checklist way. Uhm, what considerations would you list as relevant that have to be considered before adopting an AI?

Expert: *[13:59 – 16:05]* Well, I mean there has to be some kind of first of all like some you know has the health care system done some kind of change management process to, to explain what the system is all about like have, you, you know what, what have you done to, uh, communicate to clinicians what you're doing, why you're doing it, what the value of it is. Uhm, uh, that's kind of the starting point. I think the next step is to, uhm, make sure that the technology is easily integrated into the existing workflows. Uhm, to clarify, well who is supposed to be using it, how do they use it? Uhm, who's accountable I if something goes wrong? That's another piece of this that clinicians are going to want to know about. Uhm, what do they do if they have a problem or they disagree? Uhm, what do they do, what are they supposed to do with the output of it? How are they supposed to explain this to patients? Do you, does the healthcare system mandate that there's, uhm, clarity about the you know transparency about the use of an AI system to patients or with patients? Uhm, and there has to be some kind of ongoing iterative process of, uhm, evaluating the adoption and getting feedback. There should be like a continuous feedback loop I think, , uhm, in terms of the implementation cycle. So that you could continuously improve it. Certainly having a checklist to start with of how you wanna approach this with the hope that that's going to help, uhm, with, uh, the implementation is a great idea. And, uhm, it's kind of a starting point because you never know the uniqueness of the environment that you're working in and the more feedback you get, the more you're gonna improve the process. So I think that there's this balance between kind of thinking, okay, this is like where we're starting and then having an openness to, uhm, continuous feedback and iterating on the process to improve it, is, is helpful.

Interviewer: *[16:07 – 16:22]* Right. And you've mentioned quite a few workflow integration considerations. And if you think about the usability of the, uhm, clinical decision support system,, uhm, what are specific user experience or usability considerations that you could think of?

Expert: *[16:25 – 17:53]* Uhm, I mean in terms of the user experience the integration with their own electronic health record is essential. Uhm, I think also how the, you know, how the information is shared with the clinician and then being able to have some kind of translation of that information for patients is also important. I, I think if it's, if it's something that ultimately is meant to be delivered to patients, creating something like translating that in a way that patients can understand, uhm, to help. It kind of depends on what the system is trying to do. But if you're trying to like influence treatment decisions, if it's something diagnostic and that is supposed to influence what, how a patient decide, what treatment a patient decides to get, well having something that is put out in a language and a way that patients can understand, I think would be helpful too. So I think kind of thinking about the entire, uhm, life cycle, uh, from a

usability perspective is, is really, uh, is, is very important. It's not just about, uhm, it's not just about developing the technology and giving it to the clinician. It's about how do you distil that information for the clinician to make sense of it. How do you then take that information and distil it in a meaningful way for the patient so that it can actually influence clinical care? [interruption: interviewer started to introduce a new question but the expert was not finished answering]

Interviewer: *[17:55 – 17:59]* Got it. And um at the beginning you…sorry.

Expert: *[17:58 – 19:52]* I, I also I also think it's critical that these systems have something built in for feedback by clinicians, right? And, and, and so that if there's a problem with the system, clinicians know who to go to. If the clinician disagrees with the system, what do they do? And how does the system respond? Uhm, just we know that there's been some studies where clinicians are so confident in their own perspectives that they ignore the AI output even when the AI is right. And, and that's a serious problem because they, they ultimately don't trust the output of the AI. Uhm, and so the question is how do you overcome that? And I think part of the way you overcome that is by ensuring and, and this is part of the user experience, I think, that the clinician, helping the clinician to understand how the AI got to its, uhm, its output and why the clinician should actually believe it. Uhm, it, it actually reminds me, uhm, Sergej, I don't know if you've used any of the more advanced, uh, large language model tools that are out there now. So, uhm, for example, uh, Gemini, uhm, 2.5 flash thinking. Like there's, there's, there are models now that show you the, the thinking of generative AI large language model systems. You, it show, they they'll show you the resources that they've, they're reviewing. Uhm, they'll show you what are, you know how they're arriving at the key concepts and that kind of process I think will go, showing the rationale, uhm, of a clinical decision support system will go a long way towards helping, uh, clinicians ultimately adopt the, the AI.

Interviewer: [19:54 – 20:20] Right. And at the beginning of the interview, you differentiated between clinician and patient facing, uh, technology or AI. And if we think about CDSS, which from my perspective is usually clinician-facing. Uhm, what considerations, uhm, can you think of, uhm, regarding the clinician patient relationship when using CDSS?

Expert: [20:19 – 22:25] Well, uh, so I mean I, I think again it, it depends exactly it depends on what your, what the clinical decision support is doing. Uh, I, I think, I, I think coming up with some language for how clinicians can explain the use of AI in the particular context to patients in a way that is going to reinforce their relationship and not, uhm, not make the clinician feel, you know, that, uh, that the patient may not trust their own decision, the clinician's decision

making or thinking or expertise in a way that doesn't undermine the clinician's expertise, uh, is I think one of the challenges. Like some clinicians may feel that patients, uhm, you know may not have the same respect maybe for the, the clinician if they're use, if they're if they have to rely on the AI. Uhm, and I think the, I think that giving clinicians the, the language to explain that this is a new tool, it is, uhm, going to you know be for their benefit. Uhm, it, it should you know help, help the clinician focus on you know the most you know some of the, the, the most challenging patient outcome areas. Uhm, and that there's still the human touch, uhm, that the clinician is going to, you know, maintain within the context of the relationship. But this is just another tool, not so different than, you know, maybe an MRI when we first had MRIs or CAT scans, right, to help, you know, diagnose cancer. Uhm, it's just maybe, it's an advancement on that. So you I could see that there, uhm, clinicians may feel threatened in some way by it. But I think that's all in how you explain it.

Interviewer: *[22:27 – 22:43]* Okay. And before we move on to the checklist design, uhm, have, is there any other category of AI-based CDSS adoption that came to mind perhaps, uhm, as compared to the beginning of the interview? Any category you've missed?

Expert: *[22:45 – 22:50]* Uhm, category in terms of like the trustworthy, uh, things? Or types…

Interviewer: *[22:51 – 22:55]* In terms of the adoption considerations as a whole. So going beyond trustworthy AI.

Expert: *[22:57 – 23:41]* Uhm, no, I mean I think that, I think that the things that we we've talked about in terms of the user experience and the integration into workflow are a critical part of that of you know whether or not this is ultimately something that can be adopted. So it's not just like safe, you know, it's not just safety and the quality and the generalisability or fairness and privacy, security. Like those are all important things for healthcare systems to be thinking about but healthcare systems also need to be thinking about how do we deal with the patients? What's the consent process? Is there consent for it? Uhm is, uhm, you know how do we, how are we gonna deploy this? How are we gonna implement it? Right? It's much more complex.

Interviewer: *[23:43 – 24:07]* And if you compare from what I've heard clinical decision support system is still broad but if you compare it to AI in healthcare in general, what are, how, what are some specific considerations to clinical decision support systems or what, yeah, what, what should be considered for this technology in particular as compared to in and healthcare in general.

Expert: *[24:11 – 25:28]* I mean you know when you talk about AI in healthcare in general, right, there are many non-patient-facing, non-patient areas where AI can be integrated that don't have the same kind of risk or potential for harm. So I think that that's the, you know the, one of the key differentiators here. Uhm, I think the other is that, uhm, some of the AI in healthcare is not necessarily even interfacing with clinicians at all. It can be on the administrative side, right? It can be on the research side, right? There are many ways in which we can use AI in healthcare that have nothing to do directly with patients. Uhm, we can use it for administrative tasks. We can you know use it for summarising notes, right? We can, right, there's lots of ways in which we can use AI that are not about making clinical decisions that in, that are potentially high risk for patients in terms of clinical outcomes. Uh, so I, I think that, uhm, you know the categories that we talked about earlier are, are pretty differentiating or essential in terms of thinking about safety and quality and bias and things like that.

Interviewer: *[25:30 – 25:47]* Right. And if you now think about the checklist with AI adoption considerations that a healthcare provider would use before actually procuring and implementing an AI-based clinical decision support system. Uhm, how would you approach both developing and using such checklist in practice?

Expert: [25:51 – 27:30] So I mean I think, uhm, I think the checklist has to be at the health system level. Uh, clinicians are not gonna use this checklist in my opinion. Like some who are particularly interested in this area are gonna be curious about it and think more about it. But for the most part, I think it is at the healthcare system level to make sure that they're doing the right thing and that they're going to be most successful in terms of the implementation of the AI system in their environment. Uh, so I mean I think that is kind of kind of where the, where it, that's the group that you kind of need to be thinking about as you're developing this. Uhm, and recognising that those groups or that kind of segment of healthcare is going to be concerned about: Is this something that is going to add clinical value and is it worth the return on investment. Right? There's two sides to it and, uh, healthcare systems have considerations about quality and outcomes for the patients and efficiency of care, uhm, but they also have concerns about the financing of healthcare and so figuring out where there's a win-win I think is, is part of it and, and so a checklist in that environment has to be, speak to the, the concerns of that healthcare system in, along those dimensions as well as be simple and easy to use and not overly complex.

Interviewer: *[27:32 – 27:37]* And when you say simple and easy to use. What are some attributes that you have? And yeah.

Expert: *[27:39 – 28:48]* Uhm, so I, I mean I think first of all it just can't be too long and cumbersome is, is the first thing. You know, you can't have a checklist that has 30 questions on it even I think is probably too much. Uh, so the, the more simple it is the better. Uhm, something that is, uh, electronic and has some kind of, uh, interface electronically that gives you an output is in my opinion ideal that helps people kind of take it to the next step with the decision-making. Uhm, so unless it's like, you know so that you kind of understand what you can do with the output of the checklist, right. It's not just: Okay, I have this or I don't have this. And then okay do I have to have the things that I don't have. Is this supposed to just give me some guidance for what I need to do or am I creating a checklist that's also going to help me decide between different interventions that I might want to use? It kind of depends also on what, what's the point of the checklist. Is it to evaluate a particular intervention or is it to help like pre-deployment with whether or not you wanna, you should be deploying the intervention or not? And what do you need to have in place before deployment?

Interviewer: *[28:49 – 29:25]* So it's more for pre-deployment but I was thinking quite broadly about adoption as to cover both the procurement decision whether to adopt it or not but also implementation aspects, such as usability and workflow integration. Because I think procurement and the implement, and the actual implementation after you decide to procure it, uhm, overlap quite a bit, uhm, so that's why I said adoption to cover both this whole adopt or not question but also what should be considered when, when I decided to adopt it.

Expert: *[29:27 – 29:43]* Yeah, I think that those are uh very important. Uhm, and I think something that kind of, uh, integrates or synthesises the information in the checklist to some kind of output for decision: Should I procure this or should I not procure it? Uhm, would likely be helpful.

Interviewer: *[29:45 – 29:58]* Right. And you talked about a question number. And if I would just ask you out of the blue, uhm, if you could name me a range of questions, a range of question number, uhm, what would be an appropriate number?

Expert: *[29:59 - 30:08]* So you know, that's probably something that is worth exploring a little bit
more. Uhm, but I would think between 10 and 20.

Interviewer: *[30:09 – 30:11]* Right. And…

Expert: *[30:11 – 30:59]* But part of it is also like how long it takes to answer them. And you know, sometimes people have 10 or 20 questions, but then they really have multiple questions embedded in those other questions. And it, you know, so I think it's also like, the reality is you'd be surprised that healthcare systems, uh, as well as developers of AI, there's just, they're pulled in many. many different directions and the amount of time that they have to really think about this kind of stuff is shockingly, uhm, small and, and you know just not, not nearly as much time devoted to it as you might think. And there's an eagerness to just, you know, make decisions quickly. And so there, there isn't, uhm, as much thought as perhaps there should be.

Interviewer: [31:00 – 31:33] Right. And you mentioned that the checklist should be used at the health system level for the most part. Uhm what do you think about having a checklist and then indicating, so also from my perspective it should be used um by the or on the organisation level for the most part. But what do you think about, uhm, indicating: “This question should be used by the management personnel or at the health system level but there are also some questions that should be used or answered by the healthcare professional.” So, what do you think about that?

Expert: *[31:34 – 32:12]* Uh, yeah, I mean I think certainly if the health care system doesn't, you know typically health care systems have health care professionals who are part of the leadership team assessing it, uhm, who can represent the health care professionals So I, I think you know healthcare systems need to bring in the expertise like of whatever clinical care providers are relevant, right? So if I'm talking about evaluating, uhm, a uh, an AI patient care decision support tool in oncology or in cardiology or radiology right, I should have the clinical expertise of that domain involved in the process.

Interviewer: [32:12 – 32:30] Right. And how would you approach answering these checklist questions if you, if you are the, yeah, basically responsible for implementing or for procuring an AI system and you have the checklist to use?

Expert: *[32:33 – 33:18]* Uh, so obviously it depends on the question that you're talking about. Uhm, uh, questions about the, the quality of the tool have to be kind of the whoever is developing the tool, uhm, has the burden of sharing the evidence to demonstrate that. So some of the questions are coming from developers and people that have studied the tool. Uhm, if a tool is developed internal to a health care system, they're gonna be people that have to sort of serve that role. Uhm, some of the questions have to do, uhm, with uh people internal to the health care system in terms of how you're going to integrate it into the workflow. Uh, so I think it kind of depends on the nature of the question in terms of who answers it.

Interviewer: ***[33:19 – 33:30]*** Right. Yeah. Thanks so much. Uhm, yeah. Do you have any other topics or categories to share? Uhm, or is there something that I've missed with my questions?
Expert: *[33:31 – 33:36]* So, uhm, Sergej, I, I mean, I don't know if we want to turn off the recording or you want to continue the recording.

The recording continued but the interview was finished at this point. The interviewer and the expert discussed the interviewer's future study plans and working together on (follow-up) research in the area of AI integration into healthcare settings.

Interview 4 – 7 May 2025 – via Microsoft Teams

Expert 4: CEO and Co-Founder of a company offering telemedicine solutions, formerly a senior physician in intensive care

Interviewer: *[00:01 – 00:20]* Uhm, ja, dann würde ich in die erste Frage reinstarten. Uhm, so if you were to define different categories for all the questions or considerations that healthcare providers need to think about and evaluate before adopting an AI-based clinical decision support system, what would those categories be?

Expert: *[00:21 – 00:24]* Is there any like, is there a special definition of categories?

Interviewee: *[00:26 – 00:44]* Uhm, no, you can define them yourself. So basically any considerations that you would have before adopting an AI-based CSS that you would put into a category. So I could give you one example, for example, trustworthy AI could be one category, but yeah, you could basically define them yourself.

Expert: *[00:45 – 01:32]* Yeah, I, I think, uh, the most important category is it has to create efficiency or a benefit for the user or the, uhm, use case. It it is, uh, it will be useful. This is I think the most important, uhm, category. And it had, the usage itself has to be as easy as possible, uhm, for the users. So if it's highly complex, if you don't understand what the algorithm is doing in general, what its endpoint is or what it is predicting or whatever it does, uhm, it will not work. I think this is ,those are the two, uhm, most important features an algorithm or AI has to have.

Interviewer: *[01:33 – 01:49]* And if you would break those categories down, what are some considerations within these categories? So usefulness and yeah, easy, uhm, ease of use. What

are some considerations or questions a healthcare provider need to think about before adopting an AI?

Expert: *[01:50 – 02:32]* Well, I mean, in, in addition to that, it's important to know what the AI is, uh, has been trained on. So what is what is, what is its capability? What can it do in in real life? Is it trained on, on data of for example, in the healthcare business where I am working, is it trained on healthcare data and is it really matching to the question or to the group or type of patients I have or not? I think, uhm, this is, uh, this is, uhm, prior number one to, uhm, to get, to clarify.

Interviewer: *[02:33 – 02:50]* Right. And you said ease of use is, does this go into the same direction as usability? And what are some traits or attributes that would make a clinical decision support system usable?

Expert: *[02:51 – 03:20]* Uhm, well, it's, it has to like I, I'm, I'm working with hospitals. So if you try to sell or try to develop an algorithm that is, uhm, creating benefits or efficiency in, in the hospital, it has to match or has to fit into the workflow. So, uhm, I, I fully agree that that if you want to, uhm, develop your workflows down the road, be modern, deliver high quality of treatment, uhm, you also have sometimes adapt your, uhm, your workflows. And if you have a, a new algorithm, a new software that is supposed to support you in terms of being a clinical decision support, it has to fit into your workflow. And I think this is, uhm, this is important if it, if it comes to, whether it will be used or not. So just, just having it does not mean you will use it and it's going to create, uhm, a benefit.

Interviewer: *[04:06 – 04:27]* Right. So the workflow integration you mentioned, is pretty interesting. And if I'll try to get even more beneath the surface, what are some considerations within workflow integration or maybe phrased differently, what should be done to be, to integrate a CDSS pretty well into the workflow so that it fits well?

Expert: *[04:28 – 06:40]* Well, I think it, it has to be a, if it's, if it's about clinical decision support, it has to, it has to address a, a field, an area or a question, a certain question that really appears during my daily workflow. So if it's a high sophisticated special question that is, that has a very low incidence in my clinical workflow, uhm, I'm, it, it might not be the first thing I think of if I, if I have those kind of questions on my desk. So it has to be or it has to address a clinical, a real clinical question that really occurs during my daily work. And, uhm, this is, this is important that that I identify a very, let's say a tricky question. Combination of different disease lead may lead to a sophisticated treatment and I want to know about a certain risk. I want to predict a certain

risk of my patient surviving this combination or this combined treatment for different or for a combination of different diseases at the same time. And I think, uhm, it is, it is highly important that, uhm, that I think that the AI will be a real, really helpful. If I think it's just a burden that I have to use it because I bought it or someone else did, uhm, it will not create a benefit for myself. It has to, it has to for me, I'd have to feel like: Okay, there is a tricky question that comes up quite, quite often during my daily workflow with different patients. And this tool, software AI will be helpful in addressing this question and supporting me in terms of a clinical decision support to find my final decision as a, as a physician. I think this is this is this has to, uhm, this is of importance. So let's, let's put it this way.

Interviewer: *[06:41 – 07:02]* Right. And if you think about the change in workflow, so standard care versus AI based clinical decision support system, what are some questions healthcare providers need to ask or need to ensure, uhm, or need to ask to ensure that these workflow changes, uhm, yeah, basically work well.

Pause, waiting for response: *[07:03 – 07:10]*

Expert: *[07:11 – 07:22]* Well, I think, I mean what I have to or what the industry has to ask or can you can you can you put it in a different light your question?

Interviewer: *[07:23 – 07:44]* Sure. What the healthcare provider needs to ask to ensure that by integrating AI-based CDSS, uhm, the workflow will still work well. So basically, uhm, maybe just to throw one example in. What are the changes in responsibilities by physicians and healthcare professionals or, yeah.

Expert: *[07:45 – 10:53]* Well, that's a that's a hard question actually. Of course. I mean, who is responsible of using the AI and then of interpreting the results? Uhm, who is liable? I mean, what about liability? If the computer tells me, I mean I may be biased. If the computer tells me mortality risk is 90% in a certain patient, I may treat them different than, than a patient that has a mortality risk of 60% because I feel like, well, 90 is pretty high. He will die whatever I do. And then it's fulfilling prophecy. The patient dies because I didn't put much effort into the treatment. So I think liability is important. Who's who's who's responsible of acting. But also learn what does it mean? I, I, I'm as a physician, I'm not very well trained in interpreting clinical system support that is AI-based. So what does it mean if it says: “The mortality risk is 90% in a certain patient?” What does this mean? What does the algorithm mean by that? How do I interpret, uhm, and only if I understand this, uhm, very well, I, I the algorithm itself will be of will be, will be of help definitely. So maybe this is something I have to clarify together with the, with a

physician for example, that is the end-user of the algorithm and, and has to has to work with it. And I also have to transfer the knowledge that the algorithm or the computer. If it's a clinical decision support, it does not make a diagnosis. It's not a diagnostic tool. It's a prediction tool that is a decision support. So finally I have to put everything together on information I can get. Information I read out of the lab, information the algorithm gives me, and then draw my own final decision on the case. So this has to be made very clear to the users, otherwise they may overestimate the support of a clinical decision support, which is totally, I mean, I, I, I don't, I, I don't need to tell you that, but it's, it's totally different to a, for example, analysing algorithm that is analysing CT scans of the brain and then giving me a diagnosis, whether there's a tumour or not, which puts an algorithm in a totally different light compared to an algorithm that is predicting a risk and then will be interpreted as a clinical decision support. Because I as a physician have the final decision whether there is a risk or not. I can even overrule the algorithm and say, well, you know what? It's, this is a false positive. There is no risk. I don't see any risk. And then, uhm, and, uhm, and write this down in the EHR. So I think this is important of understanding what the algorithm really does.

Interviewer: *[10:54 – 11:24]* Right. So to summarise, you mentioned the liability, then you mentioned training physicians to be able to understand and interpret algorithm output. And the third consideration that you mentioned was basically this part of this human computer interaction. So you should understand it's a support and how you work with that support. Uhm, and what are some other workflow change considerations that a healthcare provider needs to take into account besides these great considerations that you mentioned?

Expert: *[11:25 – 13:28]* Well, now we were very focused on the user side, uhm, but we also have to take under consideration that there's a patient like someone who, who is, uh, who is treated by, by a physician, by a healthcare system that is using AI-supported clinical decision support. So I have to make very clear, uhm, if you are a patient in my facility, we will use AI supporting us and finding the correct treatment for you or the correct diagnosis. Uhm, and, uh, I, I, I also think that as, as best or better I can do that as like all trust or higher trust my team and the patient will put into the system. And I think if there's mistrust, you shouldn't use it. You should only use those kind of supportive systems if there's a kind of a trust. And this creating trust is about explaining the, the idea behind it and what it does. And also the patient has to understand, well, it's not the computer that is, that is making a decision about to live or die because these decision makers are the physicians in the hospital. But they have so many information at the same time that they may need clinical decision support with an AI-based model that supports the physician to find the best decision for me. And, uhm, I think that's quite clear. So even it's not only about the users, it's also about the, let's say the patient or the patient

side, uhm, that has to be addressed if we think about change management or how to implement an AI algorithm.

Interviewer: *[13:29 – 13:53]* It's highly interesting that you mentioned patients as well. And I think you've mentioned a very important consideration, the transparency. So you should communicate to the patient what is the CDSS, why is it used… and what are some other questions that need to be asked and answered regarding the physician-patient-relationship when adopting an AI-based clinical decision support system?

Expert: *[13:54 – 17:29]* Well, uhm, many AI systems are cloud based. Uhm, you also have to take into account data security, data privacy, stuff like that. So all the regulatory aspects, uhm, and you know, if you want to make the picture very big, it's also an ethical question. My personal opinion is I as a patient have a ethical personal right to, uh, have AI running over my data because we have 20% misdiagnosing by physicians. And I want to reduce this failure of diagnostic, failure of correct treatment by having an additional computer system software checking on everything and make sure or supporting at least the decision maker with uh, with, uhm, information to come to the right or correct decision in my case. So I personally feel it's totally unethical if there is a algorithm that is well trained for a certain question, not to use it. So I would, I would really prefer having it up and running, getting more information about its performance, uh, compared to not doing it because we fear that the computer is taking over or, uh, we fear it's, it's, it's not ethical in using it. I think it's vice versa. We, we have so many data, millions of data of information coming in, in a hospital or in the healthcare system, uh, about one case. Uhm, and it's not possible for a human for a physician to put everything together. But having, uhm, AI system supporting you might, might lower your misdiagnosing rate significantly. This is what patients, this is what patients usually don't know. Nobody, no one talks about it. But this is what patients should be informed of. And then I mean, I use Google Maps and it has an AI and tells me make a left, don't go right because there's traffic congestion and I make a left. I trust the system. So we, and then we have AI in the like daily life in every, in almost every, every tool, every application. Or many applications. But if it comes to healthcare and my personal health, we, we feel that we cannot trust those systems, which is only about informing the patients what it is, why we need this. Because of healthcare, it's so complex and we do 20% of misdiagnosing what we would like to reduce. So it's all about explaining why those systems are so good and are so supportive. And yes, we may also have some systems that create more of errors and we'll also do wrong diagnostic as, as well as a physician will create in his, uh, working life. So it's kind of, we have to talk about it very open. So this is the more ethical aspect maybe. It's not a direct user aspect, it's more an ethical question.

Interviewer: *[17:30 – 18:16]* But it's a really interesting perspective that yeah, the patient, if there's the evidence for an AI in healthcare, that patient has the ethical right to, to be treated with or to have this AI in its treatment. Uhm, and now if we think about the start of this interview, you mentioned, uhm, the usefulness of the AI and the ease of use or yeah, the usage, uhm, or usability as two categories. And now that we've talked about some other considerations, how would you revise those categories? Uhm, would you add some adoption categories or? Sorry, I think the question was phrased a bit badly, but, uhm yeah.

Expert: *[18:19 – 18:25]* How would I? Can, can you say that again in like other words, maybe?

Interviewer: [18:26 – 19:25] Yeah, I'm sorry. Uhm, at the beginning you, I asked you about adoption categories. So if you think about all the considerations you need to take into account before adopting an AI-based clinical decision support system, how would you put them in different categories? And you mentioned the usefulness, uhm, if I'm not mistaken and you mentioned the usage or the usability, uhm, and then we talked about some other considerations like workflow integration, uhm, the physician patient relationship. And now my question is, how would you revise these categories? Would you add some, uhm, or maybe if you can just break it down in summary for me. For example, if we look at work for integration or physician patient relationship, is this part of one of these two categories that you mentioned at the beginning or are these separate adoption categories? Yeah. So maybe if you can just give me this breakdown of all adoption categories you think need to be taken into account.

Expert: [19:25 – 23:42] Well, let me let me make an example to make it more, more, uh, colourful kind of. We, we developed an algorithm that is a neural network and it's what it does, it is, it uses real time vital signs such as heart rate, blood pressure, saturation, uhm, breathing frequents and temperature in real time and is processing it and tells us a, predicts a risk of developing a sepsis, uh, within the next ten hours. At, at this moment where the patient does not show any clinical signs of sepsis. So the patient itself might look very stable to a physician, but the algorithm shows a high risk of deterioration and going in, going down the road into a systematic or systemic infection. Uhm, and putting the patient in, uhm, in a life threatening situation. So this is a, it's so important to have a good prediction model for this question of risk of sepsis because sepsis, 50% of ICU patients will develop a sepsis within the time on ICU. So it, it addresses a lot of patients and it's like a chameleon. So it's really hard to spot or, or find out about a patient that is of risk of sepsis. Usually we are behind the development. So you, you get the, you get the idea of there's a, there's a sepsis, if it's already too late. It is already up and running. The patient has already a very low blood pressure, very high heart

rate, uh, high breathing rate and low, uh, saturation and so on and so forth. Uhm, and, uh, you will you will be kind of running behind this life threatening development. So it, it addresses a very like a current important question. And this is a question where the physician wants the algorithm to screen all patients on his ward behind the curtain, kind of, and then tells him by flagging them like this is a patient of high risk I, I have to re reassess and, uhm, check maybe on the patient whether he is maybe, uhm, a correct, a positive patient and needs, needs to have a different treatment. And with that, we have to, uh, first tell the patient: Okay, if you're a patient in our hospital, we're going to use this algorithm. We have to train the physicians on what to expect from this kind of thing. So it's not going to tell you a diagnosis: “This patient has a sepsis.” It tells you he has a risk of developing it. So it's up to you preventing the patient from deterioration. Uhm, and it's just, it's just focusing your attention on that patient. This is everything it does. So the expectation has to be quite clear. And then, uhm, we start like rolling it out, uhm, and also tell them about, there will be definite, it does not have an accuracy of 100%. No medical test system has an accuracy of 100%. Even HIV test might be false positive in very, very few, uh, patients. So the algorithm telling you there's a risk might be wrong. You have to take this under consideration as well. And this is what I mean in about adopting a workflow, informing about everything. And, uhm, if the hospital is very critical, we also talk about ethical aspects, whether it's ethical to wait until you see he goes down the road or to maybe predict his risk and then take care of him in an earlier stage. So, uhm, I don't know whether I answered your question because I, I, I, it, it sound to me you wanted to have kind of a summary of, uhm, of the, the points we, we put together.

Interviewer: *[23:43 – 24:25]* Yeah, uhm, it did. But maybe, uhm, a follow up question. So my, my goal with my thesis is to develop a checklist for the healthcare provider to have with questions that they should ask, uhm, regarding the tool. And if you would think about this checklist and you would have different categories within this checklist such, uhm, as workflow, integration and usability, uhm, maybe if you, or you would have categories for different questions, for example four questions in one category. And my question would be what categories should be covered in a checklist regarding adoption of AI in a clinical setting?

Pause, waiting for response: *[24:26 – 24:30]*

Expert: *[24:31 – 24:33]* What category should be captured.

Interviewer: *[24:34 – 24:37]* Right. So what categories of considerations?

Expert: *[24:39 – 24:52]* Why don't you think? I mean, we mentioned already a couple of

categories. Why don't you think? I mean, shall we put out like take out one or two or why? Why aren't they all important? Uhm, maybe I don't get the question.

Interviewer: *[24:53 – 25:24]* No, sure they are all important. I think just at the beginning when I when I asked you this question, you mentioned usefulness and usability, and then you mentioned some other categories. And I think my question was more, uhm, if you could just re-summarise them. Uhm, because you could also say that, for example, workflow integration is a subcategory. So my question would be basically to have all yeah, basically for you to categorise them once again imagining you would put them in a checklist.

Expert: *[25:25 – 26:27]* I think most thing, most important is the usefulness like. And then subcategory to usefulness is whether it's going to integrate into the workflow, whether it addresses a very important question that really is a question like that you have every day. So if it's not the case, you will not use it. So usability is highly important and then, uhm, safety like data safety, data security, like how the system is built. Is it a bedside? Is it on premise? Is it running on premise? Is it cloud based? And then how it's working, what does it do? And how accurate? Like what, what about the accuracy? How good is it? It does not have to be, there will be no test 100%, but even 80% will be great or 70%. It depends on the question I have and the algorithm is supposed to predict. Even 70% can be great, but I have to keep in mind, okay, there's 30% mismatch.

Interviewer: *[26:28 – 26:41]* So you have usefulness and works integration slash usability as a subcategory. Then you have safety as a second category and accuracy. Uhm, do you think there's any category missing or are these the most important ones for you?

Expert: *[26:42 – 28:46]* Well, actually for the provider, healthcare provider, the most important category is, uh, pricing and reimbursement. Because, and this is the biggest issue in terms of why, uh, do we have just a few algorithms on the market up and running. It's not because we don't like it. It's because there's no, it's because, especially in Germany because there's no reimbursement. The insurance don't does not reimburse if I have a clinical decision support up and running because we don't pay for quality, we just pay for treatment. This may change in the future, but, uhm, it's that's different in the U.S.. For example, if you have acute kidney injury coming up during your hospital stay, the hospital will be punished and get less money because, uh, the law enforcement feels like, okay, this this acute kidney injury occurred and developed during the hospital stay and they couldn't prevent. So we're gonna punish the hospital for bad quality because this is a this is a setting the patient does not had, he doesn't, uhm, had when he was, uh, before he was admitted to the hospital. So this is, this is about paying or punishing

for bad or good quality. But that's not the case in Germany. So in Germany, uhm, we might get there in the future, but or we may get there, uhm, but so far it does not play a role and hospitals feel like they don't. That sounds weird, but hospitals did not fear they are, they are responsible for good medical quality. The first point they check is about reimbursement. If they get reimbursed, they do it. If not they, it's, it's tough to implement, let's put it this way. And that's one of the biggest reasons or main issues of, of the current market of clinical decision support or why we basically don't have it in Germany.

Interviewer: *[28:47 – 29:06]* Right. So, now you've mentioned usefulness, safety, accuracy and pricing, slash reimbursement. Is there any other category that of considerations that healthcare providers need to take into account before adopting an AI-based CDSS?

Expert: *[29:07 – 30:01]* Hmm. Probably not so. Looking at it from an industry perspective, uhm, I really don't like the regular regulation that we, that the algorithm cannot be self-learning continuously because this is by law not, not possible. So we have to, if you deploy the algorithm after certification, it has to stay with its knowledge that it has reached at this point. This is it. So it's not like in the industry you have an picture software that analyses pictures and identifies people or animals as long, as longer as it runs as better as it gets. This is not the case in the healthcare system so far due to regulatory things. But, uhm, I don't think this is something the healthcare provider will think about.

Interviewer: *[30:02 – 30:18]* And now my checklist for, uh, healthcare providers is still in its development. But if I would ask you, how would you approach developing but also using a checklist with adoption considerations for healthcare providers regarding AI-based CDSS.

Expert: *[30:19 – 30:21]* How would I use it or how?

Interviewer: *[30:22 – 30:31]* How would you approach developing a checklist? Uhm, so what should be in this checklist? How would you make it practical? But also how would you use this checklist in practise?

Expert: *[30:32 – 32:00]* Well, I, I think all points we mentioned should be on the checklist. So usability, safety, pricing. And then, uhm, I mean, that's a tough question. I also will put on the checklist whether there's any comparable product. So is there the possibility of compare one, two, three different products, uh, between each other, which will rarely the case if it comes to AI in healthcare outside of radiology. Radiology is kind of a, in a certain very high developed area and the and because of the analysing of pictures it, it might be a little different, uhm, and they are a little ahead of us of the rest of the medical field. That's for the clinical medicine. But,

uhm, I think everything we, we mentioned should be on this checklist. And I would definitely, uhm, go down and check mark, check the boxes and come to a conclusion whether we, uh, want to use it or not in our healthcare system and whether we think it's gonna benefit our workflow. We'll, we'll, uhm, like release workload from the physicians or nurses. That's one very important aspect in the hospital if you wanna be successful. But also, as mentioned initially, whether I think there's a medical benefit for the patient, this needs to be put first. If there's no benefit, because this question does never occur, there's no sense of implementing the, the algorithm, even if it's even if it feels very fancy to have it.

Interviewer: *[32:01 – 32:05]* And what is an appropriate, uh, range for a number of questions?

Expert: *[32:06 – 32:53]* Uhm, well, hm, good question. I think ten, around ten, twelve, max twenty. It should not be too long because it's not complex. It's more or less a basic baseline decision about as I said, it's you know, we're talking about clinical decision support. I decide finally and not the computer. So it's, I think from the liability aspect, it's, it's in the middle of a field compared to we do implement a diagnostic algorithm, uhm, that puts a diagnosis into the system. It's by itself. For this, we may have more questions or should have more questions in terms of safety, security and accuracy.

Interviewer: *[32:54 – 32:56]* And who should answer these questions?

Expert: *[32:58 – 33:09]* Uhm, yeah, while the user should answer most of the questions. And if it comes to data safety and security, it's the IT department. This is nothing a physician can answer. They have no idea about that.

Interviewer: *[33:10 – 33:15]* And when you say IT department, IT department within the healthcare facility or, uh, by the manufacturer?

Expert: *[33:16 – 33:26]* Uhm, usually the, the hospital, they will, uhm, they will run the show, they will check everything, uhm, and make sure that it's within their guidelines and borders.

Interviewer: *[33:27 – 33:39]* And what role would the manufacturer or vendor of the AI play in using the checklist? So would there be an interaction within the healthcare provider and the manufacturer or is it for you mostly the healthcare provider?

Expert: *[33:40 – 34:41]* Yes. If the algorithm comes like is, is, is running up in the cloud, there are certain touch points. So the hospital IT will send patient data, maybe pseudonymised or real data, uhm, from the hospital to the, for the cloud provider and back to, to get analysed.

And therefore there's overlapping topics between the hospital IT and the provider's IT, definitely. And the providers also, uhm, I think it's the provider's task or responsibility to, to really tell about accuracy, train the team on how to use it and so on and so forth. It's, it's like a device, like a, it's pretty similar to a medical device. And by law, the industry is, uh, forced to, they have to, we have to have, we have to document training and have to document that the staff is, uh, feel satisfied and good in using it.

Interviewer: *[34:42 – 35:05]* Great. And one last question, uhm, how would you say or how do the considerations for AI based CDSS differ from AI in healthcare in general? So what are some considerations specifically important for AI-based CDSS or maybe less important, uhm, in comparison to AI in healthcare in general?

Expert: *[35:06 – 38:34]* Well, I mean, we different, we, we do differentiate already, which is not the case usually if you're talking to like, let's say less informed people about AI. They put everything in one basket and think, well, AI is AI, which is definitely not the case. So there's a hype ongoing with ChatGPT and large language models. And of course, even like large language models can perform great and screening information of the digital documentation of patients in the hospital, for example for different risk factors and then select them and tell you something, or try to find out about the most likely diagnosis. However, large language models are, or can be dangerous in the healthcare sector system because if, if the information the algorithm is looking for has not been put down in the records, it will come up with a wrong decision or wrong information. So if we talk about AI, we have to differentiate what is the area of like the AI is supposed to work in. Am I looking for a diagnosis? So we're gonna compare millions of different CT scans of the brain and then the AI is telling me there's a very, very small tumour in the brain of the patient that I with my eyes cannot spot whatever I do. That's a totally different field compared to clinical decision support where I think the algorithms or algorithm will take different information together that I may not put together if I look at it because I don't see a like, a link between different features or variables but the algorithm does. Uhm, and, and I have to differentiate very clear in using it whether it's, it's supporting me and it's up to me to have a, to find a decision. I cannot rely on the computer that it's, it's 100% accurate if it tells me something. I have to make the decision. I think we have to make this very, very clear. What kind of algorithm is up and running and what is this capable of? Like what does it do and what is, is it not capable of and, uhm, I have to take care of? So I, I think this, this has to be differentiated and especially the different models and different types of AI, uhm, or, you know, I mean, sometimes we talk about algorithms and we do not mean AI. We think about a decision tree, which is not AI, it's just a decision tree. But it can be really, really good if it's a complex decision tree supporting me in, you know, filtering some information and then drawing a great

summary. Finally, it's pretty cool. It's pretty, it's, it's very supportive, but it's not AI. So we have to differentiate whether there's AI in place or not. This is important. And we also have to know for Germany, that's the law by now, that AI is not continuously learning. If it's deployed to the hospital, it stopped learning and it just can do them can do the task as good as it is at this moment.

Interviewer: *[38:36 – 38:56]* Right. Yeah, that's interesting. And sorry, one more question regarding the answer of the questions for the checklist. So you mentioned manufacturer would, so just for me to understand correctly, the manufacturer would provide information on accuracy, safety or how would this process of answering the checklist items look like?

Expert: *[38:58 – 38:59]* I don't think I, I got your question.

Interviewer: *[39:01 – 39:20]* Sorry. So just to go back to the question regarding who should answer the checklist items. And you mentioned the manufacturer. So maybe if you could just elaborate a bit more on how this process should look like the interaction between a healthcare provider and manufacturer in using the checklist.

Expert: *[39:22 – 40:31]* Well, I think the checklist with medical items such as whether there's a useful field or useful question I would like to the algorithm to, to running on, this is on the hospital side. And everything else has, is on the, is on the manufacturer's side. Like about the accuracy, the data it has been trained on, whether there's a study out, internationally published peer-reviewed study for the algorithm maybe underlining how good it is, or it, it's weak, weak points, maybe if there are weak spots also; this is on the industry side I think, definitely. And, uhm, this is it. I mean, it's, it's going to be a successful project if the, finally, if the physician or the nurse really think, uhm, is, is thinks that that it, it will create a less workload and a benefit for his or her patient. I think this is the, the most crucial point. If there's no benefit, there's no, no reason of implementing AI.

Interviewer: *[40:32 – 40:42]* Right. Thanks. And is there anything that I missed, any topic or any question that I might have not asked you which you think is important, uhm, regarding this topic?

Expert: *[40:44 – 41:41]* I don't think so. I mean, we even got into the ethical aspects, which I really like. I, I worked 25 years on intensive care unit as a physician. So ethical aspects are part of a daily routine on ICU. And I think, uhm, an ICU is, is, is a very good role model for AI clinical decision support because you, you get 5,000 to 10,000 single information each patient,

each 24 hours. So it's tonnes of information. And if the AI can support you in in identifying patterns, uhm, between your patients or for your individual patient, it's very, very helpful. So, uhm, no, I don't think, I and I think it this is this is a great example of where I can be implemented successful.

Interviewer: *[41:42 – 41:47]* Right. Yeah. Well, thanks so much for answering my questions.

Supplementary Material 6: Interview transcripts with codes

This material is provided as a separate document.

Supplementary Material 7: Themes and codes for the expert interviews

Note: Subthemes are *italicised*.

Themes	**Codes**
Structuring the landscape of AI-based CDSS adoption considerations	Legal and regulatory compliance (Interview 1)
	Ethics/trustworthiness (Interview 1)
	Do no harm (Interview 1)
	Value of the system/cost-effectiveness (Interview 1)
	Reimbursement (Interview 1)
	Implementation and workflow integration, including manufacturer's support and information (Interview 1)
	Evidence for the manufacturer‘s claims (Interview 1)
	Post-deployment monitoring (Interview 1)
	Legal and regulatory compliance (Interview 2)
	Technical feasibility (Interview 2)
	Social acceptability (Interview 2)
	Ethical justifiability/Do no harm (Interview 2)
	Trustworthy AI (Interview 3)
	Benefits of AI vs. costs (Interview 3)
	Workflow integration/implementation (Interview 3)
	Usability/user experience (Interview 3)
	Environmental impact (Interview 3)
	Benefit/efficiency improvement and fulfilment of a clinical need (Interview 4)
	Usability/Workflow integration (Interview 4)
	Safety (Interview 4)

	Accuracy (Interview 4)
	Reimbursement (Interview 4)
	Check for comparable products (Interview 4)
Breaking down legal and regulatory compliance	GPDR (Interview 1)
	EU AI Act (Interview 1)
	Local legislation (Interview 1)
	Regulatory compliance as a minimum requirement (Interview 1)
	Regulatory compliance as a minimum requirement (Interview 2)
	Medical device law (Interview 2)
	Data protection (Interview 2)
	Consumer protection law (Interview 2)
	Discrimination law (Interview 2)
	Data security and privacy (Interview 4)
	Legal compliance vs. trustworthiness (Interview 2)
Breaking down technical feasibility	Healthcare provider‘s technical infrastructure (Interview 2)
	Awareness of differences between healthcare provider‘s technical infrastructure and technology used for AI development (Interview 2)
	Usability (Interview 2)
Breaking down social acceptability	Fulfilment of a need (Interview 2)
	Safety (Interview 2)
	Effectiveness (Interview 2)
Breaking down ethical justifiability/ trustworthy AI	Doctor and patient trust (Interview 1)
	Bias and discrimination (Interview 1)
	Bias and discrimination (Interview 2)
	Patient autonomy and privacy (Interview 2)
	Impact on clinician-patient-relationship (Interview 2)
	Safety/do no harm and quality of the AI (Interview 3)

	Transparency (Interview 3)
	Alignment between AI system goals and effectiveness (Interview 3)
	Bias and Fairness (Interview 3)
	Privacy and cybersecurity (Interview 3)
	Generalisability/representativeness of AI validation (Interview 3)
	Clinician-patient relationship and clinicians‘ AI acceptance (Interview 3)
	Dealing with patients and their consent (Interview 3)
	Patient trust through transparency (Interview 4)
	Training data representative of implementation setting (Interview 4)
	Patient‘s ethical right to have AI in their treatment (Interview 4)
Breaking down usability/user experience	AI integration with an EHR (Interview 3)
	AI output understandable to clinicians (Interview 3)
	AI output understandable to patients (Interview 3)
	Feedback feature in the system (Interview 3)
	Explanation of how the AI generated an output (Interview 3)
Breaking down implementation and workflow integration, including manufacturer’s support and information	Usability (Interview 1)
	User training (Interview 1)
	Support channels (Interview 1)
	Vendor‘s contingency plans (Interview 1)
	Tailoring of implementation and training material (Interview 1)
	Mobilisation of people to adopt AI (Interview 3)
	Change management (Interview 3)
	Clarity about AI use (Interview 3)

	AI implementation evaluation and feedback (Interview 3)
	Responsibility for AI use and result interpretation (Interview 4)
	Training and Information for AI system users (Interview 4)
Breaking down post-deployment monitoring	Shared responsibility for post-deployment monitoring (Interview 1)
	Pre- and post-deployment measurement of outcomes (Interview 1)
Identifying differences between AI-CDSSs and AI in healthcare in general	CDSSs assist with decisions but don‘t make them (Interview 1)
	Doctor and patient autonomy and trust
	Considerations and requirements depend on the specific CDSS (Interview 1)
	AI-CDSS have higher risk and burden of proof than AI for administrative tasks (Interview 2)
	Reflecting higher burden of evidence in the checklist questions (Interview 2)
	Interpretability, usability, and impact on clinician decision-making (Interview 2)
	Risk difference between AI-CDSS and AI in healthcare in general (Interview 3)
	CDSSs assist with decisions but don‘t make them (Interview 4)
Discussing unsolved questions about AI in healthcare	Value determination (Interview 1)
	Liability (Interview 2)
	Management of changes and updates in AI systems (Interview 2)
	Evidence generation (Interview 2)
	Generative AI and new unsolved questions (Interview 2)
	Standardised procedure for handling unanswered questions (Interview 2)
	Danger of LLMs (Interview 4)

	Liability (Interview 4)
Making the checklist practical and deciding on a practical range for the number of items	Discussion about the type of checklist/guide this thesis aims to develop (Interview 1)
	Checklist development process (Interview 2)
	About 20 checklist items (Interview 2)
	Order of adoption consideration categories (Interview 2)
	Checklist should be simple (Interview 3)
	Ten to twenty checklist questions (Interview 3)
	Number of questions also depends on time required for answering them (Interview 3)
	Healthcare providers have little time in practice (Interview 3)
	Electronic checklist with output generation (Interview 3)
	Ten to twenty checklist questions (Interview 4)
Assigning the responsibility for answering the checklist questions	Healthcare providers should also answer adoption questions (Interview 1)
	Healthcare providers answer some questions (Interview 2)
	Questions about data protection to be answered by in-house clinical information officer (Interview 2)
	Healthcare provider's IT-team also answers technology-related questions (Interview 2)
	Questions about the quality of the AI to be answered by the developers (Interview 2)
	Checklist use at the healthcare system level, not by clinicians (Interview 3)
	Clinical domain expertise required in a healthcare system's AI leadership team (Interview 3)
	Questions about the quality of the AI to be answered by the developers (Interview 3)

	Workflow integration questions to be answered by the healthcare provider (Interview 3)
	Healthcare providers answer medical checklist items (Interview 4)
	Manufacturer answers non-medical checklist items (Interview 4)

Supplementary Material 8: Initial AI-CDSS adoption checklist

This checklist is meant for healthcare providers planning to adopt an AI-CDSS. It should serve both as a support for the decision of whether an AI-CDSS should be adopted and also for the implementation process.

1. What do you think about this checklist?
2. Are there any considerations missing in the checklist? If yes, which ones?
3. Do you think any question/consideration should be removed? For instance, because you think that another consideration that is currently missing is more important.
4. Which questions would you revise? For example, because you believe additional aspects of a consideration should be covered in the question. Or you think that the phrasing of a question could be improved (to make it more accurate/practical)?
5. Do you agree with who is responsible for answering the checklist questions?

Item	**Who should answer**
Has compliance with all applicable laws and regulations been ensured (incl. data protection, safety, medical devices, non-discrimination)?	Manufacturer, healthcare facility
How was evidence generated that the AI-CDSS improves the clinical outcome you are interested in?	Manufacturer
Were the system validation conditions (e.g., setting, sample) representative of your healthcare practice?	Manufacturer
Is the deployment and use of the AI-CDSS technically feasible (e.g., having the technical infrastructure, data interoperability)?	Healthcare facility, manufacturer
Is the deployment and use of the AI-CDSS economically feasible (i.e., reimbursement, awareness of costs, economic evaluations)?	Healthcare facility, manufacturer
Is the accuracy of the system high enough for its use case?	Manufacturer
Can the AI-CDSS integrate well into your existing workflows?	Healthcare facility
Do healthcare professionals find the AI-CDSS acceptable for use in their healthcare facility?	Healthcare facility
Do patients find the AI-CDSS acceptable?	Healthcare facility
How is it ensured that your staff will have the knowledge and skills to use (healthcare professionals) and manage (e.g., managers, IT team, legal team) the AI-CDSS?	Healthcare facility, manufacturer

Can it be explained how the AI-CDSS arrived at a specific output or recommendation, and do users understand these explanations?	Manufacturer
Have you considered which information about the system needs to be communicated to your staff and patients? Is the vendor's documentation for the system and evidence for claims sufficient to enable such transparent communication?	Manufacturer, healthcare facility
Is unfair bias that systematically and unjustifiably favours or discriminates against particular individuals or groups based on characteristics such as gender, age, and race avoided in the system?	Manufacturer
Have you considered and discussed with the manufacturer all other ways how the deployment and use of the AI-CDSS can harm your patients (e.g., privacy violations), staff (e.g., de-skilling), and the relationship between your healthcare professionals and patients? Have you considered measures to reduce these risks?	Manufacturer, healthcare facility
Are there plans or measures in place to respond to different adverse impacts of the AI-CDSS (e.g., technical failures, negative impacts on work)?	Manufacturer, healthcare facility
Have you considered the liability you carry for claims related to the AI-CDSS?	Healthcare facility
Is it clear how the vendor will support you during and after the deployment of the system?	Manufacturer
Do you have a change management plan in place to prepare for the adoption of the AI-CDSS?	Healthcare facility
Is there a clear plan for how your vendor and you will conduct post-deployment monitoring of the AI-CDSS?	Manufacturer, healthcare facility
Have you considered how users of the AI-CDSS and patients can provide feedback and raise concerns related to the system?	Healthcare facility, manufacturer

Supplementary Material 9: Search results per database/search

PubMed

Total results: 1,709

Full-text records screened: 221

- Duplicate: 18
- No access: 6
- Retracted article: 1
- Not applicable to healthcare delivery: 203
- Not specific to AI or CDSS: 3
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters: 1,465
- Practical procedural guidance preceding a more practical tool: 1

Scopus

Total results: 383

Full-text records screened: 71

- Duplicate: 33
- No access: 8
- Not in English or German: 1
- Not applicable to healthcare delivery: 149
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters: 190

Google Searches:

Economic aspects

Total results: 107

Full-text records screened: 41

- Duplicates: 22
- Not applicable to healthcare delivery: 37
- No access:
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters: 45
- Actionable procedural guidance preceding a more practical tool: 1
- Inclusion: 2

Usability

Total results: 92

Full-text records screened: 26

- Duplicates: 31
- Not applicable to healthcare delivery: 8
- No access: 2
- Not specific to AI: 3
- Published in a personal capacity: 1
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters: 44
- Not in English or German: 1
- Inclusion: 2

Usability (CDSS focus)

Total results: 124 results

Full-text records screened: 65

- Duplicates: 30
- Not applicable to healthcare delivery: 1
- No access: 2
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters: 87
- Inclusions: 4

UX

Total results: 66

Full-text records screened: 19

- Duplicates: 19
- Not applicable to healthcare delivery: 4
- No access: 2
- Not specific to AI or CDSS: 5
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters: 35
- Published in a personal capacity: 1

Workflow integration

Total results: 43

Full-text records screened: 14

- Duplicates: 10
- Not applicable to healthcare delivery: 3
- No access: 3
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters: 26

- Inclusions: 1

AI Readiness

Total results: 219

Full-text records screened: 88

- Duplicates: 36
- Not applicable to healthcare delivery: 59
- No access: 32
- Not specific to AI or CDSS: 1
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters: 71
- Published in a personal capacity: 6
- Inclusions: 14

AI vendor evaluation/procurement/cross-categorical tools

Total results: 115

Full-text records screened: 67

- Duplicates: 3 + 1 + 2 + 1 = 7
- Not applicable to healthcare delivery: 17
- No access: 21
- Published in a personal capacity: 2
- Not a practical procedural guidance or evaluation tool with specific considerations for AI adopters: 59
- Inclusions: 9

Supplementary Material 10: Literature sources included

Tools applicable to trustworthy AI evaluation in healthcare

Authors	Purpose	Source and Sector
Healthcare-specific Evaluation Tools		
Vollmer et al. (2020)	"help identify common pitfalls that can undermine ML/AI based applications in health"	Peer-reviewed literature, multisectoral collaboration
Reddy et al. (2021)	"assess the translational aspects and various phases of available AI systems"	Peer-reviewed literature, academic sector
Scott et al. (2021)	"assist clinicians in assessing algorithm readiness for routine care and identify situations where further refinement and evaluation is required prior to large-scale use", specifically ML algorithms	Peer-reviewed literature, academic sector
Al-Zaiti et al. (2022)	"guide to understanding and critically appraising [clinical] machine learning studies"	Peer-reviewed literature, academic sector
Wang et al. (2022)	"systematically appraise a [predictive] model's potential to introduce bias"	Peer-reviewed literature, academic sector
Crigger et al. (2022)	"evaluate whether an AI innovation meets these conditions: does it work, does it work for my patients, and does it improve health outcomes?"	Peer-reviewed literature, nonprofit sector
Szabo et al. (2022)	"promote discussion of AI trustworthiness between clinicians and technical experts"	Peer-reviewed literature, multisectoral collaboration
van Smeden et al. (2022)	"support medical professionals to distinguish the AI-based prediction models that can add value to patient care from the AI that does not"	Peer-reviewed literature, academic sector
Emani et al. (2023)	"critical appraisal of ML in neurosurgery"	Peer-reviewed literature, academic sector

Nazer et al. (2023)	"aid in mitigating bias during the development and implementation of AI algorithms"	Peer-reviewed literature, academic sector
Bacchi et al. (2024)	"proposed as a checklist of 10 questions that clinicians may employ as a starting point when faced with the question of 'Should this artificial intelli-gence algorithm be used in my practice now?'"	Non-peer-reviewed article, academic sector
Coalition for Health AI (CHAI, 2024)	"ensure that AI solutions and systems fulfill all five key, principle-based areas for trustworthy AI: 1. Usefulness, Usability, and Efficacy; 2. Fairness; 3. Safety; 4. Transparency and Intelligibility; 5. Privacy and Security"	Grey literature, nonprofit sector
Moons et al. (2025)	"to examine the quality, risk of bias, and applicability of any type of prediction model in the healthcare sector"	Peer-reviewed literature, academic sector
Non-industry-specific Evaluation Tools		
Cigref (2018)	"categorise the ethical issues linked to digital technology, from the point of view of both the user and the designer of digital solutions and/or services"	Grey literature, nonprofit sector
Open Roboethics Institute (2019)	"analyze the technology against values", Value Questions (p. 15)	Grey literature, nonprofit sector
European Commission (2020)	self-evaluation of trustworthy AI	Grey literature, public sector
Lifshitz & McMaster (2020)	starting point for legal and ethical considerations involved in the acquisition and use of an AI system	Grey literature, commercial sector
Personal Data Protection Commission Singapore (PDPC) et al. (2020a)	"help organizations assess the alignment of their AI governance processes with the Model Framework (PDPC, 2020b), identify potential	Grey literature, multisectoral collaboration

	gaps in their existing processes and address them accordingly”	
National Institution for Transforming India (NITI Aayog, 2021)	“guide to help assess the AI governance readiness of stakeholders as per the Responsible AI principles in this document” (p. 44)	Grey literature, public sector
ITechLaw (2021)	“help measure, in quantifiable and real terms, the impact of a proposed AI solution” (i.e., responsible AI impact assessment)	Grey literature, nonprofit sector
Roll’s Royce (2021)	“checklist of measures to ensure the initial design of the AI application is ethical, and that its resulting outputs remain unbiased and true to the intended design”	Grey literature, commercial sector
Torres et al. (2021)	ethics self-assessment	Grey literature, nonprofit sector
Han & Choi (2022)	“help [...] to improve and guarantee the reliability of AI service from planning to operation”	Peer-reviewed literature, academic sector
VDE Verband der Elektrotechnik, Elektronik und Informationstechnik e.V. (VDE, 2022)	“describe whether a product adheres to specific values and can be trusted”	Grey literature, nonprofit sector
Digital Dubai (2022)	“enable AI developer organisations or AI operator organisations to evaluate the ethics level of an AI system”	Grey literature, public sector
Commission Nationale de l’Informatique et des Libertés (CNIL, 2022)	“self-assessment of all relevant aspects in terms of personal data and ethics for a processing project”	Grey literature, public sector
Ministry of Economy, Trade and	“support the evaluation of conformity to the AI governance goals in the	Grey literature, public sector

Industry (Japan, 2022)	development and operation of individual AI systems" (checklist from p. 61)	
Mylrea & Robinson (2023)	"to enhance trust in the design and management of AI systems"	Peer-reviewed literature, academic sector
Saudi Data & AI Authority (2023)	assess compliance with Saudi Arabia's AI Ethics Principles	Grey literature, public sector
TrustArc Canada Inc. (2024)	"help developers and deployers of artificial intelligence (AI) systems, including deployers of AI systems procured from a third-party, to examine their AI-related activities to determine if they meet applicable best practices"	Grey literature, commercial sector

Literature sources on economic aspects

Authors	Purpose	Source and Sector
White et al. (2023)	"recommendations to support the conduct and reporting of economic evaluations for CDSS-based interventions"	Peer-reviewed literature, academic sector
Elvidge et al. (2024)	"ensure EEs [economic evaluations] of AI-based health interventions are reported in a transparent and reproducible manner"	Peer-reviewed literature, multisectoral collaboration
Schwab Foundation for Social Entrepreneurship & World Economic Forum (WEF) (2024)	"guide social innovators and other organizations through the nuanced landscape of AI integration" (chapter "2.5. Costs and metrics" was reviewed)	Grey literature, multisectoral collaboration

Literature sources on usability and workflow integration

Authors	Purpose	Source and Sector
Fossum et al. (2011)	evaluate the usability of a CDSS	Peer-reviewed literature, academic sector
Horsky et al. (2012)	"questions ['about general system attributes related to usability and	Peer-reviewed literature, multisectoral collaboration

	safety'] clinicians may pose to vendors when considering a purchase"	
Silveira et al. (2019)	to evaluate the usability of a CDSS (Table 4)	Peer-reviewed literature, academic sector
Salwei et al. (2021)	"support consideration of workflow integration during the design of health IT in order to improve usability of the technology when implemented"	Peer-reviewed literature, academic sector
Ghorayeb et al. (2022)	"to improve the usability of a newly designed or implemented CDSS or HIS in an efficient way"	Peer-reviewed literature, academic sector
Tegenaw et al. (2023)	"evaluate the user acceptance of a clinical decision support (CDS) instrument at the point of care (POC) in low-resource settings (LRS)", based on Ji et al. (2021)	Peer-reviewed literature, academic sector

Literature sources on AI maturity

Authors	Purpose	Source and Sector
Cisco (n.d.)	"helps companies understand their level of readiness across each of these pillars" (Strategy, Infrastructure, Data, Governance, Talent, and Culture)	Grey literature, commercial sector
Microsoft (n.d.)	"help identify your readiness to begin realizing meaningful business value from AI"	Grey literature, commercial sector
Scalefocus (n.d.)	"to assess your organization's readiness score for adopting AI"	Grey literature, commercial sector
Writer (n.d.)	"gain a comprehensive understanding of your organization's readiness to navigate AI trends, engage stakeholders, prioritize use cases, drive change, build skills,	

	scale initiatives, foster innovation, and ensure ethical practices”	
Intel (2018)	“to assess where you are and increase your (AI) readiness as a result”	Grey literature, multisectoral collaboration
Nortje & Grobbelaar (2020)	“assist in the implementation of AI into a business’ structures”	Peer-reviewed literature, academic sector
Pumplun et al. (2021)	“help researchers and clinicians understand the possible range of ML adoption stages in clinics and determine an overarching maturity score”	Peer-reviewed literature, academic sector
Correlation One (2023)	“delve into the six crucial building blocks that form the backbone of an AI readiness framework”	Grey literature, commercial sector
Deloitte (2024)	“to achieve enterprise AI readiness and maturity”	Grey literature, commercial sector
Future Processing (2024)	“for organisations to ascertain their preparedness for AI deployment”	Grey literature, commercial sector
CluedIn (2024)	“provide actionable advice for data leaders and practitioners on preparing their organizations to incorporate AI”	Grey literature, commercial sector
Kavanaugh et al. (2024); Infosys (n.d.)	“reduce apprehension about AI, close gaps, and garner more benefits from AI”	Grey literature, commercial sector
Passerelle (2024)	“to determine if you have the data governance, data quality and data management in place to ensure your application output can be trusted”	Grey literature, commercial sector
Virginia Office of Data Governance and Analytics (2024)	“A data AI checklist that organizations can leverage to see how prepared their data may be for AI”	Grey literature, public sector
Rohn (2025)	“Drawing from established models like the Fusemachines AI Readiness	Grey literature, commercial sector

	Framework and Cisco's 2024 AI Readiness Index, we've synthesized the most actionable components enterprises must align to unlock sustainable AI success"	

Literature sources on AI vendor reliability and support

Authors	Purpose	Source and Sector
Healthcare-specific Records		
Aidoc (n.d.)	"to signal if an AI vendor is the right long-term partner for your health system"	Grey literature, commercial sector
California Telehealth Resource Center (2024)	"provides a set of questions for a health care provider to ask AI tool developers/vendors"	Grey literature, nonprofit sector
Chae et al. (2024)	"Risk Stratification Questions to Ask Potential Commercial AI Vendors"	Peer-reviewed literature, academic sector
Sutten (2025)	"dive deep into the different parameters for assessing the AI vendors before integrating them into healthcare operations"	Grey literature, commercial sector
Non-industry-specific Records		
Mimecast (n.d.)	"to enable corporate buyers to accurately assess AI for use cases across the enterprise without needing a data science degree to do it"	Grey literature, commercial sector
IT Convergence (2023)	"to ensure a vendor selection that aligns with your business goals and delivers the desired outcomes"	Grey literature, commercial sector
Hosch & Morris, PLLC (2024)	"provides a solid starting point for evaluating AI vendors"	Grey literature, commercial sector
Norton (2024)	"checklist of things to consider when you're evaluating AI vendors, to ensure that you choose the right solution for your business needs and	Grey literature, commercial sector

	maximize value from the technology as quickly as possible”	
Stout (2025)	measures to be covered in contracts with AI vendors to “establish crucial legal protections”	Grey literature, commercial sector

Cross-categorical literature sources

Authors	Purpose	Source and Sector
Van de Velde et al. (2018)	“facilitate a deeper and more accurate understanding of which factors make CDS more (or less) effective and to guide CDS implementation by preventing key factors from being overlooked”	Peer-reviewed literature, multisectoral collaboration
Cresswell et al. (2019)	“help to ensure that DSS are implemented/used in a way that is person centred, safe, reliable and inclusive whilst at the same time supporting service re-design and innovation”	Grey literature, multisectoral collaboration
NHSX (2020)	“sets out the important questions you need to consider in order to make well-informed buying decisions about “off-the-shelf” AI products”	Grey literature, public sector
Fasterholdt et al. (2022)	“to support decision-makers when deciding whether a mature AI application should be implemented into clinical practice or not, i.e. support adoption decisions”	Peer-reviewed literature, multisectoral collaboration
Riester & Zullo et al. (2023)	“to be used as a checklist by pharmacists or their trainees (eg, residents) when developing and implementing a pre diction tool” (only the checklist in table 3 was included)	Peer-reviewed literature, academic sector
Apfelbacher et al. (2024)	“guideline for the implementation of AI applications in university hospitals”	Peer-reviewed literature, academic sector

Hendriks et al. (2024)	“guide CDSS integration more successfully in the clinical workflow to support MDTs [multidisciplinary teams] in the future”	Peer-reviewed literature, multisectoral collaboration
Awad et al. (2024)	“guides the appropriate consideration of CDS and options”	Peer-reviewed literature, multisectoral collaboration
Stade et al. (2025)	“To ensure the responsible deployment of AI mental health systems and to support decision-making regarding use of AI”	Peer-reviewed literature, academic sector
You et al. (2025)	“framework for healthcare organizations to implement AI technologies safely and with impact, beyond scientific research, using in-house developed tools or vendor-based solutions”	Peer-reviewed literature, academic sector
Bottacin et al. (2025)	“Healthcare professionals implementing AI tools can use MedinAI to assess model documentation completeness, ensuring a comprehensive view for informed and precise decision making.”	Peer-reviewed literature, academic sector
Owoyemi et al. (2025)	“addresses the sociotechnical aspects of AI deployment in health care and provides a structured, holistic guide for teams involved in the life cycle of AI systems”	Peer-reviewed literature, academic sector
Rusanov et al. (2025)	“selecting and evaluating AI auto-segmentation systems in clinical radiotherapy”	Peer-reviewed literature, multisectoral collaboration

Supplementary Material 11: Regions of guidance and evaluation tool developers

Author	Region
Elvidge et al. (2024)	Multiple countries
Schwab Foundation for Social Entrepreneurship & WEF (2024)	Multiple countries
White et al. (2023)	Australia
Fossum et al. (2011)	Multiple countries
Horsky et al. (2012)	US
Silveira et al. (2019)	Brazil
Ghorayeb et al. (2022)	UK
Tegenaw et al. (2023)	Multiple countries
Nortje & Grobbelaar (2020)	South Africa
Pumplun et al. (2021)	Germany
Chae et al. (2024)	US
Van de Velde et al. (2018)	Multiple countries
Fasterholdt et al. (2022)	Multiple countries
Riester & Zullo et al. (2023)	US
Apfelbacher et al. (2024)	Germany
Hendriks et al. (2024)	The Netherlands
Awad et al. (2024)	Australia
Stade et al. (2025)	US
You et al. (2025)	US
Bottacin et al. (2025)	Brazil
Owoyemi et al. (2025)	US
Rusanov et al. (2025)	Australia
Reddy et al. (2021)	Multiple countries
Scott et al. (2021)	Australia
Han & Choi (2022)	South Korea
Vollmer et al. (2020)	Multiple countries
European Commission (2020)	EU/Multiple countries
Open Roboethics Institute (2019)	Canada
Roll's Royce (2021)	UK
Lifshitz & McMaster (2020)	Canada
PDPC et al. (2020a)	Singapore
NITI Aayog (2021)	India

ITechLaw (2021)	Multiple countries
VDE (2022)	Germany
Digital Dubai (2022)	United Arab Emirates
CNIL (2022)	France
Torres (2021)	US
Saudi Data & AI Authority (2023)	Saudi Arabia
Cigref (2018)	France
Al-Zaiti et al. (2022)	Multiple countries
Wang et al. (2022)	US
Crigger et al. (2022)	US
Szabo et al. (2022)	Multiple countries
Nazer et al. (2023)	Multiple countries
Emani et al. (2023)	US
Van Smeden et al. (2022)	Multiple countries
TrustArc (2024)	Canada
Moons et al. (2025)	Multiple countries
Mylrea & Robinson (2023)	US
Bacchi et al. (2024)	Australia
Coalition for Health AI et al. (2024)	US
Ministry of Economy, Trade and Industry (Japan) (2022)	Japan
Salwei et al. (2021)	US
Cisco (n.d.)	US
Microsoft (n.d.)	US
Scalefocus (n.d.)	Bulgaria
Writer (n.d.)	US
Intel (2018)	US
Correlation One (2023)	US
Future Processing (2024)	Poland
CluedIn (2024)	Denmark
Infosys (n.d.)	India
Passerelle (2024)	US
Virginia Office of Data Governance and Analytics (2024)	US
Rohn (2025)	US
Aidoc (n.d.)	Israel
California Telehealth Resource Center (2024)	US
Sutten (2025)	US

Mimecast (n.d.)	UK
IT Convergence (2023)	US
Hosch & Morris, PLLC (2024)	US
Norton (2024)	UK
Stout (2025)	US
Cresswell et al. (2019)	UK
Deloitte (2024)	UK
NHSX (2020)	UK

Supplementary Material 12: List of AI-CDSS adoption considerations

Note: Blank cells for references mean that this question was formulated based on a general understanding of AI-CDSS adoption considerations gained through the literature and expert interviews, and not based on a specific reference.

Category	Consideration	References
Regulatory and Legal Compliance	"What is the intended use of the product? What can it be used for and under what conditions can it be used? What can it not be used for?"	NHSX (2020)
	Has compliance with all applicable laws and regulations been ensured? This includes, among others, approval for market entry (e.g., CE certification in the EU, FDA approval in the U.S.), as well as laws and regulations on • data (e.g., GDPR, HIPAA) • medical devices, • safety, • human oversight, • and non-discrimination.	
	Do you have designated staff for keeping your organisation current on regulatory developments?	TrustArc Canada Inc. (2024)
	"Is the AI model developed so it is easy to adapt to new legislation, as an example proposal about an AI regulation?"	Fasterholdt et al. (2022)
	Have you established a process to assess whether modifications to the AI-CDSS trigger new regulatory compliance requirements or alter existing obligations?	
	Does the AI-CDSS "account for legal and regulatory differences across markets and regions when making medication recommendations" or recommendations in a different area (e.g., surgical procedures)? "If a recommended service falls outside legal parameters, the model should generate an alert, informing healthcare providers and preventing legal risks while ensuring compliance."	Bottacin et al. (2025)

	Have you considered the liability you carry for claims related to the AI-CDSS?	
Utility	"What is the problem you are trying to solve?"	NHSX (2020)
	"What is the rationale for choosing AI to solve your problem? What is it about AI - over and above other solutions - that makes it a powerful choice?" The adoption of AI-CDSS should be driven by its proven benefit in effectively addressing a clinical need, rather than implementing 'AI for the sake of AI'.	
	"What is the baseline you are looking to improve, and what metrics matter in measuring this improvement?"	
	"Are outcome measured important, clinically acceptable, transparent, feasible and usable?"	Cresswell et al. (2019)
	"What is the evidence base for demonstrating the product's effectiveness? Is the standard of this evidence sufficiently robust, taking into account the function and associated risk of the product?" • Has the use of the AI-CDSS been shown to improve the quality of care for a clearly defined patient population? (incl. improved patient outcomes, and prevented clinical errors) The system's effectiveness should be measured in an environment representative of your healthcare practice, including a sample representative of the target population.	NHSX (2020); Scott et al. (2021); Ghorayeb et al. (2022); Bacchi et al. (2024)
	What comparator was used to assess the effectiveness of the AI-CDSS (e.g., standard of care, other technology, other CDSS)?	You et al. (2025)
	"What insight is available on the product's effectiveness in other health and care settings?"	NHSX (2020)
	"Is the reported gain in statistical performance with the ML/AI algorithm justified in the context of any trade-offs?" Trade-offs include, but are not limited to costs (e.g., computational costs) and complexity (i.e., making explanations of clinical decisions more difficult).	Vollmer et al. (2020)

	Does the use of the system benefit healthcare professionals and help them work more efficiently? Does the system make it easier to • make efficient decisions, • prioritise the daily workload, and collaborate with colleagues?	Ghorayeb et al. (2022)
	Does the AI-CDSS empower healthcare professionals in their work?	ITechLaw (2021)
	Does the AI-CDSS provide "a useful summary view of the patient's current health status?"	Ghorayeb et al. (2022)
	"What is the associated risk of not implementing the CDS? Consider severity and likelihood."	Awad et al. (2024)
	Have you planned a dynamic process to increase the utility of the AI-CDSS post-deployment? For example, by learning from experience and user feedback, or by understanding the value of system updates to enhance the efficiency of working with the system? This process should include regular measurement of relevant outcomes.	Riester & Zullo et al. (2023)
	"Are there processes in place to evaluate the availability of more effective AI methodologies and to determine when it's appropriate to transition to a newer AI solution?"	CHAI (2024)
Trustworthy AI: Human Agency and Oversight	Are measures in place, through system design or user instruction and education to prevent over-reliance on or overconfidence in the AI-CDSS?	European Commission (2020); Saudi Data & AI Authority (2023)
	Are tasks clearly divided between the AI-CDSS and healthcare professionals, with a defined hierarchy of decision-making?	Lifshitz & McMaster (2020); Roll's Royce (2021)
	"Could the AI system affect human autonomy by interfering with the end-user's decision-making process in any other unintended and undesirable way? Did you	European Commission (2020)

	[or the vendor] put in place any procedure to avoid that the AI system inadvertently affects human autonomy?"	
	Did the AI-CDSS design consider human psychology to prevent confusion of the user, manipulation, cognitive biases (e.g., confirmation bias), and any potential stress or cognitive overload resulting from complex or conflicting recommendations?	Saudi Data & AI Authority (2023)
	"Did you [and/or the vendor] evaluate the suitable level of human control and oversight for the AI system and its use case? • Can you explain how humans control, oversee, and interact with the AI system? • Did you implement systems and processes to guarantee human control or oversight?" • Are responsibilities and authorities in overseeing the AI-CDSS, handling errors, and contesting and remedying/overriding decisions clearly assigned?	Lifshitz & McMaster (2020); NITI Aayog (2021); Saudi Data & AI Authority (2023)
	Which oversight approach applies to the AI-CDSS you are considering adopting: human-in-the-loop, human-on-the-loop, or human-in-command? And are users trained on how to exercise oversight? Are the necessary oversight capabilities for the chosen approach (e.g., transparency of decision criteria and post-hoc change options for human-on-the-loop) implemented and accessible to users?	European Commission (2020); VDE (2022)
	"Has the development of Key Performance Indicators (KPIs) of the AI system been considered to help determine human involvement's relevance in case the system deteriorates?"	Torres et al. (2021)
	"Did you establish any detection and response mechanisms for undesirable adverse effects of the AI system for the end-user or subject?"	European Commission (2020)
	"Did you ensure a 'stop button' or procedure to safely abort an operation when needed?"	European Commission (2020)

	Is there a mechanism in place to switch to a process where the AI-CDSS is not used, in case of problems with the behaviour of the system?	Ministry of Economy, Trade and Industry (2022)
Trustworthy AI: Technical Robustness and Safety	"How does the vendor evidence model robustness? Can the model make reliable predictions, given that data is subject to uncertainty and errors? Does the model remain effective even in extreme or unexpected situations?"	NHSX (2020)
	"Could the AI system have adversarial, critical or damaging effects (e.g. to human or societal safety) in case of risks or threats such as design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?"	European Commission (2020)
	"Is the AI system certified for cybersecurity (e.g. the certification scheme created by the Cybersecurity Act in Europe)19 or is it compliant with specific security standards?"	European Commission (2020)
	How exposed is the AI-CDSS to cyber-attacks? Has the vendor shared how the system's vulnerability to potential forms of attack (e.g., data poisoning, model invasion, model inversion) was assessed?	European Commission (2020)
	What measures has the vendor implemented to ensure AI-CDSS' integrity, robustness, and security against possible attacks during its lifecycle (e.g., red teaming and pen testing, monitoring for anomalies)? Is there designated personnel to monitor the security of the system?	European Commission (2020); CHAI (2024)
	Is there a procedure in place for you to raise security concerns related to the AI-CDSS?	CHAI (2024)
	"Does the organization have emergency protocols in place for potential cyber-attacks and data loss, among others, such as immediate technical changes or human intervention to reduce risks?"	Torres et al. (2021)
	Has your organisation, along with the users of the AI-CDSS, been informed about how long the vendor will provide security coverage and updates for the system?	European Commission (2020)

	"Has a definition been set of what is considered to be a safe and reliable AI System, and is this definition commonly used and implemented throughout the full lifecycle of design, development, deployment, operation and use of the AI System? • Have quantitative analysis or metrics been applied to measure and test the applied definition? • Are there regulatory requirements that impact the above definition of safety and reliability (e.g. medical devices regulations)?"	ITechLaw (2021)
	Have safety risks (e.g., technical defects, incorrect medication suggestions), risk metrics, and levels of risk been defined? Has a process for continuous monitoring and assessment of safety risks (incl. mechanisms to override AI operation if necessary), along with a procedure for reporting such risks to the vendor, healthcare staff, patients and other stakeholders?	European Commission (2020); Lifshitz & McMaster (2020); ITechLaw (2021)
	Do you share the responsibility for safety with the vendor?	Cresswell et al. (2019)
	Does the vendor have a designated contact person for concerns and emergencies related to safety?	Apfelbacher et al. (2024)
	Have you asked your vendor the following questions about threats to the AI-CDSS? "Did you identify the possible threats to the AI system (design faults, technical faults, environmental threats) and the possible consequences? • Did you assess the risk of possible malicious use, misuse or inappropriate use of the AI system? • Did you define safety criticality levels (e.g. related to human integrity) of the possible consequences of faults or misuse of the AI system?" What specific measures has the vendor taken to prevent or mitigate these threats?	European Commission (2020)

	Has the vendor assessed the extent to which the AI-CDSS outputs depend on the system's stable and reliable behaviour, and were the reliability/testing requirements aligned with the necessary levels of stability and reliability?	European Commission (2020)
	Has the vendor planned fault tolerance (e.g., a duplicated system) to ensure continued operation in case of failure?	European Commission (2020); ITechLaw (2021)
	"Did you [the vendor] develop a mechanism to evaluate when the AI system has been changed to merit a new review of its technical robustness and safety?"	European Commission (2020)
	Has the vendor shared the methods and outcomes of the AI system's internal validation, including performance metrics?	
	Has the vendor shared the methods and outcomes of the AI system's external validation, providing evidence of the system's generalizability to new data?	
	"Could a low level of accuracy of the AI system result in critical, adversarial or damaging consequences?"	European Commission (2020)
	"What are the implications of false positives? What are the implications of false negatives? Are the appropriate decision makers aware of the balancing of risks between the two?"	
	"Did you [the vendor] put in place measures to ensure that the data (including training data) used to develop the AI system is up-to-date, of high quality, complete and representative of the environment the system will be deployed in?"	European Commission (2020)
	How do you ensure that data collected (e.g., from patient records) and used during the operation of the AI-CDSS is also current, accurate, and complete?	Open Roboethics Institute (2019)
	Have measures been put in place to monitor, document, and increase the accuracy of the AI-CDSS, including a threshold for inaccurate outputs and measures in place when this threshold is surpassed?	European Commission (2020)

	Have measures been put in place to monitor the performance of the AI-CDSS during its use and address performance issues? This goes beyond accuracy monitoring and extends to other metrics and system attributes, such as the response time and latency of the system.	
	"Have potential long-term risks associated with the model's performance (that is, risks not measurable during the pilot stage but potentially arising in deployment) been identified?"	CHAI (2024)
	Can the operation of the AI-CDSS invalidate its training data or underlying assumptions (e.g., through data drift), potentially leading to adversarial effects?	European Commission (2020)
	"Could the AI system cause critical, adversarial, or damaging consequences (e.g. pertaining to human safety) in case of low reliability and/or reproducibility? • Did you [the vendor] put in place a well-defined process to monitor if the AI system is meeting the intended goals?22 • Did you test whether specific contexts or conditions need to be taken into account to ensure reproducibility?"	European Commission (2020)
	"Did you [the vendor] put in place verification and validation methods and documentation (e.g. logging) to evaluate and ensure different aspects of the AI system's reliability and reproducibility?"	European Commission (2020)
	Will the AI-CDSS maintain reliable performance (i.e., robustness) when encountering variations or unexpected data inputs, such as outliers, missing values, or noisy data?	Ministry of Economy, Trade and Industry (2022); VDE (2022)
	Has the vendor established a procedure for managing cases where the AI-CDSS produces outputs with a low confidence score?	European Commission (2020)
	"Is your AI system using (online) continual learning?	European Commission (2020)

	• Did you consider potential negative consequences from the AI system learning novel or unusual methods to score well on its objective function?"	
	Has your vendor developed a contingency plan to address unexpected situations, adversarial attacks, and other adverse events (incl. defining and explaining how the system could fail)?	European Commission (2020); Lifshitz & McMaster (2020)
	Has the vendor defined tested fallback plans to address errors of the AI-CDSS, adversarial attacks, and other unexpected situations? Are governance procedures established to trigger these fallback plans?	European Commission (2020); Saudi Data & AI Authority (2023)
Trustworthy AI: Privacy and Data Governance	"Did you [and the vendor] consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection?"	European Commission (2020)
	Have mechanism been established "that allow flagging issues related to privacy concerning the AI system?"	
	"Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?"	
	"Did you [or the vendor] put in place any of the following measures some of which are mandatory under the General Data Protection Regulation (GDPR), or a non-European equivalent? • Data Protection Impact Assessment (DPIA)[23]; • Designate a Data Protection Officer (DPO)[24] and include them at an early state in the development, procurement or use phase of the AI system; • Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access and making modifications);	

	• Measures to achieve privacy-by-design and default (e.g. encryption, pseudonymisation, aggregation, anonymisation); • Data minimisation, in particular personal data (including special categories of data); Did you implement the right to withdraw consent, the right to object and the right to be forgotten into the development of the AI system? Did you consider the privacy and data protection implications of data collected, generated or processed over the course of the AI system's life cycle?"	
	"Did you [the vendor] consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data?"	
	"Did you align the AI system with relevant standards (e.g. ISO25, IEEE26) or widely adopted protocols for (daily) data management and governance?"	
	Are documented data breach notification procedures in place?	
	Is it clearly defined who can access the AI-CDSS? How is it controlled that no unauthorised access is taking place?	VDE (2022)
	Where is the AI-generated content (e.g., treatment recommendations, diagnostic insights, or clinical alerts) stored?	Norton (2024)
	"On decommissioning the product, what will happen to any data that is stored outside of your organisation's systems? Will it be deleted, or archived?"	NHSX (2020)
	"How will you ensure that you have access to any data or analysis you require that is due to be deleted or archived?"	
	"On decommissioning the product, how will you ensure that the vendor's access to any part of your organisation's infrastructure is revoked in full?"	

Trustworthy AI: Transparency	Has traceability of the AI-CDSS during its entire lifecycle been ensured? • "Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)?" • "Can you trace back which AI model or rules led to the decision(s) or recommendation(s) of the AI system?" • "Did you put adequate logging practices in place to record the decision(s) or recommendation(s) of the AI system?"	European Commission (2020)
	Can the vendor share how the system was designed and developed (i.e., development of the rule set for rule-based AI systems; training data and methods for learning-based AI systems)?	Lifshitz & McMaster (2020)
	Can the vendor share how the system was tested and validated (i.e., scenarios/situations used for rule-based AI systems; test data used for learning-based AI systems)?	
	Can the vendor share documentation on the composition of the AI-CDSS, including software and hardware components, and are procedures in place to track updates and changes? This is relevant during incident management to determine whether a specific component of the system was faulty and caused or contributed to an error.	VDE (2022)
	"How does the vendor evidence model explainability? Can predictions [or recommendations] made by the model be explained in terms that both a trained user of the product and a patient/service user would understand?"	NHSX (2020)
	Are healthcare professionals continuously surveyed if they understand the explanations of how an AI-CDSS arrived at a specific output or recommendation?	European Commission (2020)
	"Where explainability cannot be practically achieved, did your organization [the vendor] consider lesser alternatives?"	PDPC et al. (2020a)

	Does the vendor provide accurate, up-to-date, understandable, and easily accessible Terms of Use?	ITechLaw (2021)
	Will necessary information about the system be provided to those involved with (e.g., healthcare professionals, legal staff) or affected by the AI-CDSS (e.g., patients) in an understandable (ideally tailored to audience, purpose, and context) and easily accessible manner? This information should include, but is not limited to: • the purpose of the AI-CDSS and its intended use (patients should also be informed that the care decision affecting them is supported by AI), • rationale for AI deployment/use and the system's benefits, • the system's attributes and associated information (e.g., degree of reliability, accuracy/ performance, confidence value, cybersecurity), • limitations and risks (e.g., biases) of the system, • functionalities of the AI-CDSS and how it works (e.g., output generation), • system updates and changes (incl. schedules for updates, their duration, and information on the updates/changes done), • the data collected and used by the system, • data processing (e.g., informing staff using the systems, informing patients how their data is used), • impact of AI outputs on individuals and how healthcare staff should use these outputs, • rights related to the use of the system, • deficiencies (e.g., errors) and problems of the AI-CDSS (e.g., attacks) when detected, • AI deployment and management plans and decisions (e.g., justification if no dedicated group was established for AI-CDSS monitoring),	European Commission (2020); Lifshitz & McMaster (2020); PDPC et al. (2020a); ITechLaw (2021); Torres et al. (2021); VDE (2022); CHAI (2024)

	• and clarification of concepts such as 'AI fairness' in a taxonomy or glossary.	
	"Are any other disclosures made with respect to the transparency and explainability of the AI System (e.g. videos, icons, symbols, white papers, dashboards, or counterfactual interfaces)?"	ITechLaw (2021)
	Do both the vendor and your healthcare facility have clear policies on how transparency around the AI-CDSS is maintained, including cases in which certain stakeholders should receive additional information?	
	Are the communication and feedback channels — both for communication within your healthcare facility and with your vendor — supervised by qualified personnel?	Torres et al. (2021)
Trustworthy AI: Diversity, Fairness, and Non-discrimination	"How does the vendor evidence model fairness? What measures are in place to prevent the model from discovering hidden patterns of discrimination in its training data, reproducing these patterns and making biased predictions as a result?" The ALTAI (European Commission 2020) provides further guidance on which considerations should be addressed.	NHSX (2020)
	Is the risk of unfair bias that systematically and unjustifiably favours or discriminates against particular individuals or groups based on characteristics such as gender, age, and race mitigated in the AI-CDSS? Measures to avoid such bias include, but are not limited to: • having a clear definition of fairness, • having diverse teams in the design, development, and deployment of the AI-CDSS, • using training data that is representative of the target population, • scrutinising algorithmic design and modelling choices, as well as data labelling, • training and educating users of the AI-CDSS to avoid biases that could be introduced by their interactions with the system (e.g., different	Open Roboethics Institute (2019); European Commission (2020); Lifshitz & McMaster (2020); PDPC et al. (2020a); ITechLaw (2021); NITI Aayog (2021); Torres et al. (2021); CNIL (2022); Crigger et al. (2022); Han & Choi (2022); Ministry of Economy, Trade

	interpretation of outputs for different sub-groups) • testing the AI-CDSS for specific target groups and use cases, • testing and monitoring for biases with defined fairness metrics during the entire AI-CDSS lifecycle, • establishing a mechanism for flagging and communicating bias-related issues, • and establishing procedures to respond to and mitigate biases.	and Industry (2022); Szabo et al. (2022); VDE (2022); Wang et al. (2022); Emani et al. (2023); Nazer et al. (2023); Saudi Data & AI Authority (2023); CHAI (2024); Moons et al. (2025)
	"When it comes to talent management, has your company started to think about 'accessibility' of AI technologies for employees who are differently abled?" Is information about the AI-CDSS and its user interface usable and accessible to those employees?	Cisco (n.d.); Cigref (2018); European Commission (2020); Lifshitz & McMaster (2020); ITechLaw (2021)
	Have other potential barriers preventing equal access to the AI-CDSS been considered for all intended users (e.g., training gaps, role-based restrictions)?	CHAI (2024)
	"Are the benefits and efforts relatively equal for all stakeholders?"	Cresswell et al. (2019)
	Have you and the vendor considered if the use of the system can disproportionately benefit or disadvantage certain patient groups (as classified by characteristics such as gender, race, and socioeconomic status) and contribute to unequal access to healthcare?	European Commission (2020); Lifshitz & McMaster (2020); Crigger et al. (2022); CHAI (2024)
	Were healthcare professionals actively involved in the design and development of the AI-CDSS? Did the vendor also include other relevant stakeholders, such as patient representatives?	European Commission (2020); Lifshitz & McMaster

		(2020); Torres et al. (2021); Saudi Data & AI Authority (2023); CHAI (2024)
	Is there a mechanism for stakeholders (healthcare professionals, AI experts, patients, etc.) to provide feedback regarding the deployment and use of the AI-CDSS (organisation-level feedback)?	Scalefocus (n.d.)
	"When and how should patients be involved in data collection, analysis, deployment, and use?"	Vollmer et al. (2020)
	Are those who have been involved in providing feedback related to the AI-CDSS representative of different social groups, backgrounds, and experiences?	Lifshitz & McMaster (2020)
Trustworthy AI: Societal and Environmental Well-being	In addition to considering risks related to the trustworthy AI requirements above (e.g., privacy, transparency, non-discrimination), have you considered if the deployment and use can risk violating any other fundamental human rights of your patients and employees? Has an impact assessment been conducted for stakeholders potentially affected by the AI-CDSS?	
	"Are there potential negative impacts of the AI system on the environment? Which potential impact(s) do you [or the vendor] identify?"	European Commission (2020)
	Are mechanisms established to assess the environmental impact of the AI-CDSS's development, deployment, and use (e.g., energy consumption, carbon emissions)? Have measures been defined to reduce the negative environmental impact?	European Commission (2020)
Trustworthy AI: Accountability	Is it possible for independent third parties to audit the AI-CDSS (e.g., by ensuring traceability and logging the system's processes and impacts)?	European Commission (2020); NITI Aayog (2021)
	"Did you foresee any kind of external guidance or third-party auditing processes to oversee ethical concerns and accountability measures?"	European Commission (2020)

	Have you evaluated how you as a healthcare provider might be impacted during third-party audits of the AI-CDSS, including potential disruptions, data access requests, or liability considerations? Have you thought about measures to address potential impacts of audits?	
	Does the development, testing, deployment, and/or use of the AI-CDSS comply with certain standards (e.g., ISO, IEEE), codes of conduct, or best practices?	CNIL (2022); TrustArc Canada Inc. (2024)
	"Did you consider establishing an AI ethics review board or a similar mechanism to discuss the overall accountability and ethics practices, including potential unclear grey areas?"	European Commission (2020)
	Do you have a process in place to monitor adherence to the six aforementioned trustworthy AI requirements, and are those involved in the monitoring appropriately trained to do so? • "Does this process include identification and documentation of conflicts between the 6 aforementioned requirements or between different ethical principles and explanation of the 'trade-off' decisions made?"	European Commission (2020)
	Have processes been established for relevant stakeholders (e.g., clinical staff, IT staff, vendor representatives, patients and patient advocates) to report potential vulnerabilities, risks, and adverse impacts (e.g., bias) in the AI-CDSS?	European Commission (2020); TrustArc Canada Inc. (2024)
	Are redress mechanisms established (either by-design in the system or on an operational level) to address potential adverse effects on individuals? Are users and third parties informed about redress methods?	European Commission (2020); Lifshitz & McMaster (2020); NITI Aayog (2021)
	In addition to the previous questions, have you considered if the deployment and use of the AI-CDSS in your healthcare facility would require any further measures to timely identify, assess, and manage risks (going beyond safety risks, as addressed under	Ministry of Economy, Trade and Industry (2022); VDE (2022)

	'Technical Robustness and Safety', and also covering risks such as workflow disruptions, human autonomy impacts, reputational risks), including the establishment of a structure that enables timely handling of problems and mitigation of system failures?	
	How are risk management responsibilities divided between you and the vendor?	
	"How are the results of all risk assessment, risk management and risk control procedures in relation to safety and reliability of the AI System factored into necessary or desirable alterations of (the design of) the AI System? How is this process documented?"	ITechLaw (2021)
	Have you developed a plan to regularly assess and update risk management systems?	Torres et al. (2021)
	"Has the development of emergency responses based on the analysis of scenarios been considered if risk mitigation measures fail?"	Torres et al. (2021)
	Are all roles and responsibilities in the deployment, use, and governance of the AI-CDSS clearly defined (e.g., legal persons in charge, persons monitoring the system, persons responsible for data governance)?	
	Have you already planned the development of a usage and governance policy for the AI-CDSS that addresses the seven trustworthy AI requirements and other considerations for the system's safe and effective use, as listed in this checklist and applied to your specific setting (e.g., by including a Code of Conduct)?	
	"Is there a de-implementation plan in place and understood by end users, outlining the process for discontinuing the use of the model when necessary?"	CHAI (2024)
Economic Aspects	Have you considered all costs associated with procuring, deploying, and using the AI-CDSS? (i.e., costs for rights of use, system running, system maintenance and updates/optimisation, evaluation activities, hardware, infrastructure, IT integration, legal and compliance matters, personnel and change management, staff training, and opportunity costs)	Schwab Foundation for Social Entrepreneurship & WEF (2024)

	"If you are replacing an older system with the new technology, have you factored in time, costs and potential complications of dealing with a legacy system?"	NHSX (2020
	Has an economic evaluation been conducted? If yes, did it • evaluate the AI-CDSS across multiple sites, • evaluate the costs and health outcomes over its full life-cycle, • consider different direct and indirect costs, as mentioned above, • and consider heterogeneity and uncertainty in measured costs and outcomes, as well as factors that could have affected the results?	White et al. (2023)
	Have the outcomes of using the AI-CDSS been quantified in monetary terms? And has an analysis of the return on investment been conducted?	Scalefocus (n.d.); Nortje & Grobbelaar (2020)
	Was the reporting of the economic evaluation in line with the CHEERS or CHEERS-AI checklist?	Husereau et al. (2022); Elvidge et al. (2024)
	"How is the pricing structured and are there variable costs or the option to lock-in rates?"	Norton (2024)
	Are you familiar with the process for obtaining reimbursement for the AI-CDSS?	
	Do you have a dedicated budget for AI deployment and use?	Pumplun et al. (2021)
	Do you have a financial strategy in place to ensure sustainable funding for the deployment and use of the AI-CDSS?	Cisco (n.d.)
	"What is the extent of cover your own indemnifier or insurer can provide in the event of product failure or human error? Do you need to purchase additional cover or extend existing cover?"	NHSX (2020)
Usability	Is the AI-CDSS easy to use? This includes easily	Fossum et al. (2011) ; Silveira

	• navigating through the system with its user interfaces and functions, • becoming productive using its main menu, • seeing all information on the screen (incl. eye-catching display with an appropriate font style and size) and finding specific information, • entering data and completing the CDSS fields, • correcting a data entry error, • learning how to use the system, • and remembering how to use it.	et al. (2019); Ghorayeb et al. (2022); Tegenaw (2023)
	Are nomenclatures and terminologies used in the AI-CDSS clear and consistent?	Horsky et al. (2012), Silveira et al. (2019)
	Are the information and suggestions complete, understandable, and unambiguous, allowing the user to act on the support?	Ghorayeb et al. (2022)
	Does the AI-CDSS clearly present both the potential benefits and harms (incl. safeguards such as warnings to prevent unsafe decisions) of recommended actions?	Van de Velde et al. (2018)
	"Have transparency measures been defined to accommodate different user-facing views of model outcomes (e.g., providing options versus automatically ranking or triaging), ensuring that bias is mitigated?"	CHAI (2024)
	"Is it clear to the users why the decision support information is provided for a given patient?"	Van de Velde et al. (2018)
	Does the system provide all necessary information and interventions without overwhelming the user and causing alert fatigue?	Fossum et al. (2011), Ghorayeb et al. (2022); Owoyemi et al. (2025)
	"Is the clinical importance and urgency of the recommended action sufficiently clear?" Can the system adjust the saliency of interventions (e.g., alerts, recommendations) based on the severity?	Horsky et al. (2012); Van de Velde et al. (2018)
	"The available functionality of the system was complete."	Fossum et al. (2011)

	Does the system have a quick response time, which is appropriate in given clinical context? (from data input to outcome delivery, and with regards to user inquiries)	Tegenaw et al. (2023)
	"Can system and user actions be accessed and analyzed?" (i.e., activity logs)	Horsky et al. (2012)
	Is it possible for users to provide feedback within the AI-CDSS (system-level feedback)?	
Workflow Integration	Are you aware of the specific tasks or actions healthcare professionals have to carry out when using the AI-CDSS (e.g., entering patient information, reviewing recommendations)?	Bottacin et al. (2025)
	"Did you adopt measures to ensure that the impacts of the AI system on human work are well understood? Did you ensure that workers understand how the AI system operates, which capabilities it has and which it does not have?"	European Commission (2020)
	Is the fit of the AI-CDSS appropriately considered within the flow of sequential (i.e., support of the next task, accessibility "in a location near the preceding and succeeding tasks"), concurrent (incl. potential redirection of the user from another task), or discontinuous tasks?	Salwei et al. (2021)
	"Does the technology account for the flow of people? (e.g., flow of people through the physical environment as they complete tasks)"	
	"Does the technology appropriately account for the flow of information that can occur sequentially, in parallel, and discontinuously?"	
	"Does the technology align with the flow of other tools and technologies that are used? (e.g., other CDS, placing orders, computer use in patient rooms)"	
	"Does the technology fit in the clinical workflow accounting for changes that occur over extended time (i.e. beyond the scope of a single use)?"	
	Is the AI-CDSS available for use when needed, with the necessary tools/technologies (e.g., EHR) and information (e.g., lab results) easily accessible? This	Pumplun et al. (2021); Salwei et al. (2021)

	includes fully digitised patient data in a unified data format (e.g., FHIR).	
	"Does the technology fit in the workflow before, within, and after the patient visit?" This includes the clinician-patient interaction.	
	Does the AI-CDSS fit in the workflow of an individual healthcare professional, the care team, and the broader healthcare organisation?	
	"Are established professional role responsibilities redistributed?" If yes, have you considered whether the roles and workflows of different healthcare professional groups (e.g., clinicians, nurses) are influenced in distinct ways?	Horsky et al. (2012)
	Have you assessed if a redistribution of role responsibilities or any other possible effect of AI-CDSS deployment and use can negatively impact the work and skills of your staff (e.g., de-skilling of the workforce)?	European Commission (2020)
	"If significant changes to your organisation's ways of working are needed to realise the benefits promised by the product, is this possible?"	NHSX (2020)
	"If implementation of the product will cause short-term disruption, how will you manage this?"	
	In addition to the previous questions: Are there any (other) potential unintended consequences for the workflow in the healthcare facility?	Horsky et al. (2012)
	"Is the model multimodal?"	Norton (2024)
	Is the way of decision support delivery (e.g., alerts, recommendations, clinical pathways) appropriate for the intended healthcare environment and tasks?	Horsky et al. (2012)
	"Is the advice applicable in the setting in which it will be implemented?"	Van de Velde et al. (2018)
	Have you assessed the barriers and facilitators that may influence your healthcare professionals' adherence to the AI-CDSS recommendations (e.g., incentives, trust, organisational context, capacity and	

	resources required for adherence to recommendations)?	
	Do the AI-CDSS and the vendor's technology stack (incl. programming languages, frameworks, and libraries) integrate seamlessly with your IT infrastructure and ecosystem?	IT Convergence (2023)
	Does the system integrate with your EHR system to help ensure data interoperability?	Aidoc (n.d.)
	Does the system align with the national standard or approach to achieving interoperability?	Cresswell et al. (2019)
	"If you want to automatically access the product's internal data, have you considered whether the product has an Application Programming Interface (API)?"	NHSX (2020)
	"Do you use an open-source or closed-model?"	Norton (2024)
	Is it possible to integrate and connect third-party additions (e.g., external software, databases, or services) to the AI-CDSS?	Horsky et al. (2012)
	Can the AI-CDSS be modified to account for the specifics of the healthcare facility?	Horsky et al. (2012)
	Has It been considered how your healthcare staff can work around perceived system limitations?	Cresswell et al. (2019)
	Can preferences of the healthcare professionals and patients be integrated into the AI-CDSS?	Hendriks et al. (2024)
AI Maturity	Do healthcare professionals find the AI-CDSS acceptable for use in their healthcare facility? This includes their • expectation confirmation (incl. system functionality and effectiveness), • satisfaction, • perception of benefits, • perception on job security in relation to the AI, • and intention to use the system.	Nortje & Grobbelaar (2020); Tegenaw et al. (2023)
	Do patients find the AI-CDSS acceptable?	
	Do the leadership team (e.g., medical director) and dedicated AI representatives (e.g., Chief AI Officer) extensively support the adoption of the AI-CDSS?	Pumplun et al. (2021)

	Does your board support the adoption of the AI-CDSS?	Cisco (n.d.)
	Does your middle management support the adoption of the AI-CDSS?	
	Have you considered potential conflicts of interest within your organisation (management and healthcare staff), with the vendor, and other stakeholders?	Scott et al. (2021)
	Has an AI strategy been largely implemented in the healthcare organisation and is it supported by all stakeholders (e.g., clinicians, medical director)?	Cisco (n.d.); Pumplun et al. (2021)
	"Is it clear who / what team is leading the AI strategy for your company or is it being managed in a more organic and decentralized manner?"	Cisco (n.d.)
	"Have realistic [and flexible] deployment plans been set and communicated?" This should include realistic timelines, the allocation of sufficient resources (incl. personnel, funding, equipment), and methods to assess implementation outcomes (e.g., usefulness).	Intel (2018); Riester & Zullo et al. (2023)
	"Has an assessment been carried out to compare and evaluate the disparities between the development environment and the organizational environment where the AI solution will be implemented?"	CHAI (2024)
	"Have you considered starting off with a pilot project with a tightly defined scope and set of success metrics, before scaling?"	NHSX (2020)
	"Do you have a change management plan in place to address the changes brought about by deploying AI technologies?"	Cisco (n.d.)
	Have you established an AI committee consisting of experts from different areas (e.g., clinical practice, data protection, ethics, medical informatics, IT, patient needs) prior to the deployment of the AI-CDSS?	Apfelbacher et al. (2024)
	Do you have a plan for involving all employees impacted by the AI-CDSS in its deployment process?	CluedIn (2024)
	Do you have the IT and AI methods expertise necessary for the adoption and use of the AI-CDSS?	Pumplun et al. (2021)

	(either by having dedicated AI experts in data science and engineering, or by seeking external AI expertise)	
	Have you assessed the need for upskilling to ensure that your healthcare staff has the necessary skills and knowledge to work with the AI-CDSS? This includes awareness of one's role and responsibilities (incl. cross-checking recommendations, responsibility for clinical errors associated with the system), as well as data literacy, technical skills (e.g., interpreting outputs, incorporating AI into workflows), and risk training to provide knowledge of ethical and regulatory considerations (e.g., patient privacy and autonomy).	Cisco (n.d.); Infosys (n.d.); Microsoft (n.d.) Scalefocus (n.d.); Writer (n.d.); Correlation One (2023); Deloitte (2024; CluedIn (2024); Virginia Office of Data Governance and Analytics (2024); Bottacin et al. (2025)
	"Are there data available and methods defined to evaluate whether the AI system is being used as intended by end users and whether variability in end-user behavior impacts treatment or outcomes of specific sociodemographic subgroups, ensuring adherence to intended use and identifying potential biases?"	CHAI (2024)
	Are other members of your workforce (e.g., legal staff, security experts, AI deployment and governance team) also prepared for the adoption and monitoring of the AI-CDSS?	
	Do you have a plan for ongoing education of your workforce, as the AI-CDSS and best practices evolve (e.g., through agreements with the vendor, in-house education, external experts)?	Bottacin et al. (2025)
	Do you have the resources and personnel to ensure continuous oversight and operation of the AI-CDSS, including during periods of staff absence such as holidays?	CNIL (2022)

	Is there a process for collaboration and open communication between AI experts, healthcare professionals, and other staff involved in the procurement, deployment, and use of the AI-CDSS?	Writer (n.d.); CluedIn (2024)
	Have you considered whether involving external partners (e.g., research institutions) would benefit your deployment and use of the AI-CDSS?	Correlation One (2023)
	Are the data assets we currently have sufficient for the AI-CDSS?	Future Processing (2024)
	Do you have the technical infrastructure necessary for the deployment and use of the AI-CDSS? This includes, among others: Readiness from a power consumption perspective and "up-to-date systems deeply integrated in data networks for regular and extensive data exchange within the clinic and external institutions (eg, other clinics and academic institutions)".	Pumplun et al. (2021)
	"If you are not buying into a managed service, do you have the IT capability in-house?"	NHSX (2020)
	"Does your organization have a formal data governance framework with defined roles, responsibilities, and processes?" (Virginia Office of Data Governance and Analytics, 2024) This should include, among others: • "Can you identify the appropriate data stewards and subject matter experts in your organization?" • "Is data ingested, stored and processed for availability?" • "Is data classified for ease of management?" • "Is it easy to find data relevant to the use case?" • "Is data known, trusted and measured?" • "Is data secure and observable throughout the data lifecycle?" (Passerelle, 2024)	Virginia Office of Data Governance and Analytics (2024); Passerelle (2024); NHSX (2020)

	• "Do you have a process to identify and classify critical data assets based on their value to the agency?" (Virginia Office of Data Governance and Analytics, 2024) • If cloud-based servers are used, knowledge of where they are based	
	"Are data ownership, access, and usage clearly defined and communicated throughout the organization?" Is it integrated into the project and change management processes?	Virginia Office of Data Governance and Analytics (2024)
	"Is there a centralized data catalog or repository to document data definitions, metadata, and lineage?"	
	Do you have quality management structures in place for the AI-CDSS?	Nortje & Grobbelaar (2020)
	Do you have knowledge management structures in place to capture, organise, share, and apply knowledge related to the AI-CDSS within your healthcare facility?	
	Will you be able to monitor and react to new potential benefits and risks of the AI-CDSS, as the AI field, best practices, and ethical views evolve? This should include an assessment if the AI governance structures and processes in place align with changing standards.	PDPC et al. (2020a)
	Have you considered if there should be a strategy for scaling the AI-CDSS (e.g., higher patient volume, use across multiple departments of your facility, adaptability to ongoing developments in AI and healthcare)?	
Vendor Reliability, Support, and Agreements	Is the relationship with the vendor effectively managed?	Cresswell et al. (2019)
	How does the vendor support the deployment of the AI-CDSS?	IT Convergence (2023); Norton (2024); Sutten (2025)
	"Will the commercial AI vendor also assist with workflow integration? If so, to what extent?"	Chae et al. (2024)
	What education and training programs and materials does the vendor provide for our staff? "The vendor	California Telehealth

	should provide role-based training programs that not only cover technical aspects but also integrate clinical context and decision-making protocols."	Resource Center (2024); Sutten (2025)
	"Is the commercial AI vendor willing to validate and refine their product to meet minimally acceptable performance standards on test data provided by the purchasing radiology practice before the final purchase?"	Chae et al. (2024)
	Do you have knowledge about the product's time on the market, any known past incidents, other customers, and their satisfaction and experiences?	Fasterholdt et al. (2022)
	Can the vendor share their product roadmap?	Norton (2024)
	Does the vendor have a long-term plan for the maintenance, updating and improvement of the AI-CDSS (incl. revalidation of the system following updates and documentation of updates)? For example: "Can patient data sets and radiologist feedback be fed back into model training for continuous AI learning that evolves with changes in the patient population?" Another example is the correction of model drift.	Scalefocus (n.d.); Crigger et al. (2022); Bottacin et al. (2025)
	If the responsibility for maintaining, updating, and improving the AI-CDSS is shared between the vendor and your organisation, have you clearly defined how these responsibilities will be divided and designated individuals in charge?	Bottacin et al. (2025)
	"What are the vendor's expectations of your organisation sending back data to support their iteration of the model or development of other products? Have you clarified what the vendor means by model iteration and development, and have you ensured that your information governance arrangements address this?"	NHSX (2020)
	Does the vendor have incident detection and response procedures in place?	TrustArc Canada Inc. (2024); Stout (2025)
	Is effective collaboration and communication with the vendor possible?	IT Convergence (2023)

	• "Evaluate the vendor's communication channels, responsiveness, and willingness to engage in ongoing discussions." • "Assess their ability to understand and interpret your requirements, as well as their capability to provide regular updates and progress reports."	
	What is the process for submitting support requests and feedback to the vendor? "Support must be provided through different channels with guaranteed response times like chatbots, service representatives, and a specialized point of contact requiring immediate attention based on urgency levels." (Sutten, 2025)	Norton (2024); Sutten (2025)
	What expertise and experience does the vendor have in AI for healthcare and in the specific area of application of the system (e.g., understanding of patient care workflows, best practices, medical terminology, and data flows in healthcare)? "Can you provide case studies or references from similar healthcare settings to our patient population and settings where your AI tools have been successfully implemented? How was success defined?" (California Telehealth Resource Center, 2024)	IT Convergence (2023); Norton (2024); Sutten (2025)
	Can the vendor easily and with detail explain the AI-CDSS, its technical foundations, and purpose?	Hosch & Morris (2024)
	Can the vendor describe their MLOps/DLOps practices to manage the lifecycle of the AI-CDSS?	You et al. (2025)
	Does the vendor have guarantees for the performance of the AI-CDSS?	Hosch & Morris (2024)
	"What guarantees are provided with respect to software downtime and probability of system failures?"	Chae et al. (2024)
	Have you reviewed the vendor's Terms of Service, warranties, and any other contractual service level agreements? "Don't assume the terms are non-negotiable; be prepared to discuss and modify terms to align with your organization's needs and risk tolerance." Hosch & Morris (2024) provide further (introductory) guidance on contractual agreements.	Hosch & Morris (2024)

	"What provisions are in place for contract termination and handover to another supplier?"	NHSX (2020)
	"To what extent will you be able to publish details of your contract?"	
	"What is your [the vendor's] policy on accountability for errors or negative outcomes resulting from the AI tool?" The accountability for the outcomes of the AI-CDSS needs to be clearly defined between the vendor and you.	California Telehealth Resource Center (2024)
	"What liability coverage do you [the vendor] provide in case of AI tool failure or inaccuracies?"	
	Are indemnity clauses clearly set out in the contract, and does the vendor have insurance coverage to support those liabilities?	NHSX (2020); Hosch & Morris (2024)
	"Is it clear what is considered as product failure versus human error in using the product?"	NHSX (2020)
	Have the ownership (incl. intellectual property rights), usage rights, and licensing terms for both the AI-CDSS and the data been clearly defined and agreed upon?	Stout (2025)
	"Have you clearly documented and justified instances of your organisation talking to / inviting specific vendor/s to bid for the project?"	NHSX (2020)
	"If you are being offered a product for free, what steps have you put in place to ensure that you remain compliant with public procurement guidelines?"	