emerging trends in machine intelligence and big data
ORIENT review is a scientific journal publisher based in Asia, dedicated to disseminating high-quality research across various academic disciplines. With a strong commitment to fostering scholarly advancement and promoting cross-cultural understanding, ORIENT REVIEW has established itself as a reliable platform for researchers, academics, and professionals across the globe.

Human-Centered AI (HCAI) Paradigms in Clinical Artificial Intelligence: An Analytical Discourse on Implementation Across AI Lifecycle Stages

Jatin Pal Singh

abstract

Traditional discussions place humans around AI in decision-making, but the Human-Centered AI (HCAI) approach puts AI around humans, making humans the main focus. This paper presents an analysis of the key aspects of Human-Centered Artificial Intelligence (HCAI) in clinical AI algorithms, examining its implementation throughout the AI lifecycle stages. The focus is on the past literature and practical strategies to embed HCAI principles in each stage, namely Data Creation, Data Acquisition, Model Development, Model Evaluation, and Model Deployment. The research delineates how these stages can be optimized to enhance human performance, support ethical design, and ensure AI solutions are aligned with human values and societal needs. This research argues that in the Data Creation phase, the emphasis lies on inclusive and diverse data collection, ensuring that the data reflects a broad spectrum of demographics. The involvement of clinicians and patients is critical in making the data both comprehensive and reflective of real-world clinical scenarios. Ethical data collection practices are underscored, focusing on transparent consent and privacy safeguards. The integrity and contextual richness of data are also highlighted, stressing the need for accurate, complete, and relevant data collection. The Data Acquisition phase demands adherence to legal and regulatory frameworks, ethical data sourcing, and transparency in the data acquisition processes. The acquisition of diverse and representative data, focusing on quality and relevance, and responsible data partnerships are essential. Special attention is given to sociodemographic factors. In Model Development, the paper advocates for a clear definition of the AI model's purpose and ethical data handling. The model should be transparent, explainable, and designed with human factors in mind. Robustness, generalizability testing, iterative development, compliance with regulatory standards, and addressing societal and ethical concerns are vital. The Model Evaluation phase calls for performance evaluation across diverse populations, integration assessment within clinical workflows, fairness and equity analysis, and user feedback incorporation. Additionally, the model's explainability, robustness, regulatory compliance, and longitudinal performance monitoring are emphasized. In the Model Deployment phase, continuous monitoring of real-world performance, facilitation of understanding among clinicians and patients, ensuring ethical use, adaptation to diverse clinical environments, maintaining data security, responding to feedback, assessing the impact on clinical practice, and compliance with evolving regulations and standards are essential. This paper argues and highlights that the successful implementation of HCAI in clinical AI algorithms requires considering not only technical and scientific aspects but also ethical, legal, and social dimensions.

Keywords: Clinical AI Algorithms, Data Acquisition, Ethical Design, Human-Centered AI (HCAI), Model Development

introduction

Optimizing performance metrics in Artificial Intelligence (AI) has traditionally been the focal point of research and development. (Mämmelä *et al.*, 2018; Roth *et al.*, 2020) showed that key indicators such as accuracy, speed, and resource efficiency dominate the criteria for evaluating AI systems. This focus is reflective of a technical and utilitarian approach, where the primary objective is to enhance the functional capabilities of AI systems. Accuracy, a performance metric, is often quantified by the ability of an AI system to produce correct results, be it in classification tasks, prediction, or decision-making processes. Speed, another vital metric, pertains to the efficiency with which AI systems process and analyze vast amounts of data, delivering outputs in a timely manner. Resource efficiency, meanwhile, addresses the optimization of computational resources, ensuring that AI systems operate effectively without excessive consumption of processing power or energy. These metrics, while crucial for the technological advancement of AI, often overshadow broader considerations beyond technical performance.

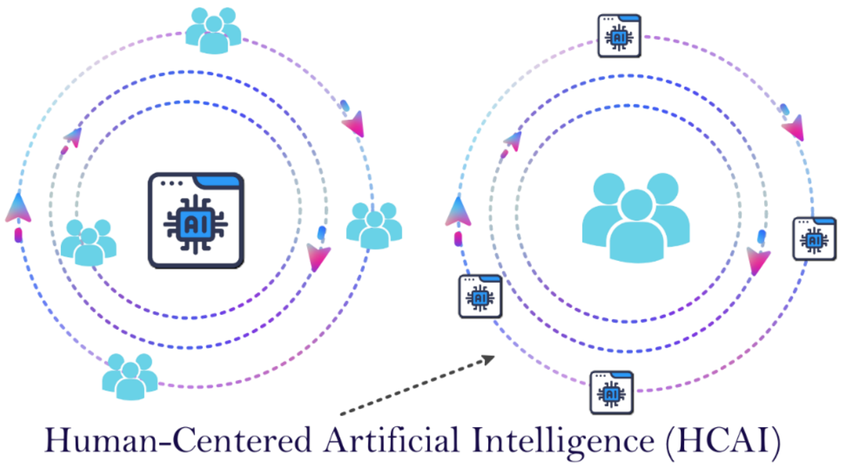

Figure 1. HCAI places humans at the center of systems design thinking. ***Source***: author [*concept from* (B. Shneiderman, 2020)]

In contrast to the traditional approach, Human-Centered Artificial Intelligence (HCAI) represents a paradigm shift, emphasizing the integration of human values and societal impacts in the design, development, and deployment of AI systems. HCAI posits that AI should not merely be an instrument of efficiency and automation but should be aligned with human needs and aspirations. This perspective advocates for a more holistic approach to AI, considering the ethical, social, and cultural dimensions of technology. For instance, HCAI emphasizes the importance of fairness in AI systems, ensuring that they do not perpetuate or amplify existing societal biases. It also highlights the need for transparency and explainability in AI, enabling users to understand and trust AI decisions. HCAI stresses the importance of user-centric design, ensuring that AI systems are accessible, usable, and beneficial to diverse groups of people. By incorporating these human-centric considerations, HCAI seeks to foster AI systems that are not only technically proficient but also ethically sound and socially responsible.

The emergence of HCAI as a response to prevalent practices in AI development shows a growing awareness of the limitations of a purely performance-oriented approach. While optimizing for accuracy, speed, and resource efficiency is undeniably important, focusing solely on these aspects can lead to the neglect of crucial human and societal dimensions. For instance, an AI system optimized for speed and accuracy might be highly effective in a specific task but could inadvertently exacerbate societal inequalities or compromise individual privacy. HCAI challenges this narrow focus, advocating for a more balanced approach that considers the impacts of AI on society. It represents a call to reevaluate the priorities and goals of AI development, urging researchers, developers, and policymakers to integrate human values and societal well-being into the core of AI innovation. This shift towards HCAI is not merely a theoretical exercise but a practical necessity, as the increasing ubiquity of AI in various domains of life demands a more thoughtful and responsible approach to its development and deployment.

The most widely used definition for HCAI is the one developed by Ben Shneiderman. This definition reads as: "*HCAI focuses on amplifying, augmenting, and enhancing human performance in ways that make systems reliable, safe, and trustworthy. These systems also support human self-efficacy, encourage creativity, clarify responsibility, and facilitate social participation*" (Ben Shneiderman, 2020).

HCAI's focus on "*amplifying, augmenting, and enhancing human performance*" marks a paradigm shift in AI development. This emphasis on enhancing human capabilities through technology illustrates a collaborative synergy between human intelligence and artificial systems. However, the definition's focus on "*making systems reliable, safe, and trustworthy*" might not fully encapsulate the complexity of human-AI interaction. Trustworthiness in AI, while crucial, extends into ethical realms, necessitating transparency and accountability in AI decision-making processes. Yet, this perspective could be seen as somewhat narrow, potentially overlooking broader societal impacts. Issues such as the digital divide, privacy concerns (Saxena, 2020), and potential job displacement due to AI advancements are considerations that seem underrepresented in this definition.

Supporting "*human self-efficacy*" and "*encouraging creativity*" positions HCAI as a tool for enhancing human agency and innovation. These aspects suggest a vision of AI that transcends mere computational efficiency, aiming to enrich human creativity and problem-solving abilities. However, "*clarifying responsibility*" in AI usage presents a complex challenge. Determining accountability in decisions made by AI systems is fraught with ethical dilemmas, especially in scenarios where AI significantly influences or autonomously makes decisions. Furthermore, "*facilitating social participation*" through AI implies an inclusive approach, yet it raises questions about accessibility and equity in the context of AI technologies. Ensuring that AI systems are accessible and beneficial to diverse populations is imperative, but achieving this in practice is a formidable task, given the existing disparities in technology access and literacy.

Although the ultimate goal of building fully intelligent entities remains elusive, the age of AI is already impacting humanity in ways that are substantial yet not well understood. In the recent past, various scientific disciplines including physics and chemistry had to reckon with the societal consequences of their scientific advances when these advancements migrated from conference discussion or a laboratory experiment into wide adoption by industry. In a similar manner, now is the time for the scientific community to grapple with the societal consequences and potential changes to the human condition resulting from the adoption of current AI systems.

The opportunity of artificial intelligence (AI) to reduce health care disparities and inequities is recognized, but it can also exacerbate these issues if not implemented in an equitable manner. This perspective identifies potential biases in each stage of the AI life cycle, including data collection, annotation, machine learning model development, evaluation, deployment, operationalization, monitoring, and feedback integration.

implementation across ai lifecycle stages in clinical artificial intelligence

The AI life cycle clinical AI algorithms, have five phases:

1. **Data Creation**: This phase involves understanding the origins of data used in training and evaluating clinical AI algorithms. Data sources can range from routine clinician-patient interactions to data from wearable devices and clinical research trials. Issues of data diversity and representativeness are pivotal here, highlighting the need for context disclosure including patient demographics and other characteristics.
2. **Data Acquisition**: This stage is concerned with obtaining and consolidating data for AI model development. Challenges include navigating legal and regulatory frameworks like HIPAA, GDPR, or CCPA, which can lead to the omission of sociodemographic or clinical information, impacting the ethical nature of the AI algorithms. Transparency in financial arrangements, privacy protections, and patient consent in data acquisition is crucial for ethical development.
3. **Model Development**: This process demands a clear definition of the AI algorithm's purpose, including identifying the problem it aims to solve, the end-user, and the beneficiary. This phase is heavily influenced by the data available for training and involves decisions on data pre-processing, model architectures, and

hyperparameters. Issues of data labeling, establishing a ground truth, and the potential for embedded biases are significant. Transparency in model development processes and the requirement for open-source code and public data availability are vital for replicability and ethical consideration.

4. **Model Evaluation**: The objective here is to assess the performance of the AI algorithm in its intended environment. Evaluation methodologies now include fairness metrics, but challenges persist, such as limited evaluations that do not fully integrate the tool into clinical workflows. Generalizability and calibration across different populations are essential, with a focus on assessing performance across diverse groups to ensure equitable outcomes.
5. **Model Deployment**: This final phase involves integrating the AI algorithm into clinical settings to improve patient care and outcomes. Key concerns include understanding the algorithm's performance across populations, its interpretation by clinicians, and its overall clinical utility. Transparency, generalizability, fairness, and accessibility remain critical. The potential for AI algorithms to create biased clinical data, which then feeds into future models, highlights the interconnected nature of the AI life cycle stages and the amplification of biases.

Table 1. definitions of human-centered artificial intelligence in literature	*reference*
HCAI focuses on amplifying, augmenting, and enhancing human performance in ways that make systems reliable, safe, and trustworthy. These systems also support human self-efficacy, encourage creativity, clarify responsibility, and facilitate social participation.	(Ben Shneiderman, 2020)
Human-centered means that a system should always consider the human partner in its deliberations. Tasks of the AI system should be performed for someone, in some context, and if the actions of the AI system affect people directly or indirectly, this should be considered in its decision-making.	(Dignum and Dignum, 2020)
Human-centered AI needs to focus on three integrated perspectives when designing AI systems: rationalistic (technology), humanistic (people), and judicial (policies).	(Riedl, 2019)
Human-centered AI is defined as a synergistic approach to align AI solutions with human values, ethical principles, and legal requirements to ensure safety and security, enabling trustworthy AI.	(Holzinger *et al.*, 2022)
HAI includes three main components: 1) ethically aligned design, which creates AI solutions that avoid discrimination, maintain fairness and justice, and do not replace humans; 2) technology that fully reflects human intelligence, enhancing AI technology to more closely resemble human intelligence; and 3) human factors design to ensure that AI solutions are explainable, comprehensible, useful, and usable.	(Xu, 2019)

Table 2. Key aspects of Human-Centered Artificial Intelligence HCAI in past literature

Key Aspect	**Definition Focus**	**Details**
Amplification of Human Performance	HCAI as a Performance Enhancer	Emphasis on amplifying, augmenting, and enhancing human capabilities, ensuring system reliability, safety, and trustworthiness.
Supportive Design	Systems Design	Focus on supporting human self-efficacy, creativity, responsibility clarity, and social participation.
Ethically Aligned Design	Ethical AI Solutions	Creating AI solutions that avoid discrimination, ensure fairness and justice, and complement rather than replace human roles.
Intelligence Reflection	Technology Enhancement	Enhancing AI technology to more closely resemble human intelligence.
Human Factors Design	AI Solution Usability	Ensuring AI solutions are explainable, comprehensible, useful, and usable.
Societal Augmentation	Addressing Societal Needs	AI aiming to augment human abilities and address societal needs, drawing inspiration from human beings.
Systemic Perspective	Larger System Awareness	Recognizing AI as part of a system including human stakeholders like users, operators, clients.

Integrated Perspectives	Rationalistic, Humanistic, Judicial Focus	Focusing on technology, people, and policies in AI system design.
Deliberation Consideration	Human-System Interaction	Systems considering human partners in decision-making, especially in tasks affecting people directly or indirectly.
Service Orientation	Commitment to Humanity	AI systems being human-centric, serving humanity and the common good, aiming to improve welfare and freedom.
Synergistic Approach	Alignment with Human Values	Aligning AI solutions with human values, ethical principles, and legal requirements for safety, security, and trustworthiness.

Implementing Human-Centered Artificial Intelligence (HCAI) in the data creation phase

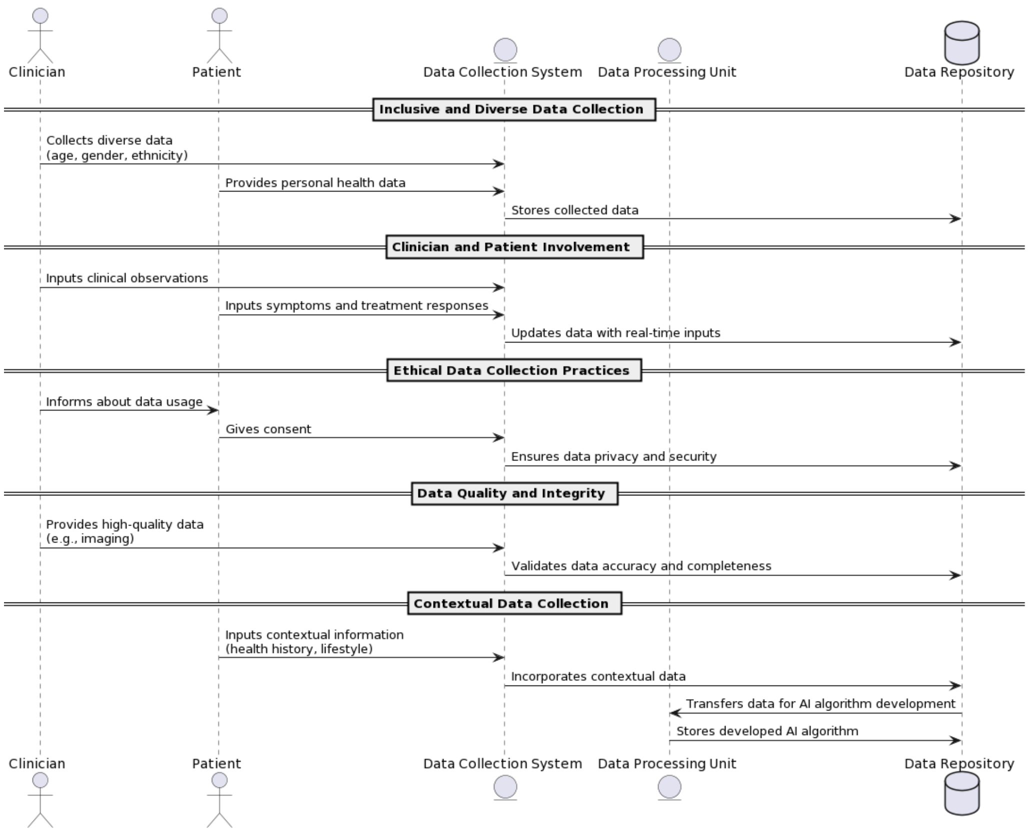

Fig 2. The diagram shows the sequence of interactions for each aspect of the HCAI implementation in the **data creation phase**. The actions include data collection, validation, and processing, emphasizing the key roles of clinicians and patients in providing diverse, ethical, high-quality, and contextual data. The final step shows the data being transferred to a data processing unit for the development of the clinical AI algorithm, which is then stored back in the data repository. ***Source***: author

Implementing Human-Centered Artificial Intelligence (HCAI) in the data creation phase of clinical AI algorithms necessitates an approach, focusing on several aspects to ensure the effectiveness and ethical integrity of these systems. One of the primary considerations is the Inclusive and Diverse Data Collection. This aspect suggests the imperative of incorporating a broad spectrum of demographics into the dataset. Such diversity spans across various dimensions, including but not limited to age, gender, ethnicity, and socio-economic backgrounds. This approach is instrumental in mitigating biases that are often inherent in AI algorithms, thereby fostering a more accurate representation of the population at large. Consider, for example, the development of a cardiology AI tool. It is imperative to integrate data from diverse genders, multiple age groups, and various ethnic backgrounds. This inclusivity is essential to enhance the tool's diagnostic accuracy and efficacy across a diverse patient demographic.

Another aspect in the implementation of HCAI in clinical AI is the Clinician and Patient Involvement in the data creation process. This involvement is important in ensuring the richness and comprehensiveness of the data, as well as its alignment with the realities of clinical settings. Clinician and patient input can be operationalized through various mechanisms, such as digital tools that facilitate the recording of clinical observations and patient feedback. For instance, deploying a mobile application that enables patients to input data regarding their symptoms and responses to treatments can be highly beneficial. This approach facilitates the generation of patient-centered data that is reflective of real-time health experiences. Such data not only enriches the AI algorithm with a wide range of health data inputs but also ensures that the algorithm is grounded in the actual experiences and needs of patients.

Ethical Data Collection Practices encompasses ensuring informed consent and maintaining privacy and confidentiality of patient data (Khanna and Srivastava, 2020). The process of data collection must be characterized by transparency, where patients are fully informed about the utilization of their data. For example, in the context of developing an AI-based diagnostic tool, it is essential to communicate to patients the specific purposes for which their data will be utilized, along with assurances regarding the confidentiality and security of their data. Additionally, Data Quality and Integrity must be emphasized, focusing on the accuracy, completeness, and relevance of the data. For instance, in the collection of imaging data for an AI algorithm in radiology, it is necessary to ensure that the images are high-quality, accurately labeled, and pertinent to the conditions the AI is intended to identify. Contextual Data Collection is vital, wherein data collection should extend beyond mere raw data to include information about the circumstances under which the data was collected. This may encompass patient health history, lifestyle factors, and environmental conditions, providing a rich context that is indispensable for the development of AI algorithms capable of precise predictions.

Implementing Human-Centered Artificial Intelligence (HCAI) in the data acquisition phase

The incorporation of Human-Centered Artificial Intelligence (HCAI) principles into the Data Acquisition phase of the clinical AI algorithm lifecycle necessitates the implementation of strategies that foreground ethical, legal, and human-centric considerations, ensuring the responsible and beneficial use of AI in healthcare.

A primary strategy in this phase is the Compliance with Legal and Regulatory Frameworks. This necessitates strict adherence to data protection and privacy laws such as HIPAA (Health Insurance Portability and Accountability Act), GDPR (General Data Protection Regulation) (Carlson *et al.*, 2020), and CCPA (California Consumer Privacy Act). In practical terms, when acquiring patient data, organizations are required to establish protocols that guarantee the anonymization or de-identification of data where necessary and safeguard it against unauthorized access. A component of this is obtaining informed consent from patients, clearly elucidating the intentions and methodologies of using their data in AI development. For instance, in the context of clinical AI, it is imperative that when patient data is acquired, the process conforms to these legal standards, ensuring patient privacy and data integrity.

Ethical Data Sourcing extends the scope of consideration beyond mere legal compliance. It involves acquiring data in ways that respect patient rights and adhere to societal norms. When sourcing data from electronic health records (EHRs), it is ought to contemplate the ethical implications of utilizing patient data. This encompasses assessing the potential impact on patient privacy and maintaining transparency with patients about the utilization of their data.

This approach is not only ethical but also reinforces the trust and cooperation of patients, which is fundamental for the success of AI in healthcare.

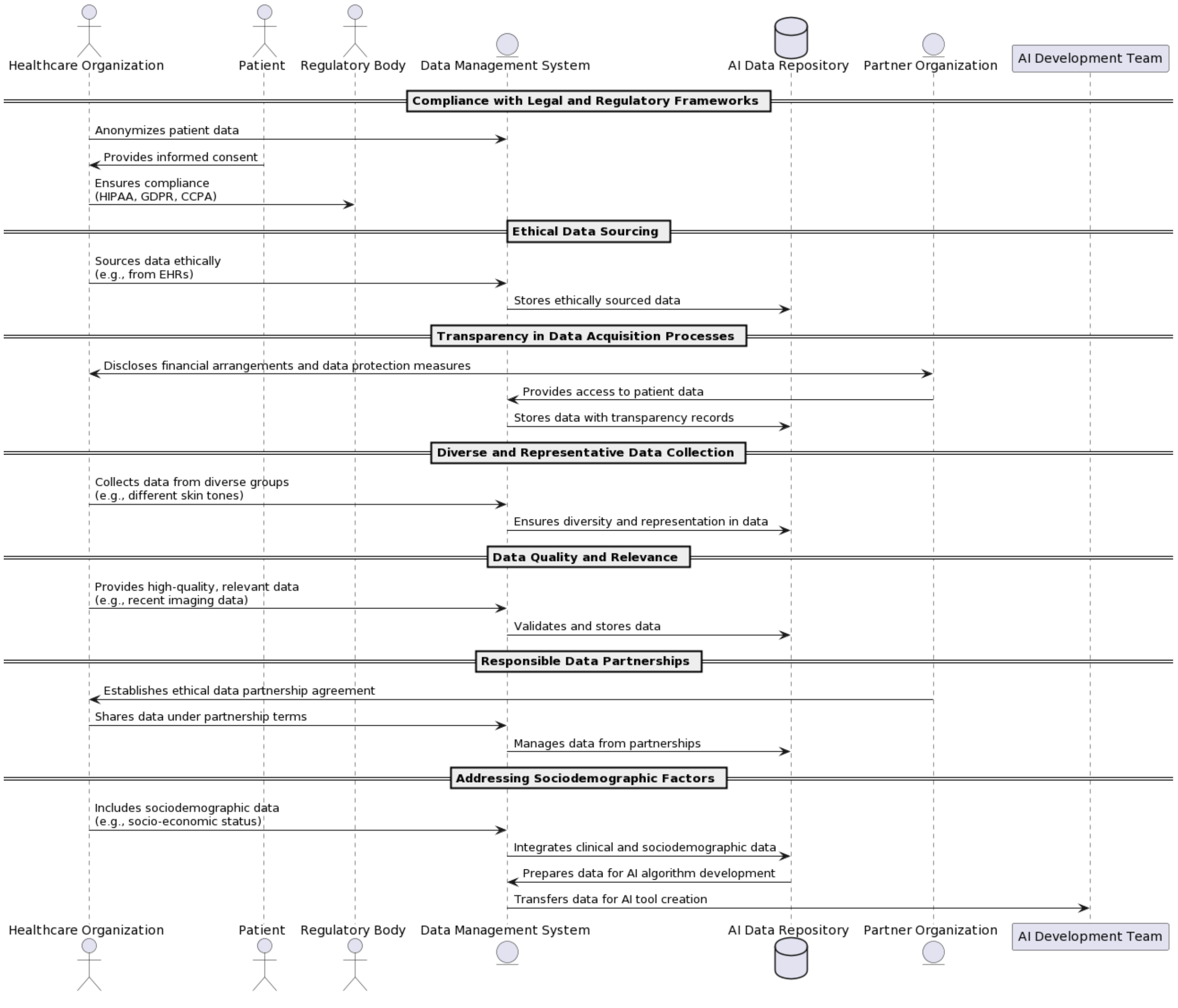

Fig 3. This diagram details the sequence of interactions across the various HCAI strategies in the Data Acquisition phase. Starting from compliance with legal frameworks, it proceeds through ethical sourcing, transparency, diversity, data quality, responsible partnerships, and consideration of sociodemographic factors. The final step illustrates the preparation of data for the development of the clinical AI tool. ***Source***: author.

Transparency in Data Acquisition Processes is instrumental in building trust and ensuring the ethical procurement of data. This involves clear communication regarding any financial arrangements involved in data acquisition and stringent measures to protect patient privacy. For instance, in cases where data is acquired through partnerships with other organizations, it is imperative that both parties disclose their roles, responsibilities, and benefits derived from the data acquisition. This transparency is key to maintaining ethical standards and public trust in the use of AI in healthcare.

Diverse and Representative Data Collection involves actively seeking data sources that encompass underrepresented groups, ensuring that the AI tool is effective and equitable across the entire population it serves. For example, in the development of an AI tool for diagnosing skin conditions, including a diverse range of skin tones is vital to ensure the tool's effectiveness across different ethnicities. Data Quality and Relevance should be emphasized during the acquisition phase. This entails ensuring that the data is accurate, current, and directly applicable to the healthcare context for which the AI tool is being developed. In the acquisition of imaging data for an AI model in oncology, for instance, it is vital to ensure that the images are recent and pertinent to the types of cancer the model is designed to detect. When acquiring data through partnerships with other entities, such as research institutions or healthcare providers, it is important to forge agreements that adhere to the principles of HCAI. This includes ensuring that partners are committed to ethical data practices and that the data exchange is beneficial to all parties involved, especially the patients whose data is being utilized. Addressing Sociodemographic Factors involves not only the collection of clinical data but also the consideration of factors like socio-economic status, education, and geographic location, which can significantly influence health outcomes. For instance, in acquiring data for a heart disease prediction tool, including socio-economic data can provide insights into risk factors that extend beyond purely medical considerations, thereby enhancing the accuracy and applicability of the predictions.

Implementing Human-Centered Artificial Intelligence (HCAI) in the model development phase

The integration of Human-Centered Artificial Intelligence (HCAI) principles into the Model Development phase of clinical AI algorithms requires a careful approach, focusing on strategies that ensure the development of AI models that are ethically sound, effective, and centered around the needs of users. Commencing with Purpose and Problem Definition, it is imperative to articulate clearly the objectives of the AI model. This involves identifying the specific clinical issue the AI is designed to address, pinpointing the intended end-users (such as clinicians, patients, or healthcare administrators), and elucidating the direct beneficiaries of the AI application. For example, in the development of an AI model for diagnosing diabetic retinopathy, it is essential to define precisely the clinical necessity, identify the target user group (e.g., ophthalmologists), and delineate how the model will enhance patient care.

Ethical Data Pre-processing and Labeling is another strategy in this phase. The model development process must ensure ethical handling of data, particularly in its pre-processing and labeling stages. This includes employing unbiased methods to label data and avoiding the introduction of human biases. For instance, when labeling radiology images, it is paramount to engage multiple radiologists to reduce individual bias and ensure a comprehensive understanding of the data.

Transparent and Explainable AI Design is essential for the development of a user-friendly AI model. The AI model should be designed to be transparent and explainable, aiding users in understanding how the model arrives at its decisions. In the case of an AI tool utilized in predictive diagnostics, this might entail incorporating features that enable clinicians to understand the factors influencing a model's prediction, thereby fostering trust and usability.

Incorporating Human Factors into the model design is used for ensuring usability and a positive user experience, particularly for clinicians or healthcare staff who will interact with the AI. This entails creating interfaces that are intuitive and present relevant clinical information in an accessible format. For instance, an AI model designed for patient triage should integrate effortlessly into existing clinical workflows and offer decision support in a clear, actionable manner. Robustness and Generalizability Testing is vital to ascertain that the model performs effectively across diverse populations and settings. This requires testing the model on a broad array of datasets to check for biases and confirm its applicability to different patient groups. For a clinical AI tool utilized in cardiology, testing the model across varied age groups, genders, and ethnicities is important to ensure its accuracy and reliability.

Iterative Development and Feedback Integration is a key aspect of the model development process. This approach allows for continuous refinement of the model based on feedback from end-users and stakeholders. It could involve pilot testing the model in a clinical setting and integrating feedback from clinicians and patients to enhance its performance and usability. Compliance with Regulatory Standards includes ensuring that the model complies with the requirements for medical devices or clinical decision support tools as stipulated by relevant authorities like the FDA (Food and Drug Administration) or EMA (European Medicines Agency). For example, an AI model for cancer detection must meet the regulatory standards for accuracy, safety, and reliability. Addressing Societal and Ethical Concerns is an integral part of the model development process. This involves considering the potential impact of the AI tool on healthcare disparities and ensuring that the model does not exacerbate existing inequalities. It is essential to contemplate the broader societal and ethical implications of the AI tool, ensuring it contributes positively to healthcare outcomes without contributing to the amplification of disparities.

Implementing Human-Centered Artificial Intelligence (HCAI) in the model development phase

In the Model Evaluation phase of clinical AI algorithm development, the incorporation of Human-Centered Artificial Intelligence (HCAI) principles ensures comprehensive, ethical, and effective evaluation of AI models. This phase involves a series of specialized strategies:

The first strategy, Performance Evaluation Across Diverse Populations, entails rigorous testing of the AI model's performance among various demographic groups to guarantee its effectiveness and impartiality. For example, a model developed for diagnosing skin diseases must be evaluated on an extensive range of skin tones. This diversity in testing is fundamental to prevent biases and ensure accuracy for all patient groups, thereby promoting equitable healthcare outcomes.

Clinical Workflow Integration Assessment is another pivotal strategy. This involves evaluating how seamlessly the AI model integrates into existing clinical workflows. The assessment can be conducted through trials in real clinical settings, where the interactions between clinicians and the model are observed, and the impact on clinical efficiency is measured. For instance, an AI tool designed for emergency room triage should be evaluated for its usability in high-pressure situations and its ability to enhance patient flow, thereby contributing to improved healthcare delivery.

Fairness and Equity Analysis ensures that the model does not perpetuate or exacerbate existing healthcare disparities. This strategy includes analyzing the model's predictions for biases based on race, gender, socioeconomic status, and other sociodemographic factors. For a predictive model in cardiology, it is essential to assess whether the model performs equally well for patients from diverse ethnic backgrounds, thereby ensuring fairness and equity in healthcare services.

Incorporating User Feedback is essential. Gathering and integrating feedback from end-users, such as clinicians, during the evaluation phase, provides insights. This might involve conducting structured interviews, surveys, or usability tests to assess the model's practicality, user interface, and overall utility. For example, feedback from nurses and doctors using an AI-based patient monitoring tool would be instrumental in assessing its efficacy in real-world clinical settings.

Explainability and Transparency Assessment is vital for user trust and understanding. Evaluating the model's explainability ensures that clinicians and patients can comprehend the basis of its predictions or recommendations.

ORIENT Review International
ORIENT review is a scientific journal publisher based in Asia, dedicated to disseminating high-quality research across various academic disciplines. With a strong commitment to fostering scholarly advancement and promoting cross-cultural understanding, ORIENT REVIEW has established itself as a reliable platform for researchers, academics, and professionals across the globe.

For an AI model used in diagnostic imaging, it is imperative that the model provides explanations for its findings in a manner that is accessible and comprehensible to radiologists and other medical professionals.

Fig 4. This diagram illustrates the sequential steps and interactions between the data scientist, clinicians, regulatory authorities, end-users, and the AI model itself during the Model Development phase. Each stage, from defining the purpose and problem to addressing societal and ethical concerns, is represented to show the flow of information and feedback necessary for developing a human-centered AI model. ***Source***: author

Robustness and Reliability Testing involves subjecting the model to extensive testing under various clinical scenarios. This includes stress-testing the model in atypical or rare clinical conditions to ensure its accuracy and consistency. For example, a sepsis prediction model must be tested across a broad spectrum of patient conditions and comorbidities to evaluate its reliability and robustness in diverse clinical situations. Regulatory Compliance

Verification is a key component of the evaluation process. Ensuring that the model meets all regulatory requirements for clinical AI tools is crucial. This encompasses compliance with standards for medical devices, data protection, and patient safety. For instance, a model developed for automated drug dosing must adhere to stringent safety and efficacy standards as set by regulatory authorities.

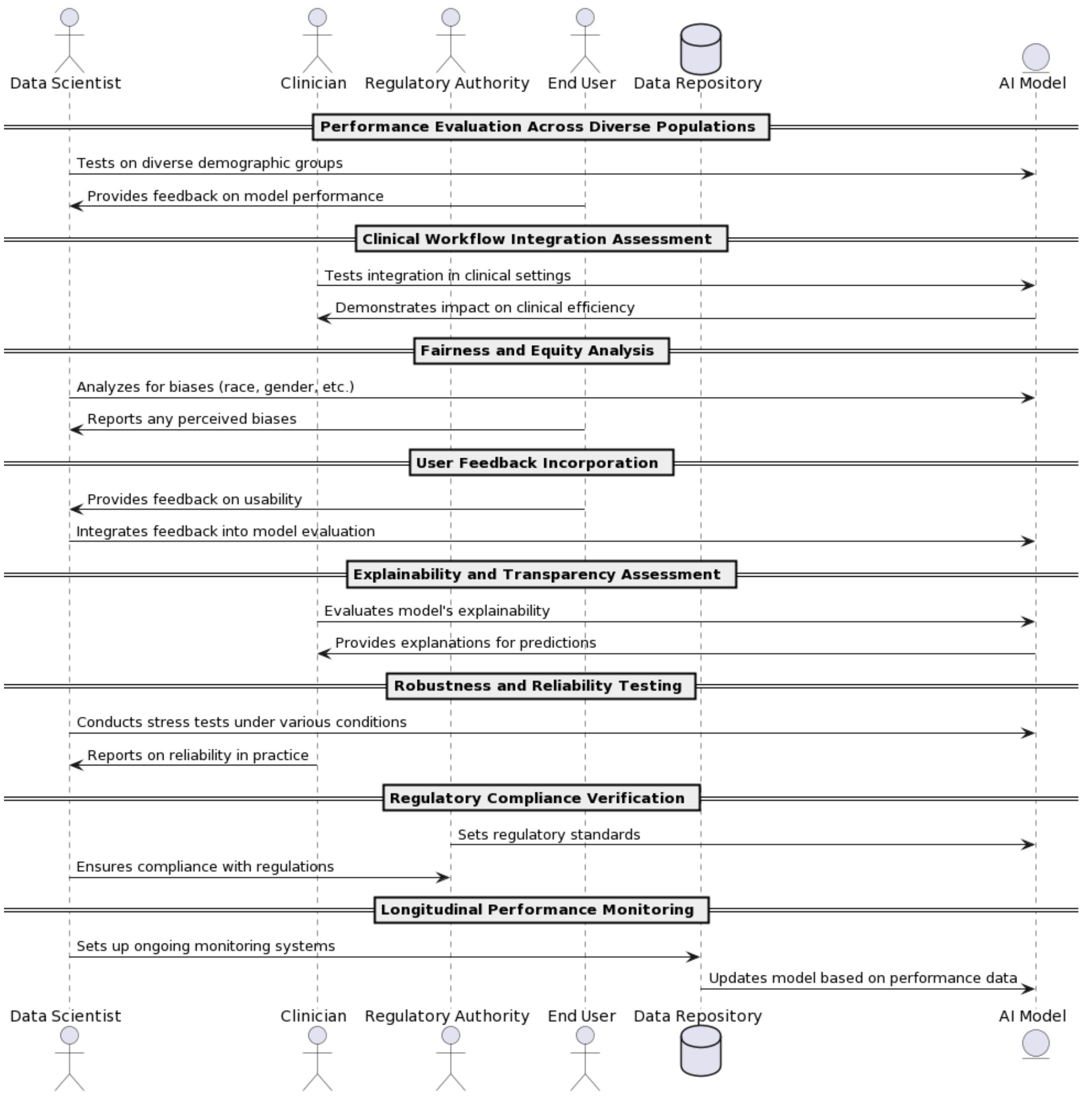

Fig 5. This sequence diagram outlines the key steps and interactions in the Model Evaluation phase, highlighting the involvement of data scientists, clinicians, end-users, and regulatory authorities. Each stage from performance evaluation across diverse populations to longitudinal performance monitoring is depicted to show the approach required in evaluating clinical AI algorithms according to HCAI principles. ***Source***: author

Longitudinal Performance Monitoring is essential to ensure the model's sustained accuracy and effectiveness over time. This involves establishing systems for ongoing monitoring of the model's predictions and updating the model as necessary to adapt to evolving clinical practices and patient populations. For an AI tool used in chronic disease management, continual evaluation is for ensuring its adaptability and relevance in changing healthcare environments.

Implementing Human-Centered Artificial Intelligence (HCAI) in the model development phase

In the Model Deployment phase of clinical AI algorithm development, the integration of Human-Centered Artificial Intelligence (HCAI) principles is indispensable for ensuring the successful, ethical, and effective utilization of these models in clinical settings.

Continuous monitoring of the AI model's performance in actual clinical environments is essential. This process entails tracking the accuracy, effectiveness, and any unintended consequences of the model. For example, in the deployment of an AI system for patient risk stratification in a hospital setting, it is vital to regularly evaluate its impact on patient outcomes and discern any disparities in its effectiveness across various patient groups. Such ongoing evaluation is vital for identifying areas of improvement and ensuring the model's sustained relevance and effectiveness. Facilitating Clinician and Patient Understanding is vital to the successful deployment of AI models. Both clinicians and patients must have a clear understanding of the AI model's functionality and limitations. This may involve conducting training sessions for clinicians and providing clear, accessible explanations for patients. For an AI diagnostic tool, it is important to elucidate how the tool assists in diagnosis, its accuracy rates, and how its findings should be interpreted in the broader clinical context. Such understanding is for the appropriate application of AI tools and for maintaining trust in AI-assisted clinical decision-making.

Ensuring Ethical Use and Preventing Misuse involves deploying guidelines and safeguards to guarantee that the AI model is used ethically and as intended. This includes measures to avert misuse or over-reliance on the AI system. For instance, in the deployment of an AI model for predicting medication dosages, checks should be in place to ensure that final dosing decisions are always reviewed and authorized by a qualified healthcare professional.

Adapting to Diverse Clinical Environments is another significant strategy. The AI model should be adaptable to various clinical environments and workflows. This might involve customizing the user interface or functionality to suit different settings. For an AI-powered patient triage system, ensuring that the system integrates seamlessly into the diverse operational workflows of various healthcare facilities is imperative for its effectiveness and utility. Ensuring Data Security and Patient Privacy remains paramount even after deployment. Maintaining the highest standards of data security and patient privacy involves regularly updating security protocols and ensuring compliance with healthcare privacy laws. For all deployed AI models in healthcare, robust data encryption and access control measures are indispensable for protecting sensitive patient data.

Responding to Feedback and Making Iterative Improvements phase should be viewed as an ongoing process, where feedback from users is continuously gathered and utilized to make iterative improvements to the AI system. For instance, feedback from clinicians using an AI-based diagnostic support tool could inform enhancements in its user interface or the incorporation of new features to augment its utility. Assessing Impact on Clinical Practice and Patient Outcomes is essential for understanding the real-world implications of deploying AI models. This involves analyzing changes in clinical workflows, patient satisfaction, and overall health outcomes. For example, the deployment of an AI system for managing chronic diseases should be evaluated for its efficacy in improving patient adherence to treatment and reducing hospital readmissions.

The deployed AI model must be updated as necessary to remain compliant with new guidelines or laws. For example, changes in medical device regulations might necessitate updates or re-certification of the AI model, ensuring its continual alignment with regulatory standards.

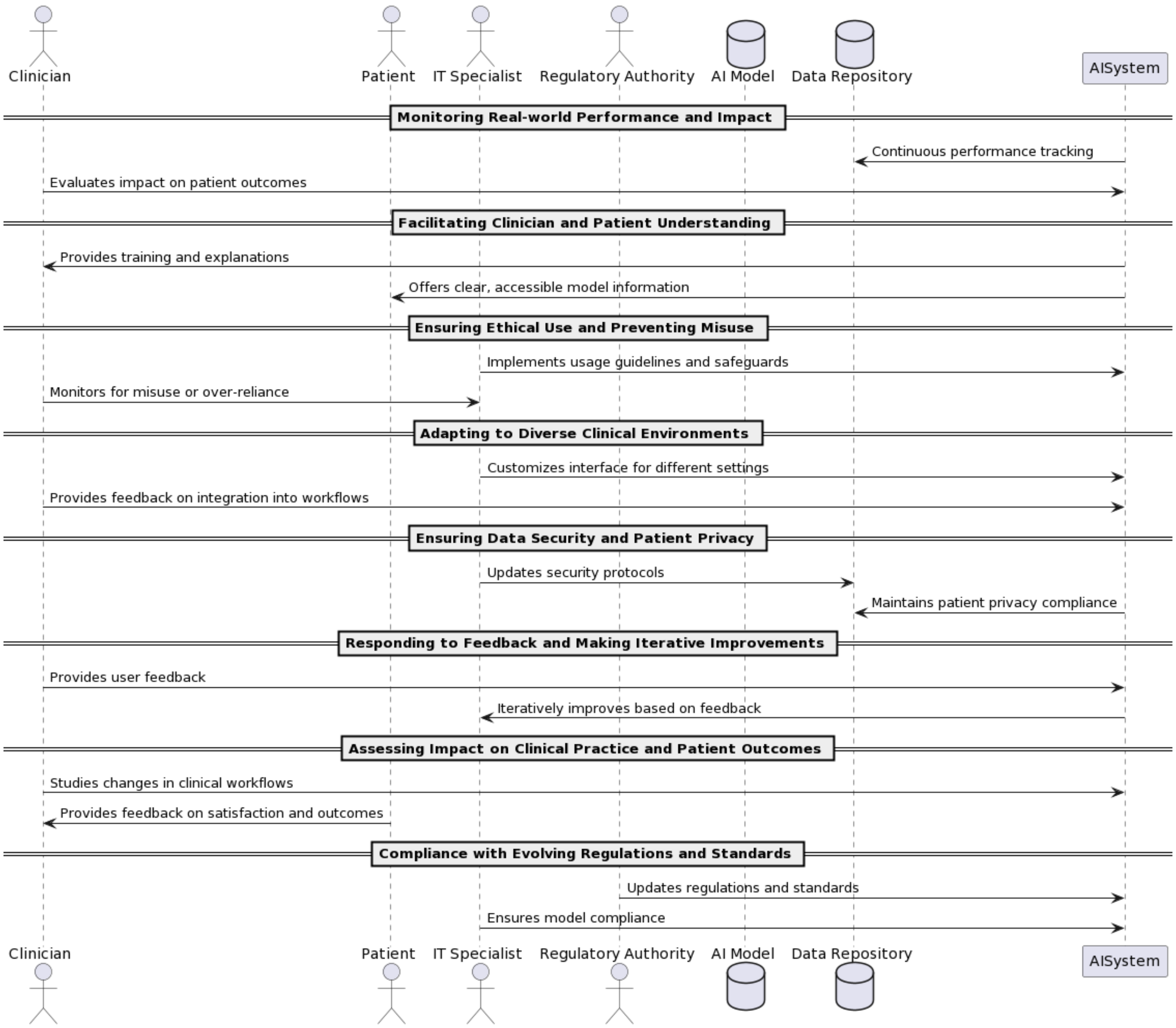

Fig 6. This diagram outlines the sequential steps and interactions between clinicians, patients, IT specialists, regulatory authorities, and the AI system itself during the Model Deployment phase. Each stage, from monitoring real-world performance to compliance with evolving regulations, is depicted to show the approach required in deploying clinical AI algorithms according to HCAI principles. ***Source***: author

conclusion

Human-Centered Artificial Intelligence (HCAI) has been a focal point in past literature, emphasizing the enhancement of human performance. Researchers have rigorously explored how HCAI acts as a performance enhancer, underscoring its role in amplifying, augmenting, and enhancing human capabilities. The literature stresses the importance of system reliability, safety, and trustworthiness in this context. System design in HCAI receives considerable attention, with a clear focus on supporting human self-efficacy, creativity, responsibility clarity, and social participation. Additionally, the concept of Ethically Aligned Design is prevalent, advocating for AI solutions that avoid discrimination, ensure fairness and justice, and complement rather than replace human roles. Intelligence reflection, a key aspect, involves enhancing AI technology to more closely resemble human intelligence. Moreover, the design of AI solutions prioritizes usability, aiming to make AI solutions explainable, comprehensible, useful, and usable. HCAI literature also discusses the augmentation of societal needs, drawing inspiration from human beings to address these needs.

The AI life cycle in clinical AI algorithms incorporates five phases. Data Creation is the initial phase, focusing on the origins of data used in training and evaluating clinical AI algorithms. This phase emphasizes the importance of data diversity and representativeness, advocating for comprehensive context disclosure including patient demographics. The subsequent phase, Data Acquisition, involves obtaining and consolidating data for AI model development, highlighting challenges like navigating legal and regulatory frameworks which impact the ethical nature of AI algorithms. Transparency in financial arrangements and patient consent in data acquisition is underscored. Model Development, the third phase, requires a clear definition of the AI algorithm's purpose, heavily influenced by data availability and involving decisions on data pre-processing, model architectures, and hyperparameters. This phase addresses issues such as data labeling and embedded biases. Model Evaluation assesses the AI algorithm's performance in its intended environment, focusing on fairness metrics and generalizability across different populations. The final phase, Model Deployment, integrates the AI algorithm into clinical settings, with key concerns including the algorithm's performance across populations, interpretation by clinicians, and overall clinical utility. This phase underscores the importance of transparency, generalizability, fairness, and accessibility, highlighting the potential for AI algorithms to create biased clinical data.

The implementation of Human-Centered Artificial Intelligence (HCAI) in the data creation phase of clinical AI algorithms necessitates a multi-faceted strategy, focusing on inclusive and diverse data collection, active clinician and patient involvement, ethical data practices, emphasis on data quality and integrity, and the incorporation of contextual data. This approach ensures the development of AI algorithms that are representative, reliable, and grounded in the complexities of real-world clinical scenarios. For instance, in cardiology, including a broad spectrum of demographics in data collection enhances the algorithm's accuracy across diverse patient populations. Similarly, integrating clinician observations and patient feedback into the data creation process enriches the data set with practical insights, enhancing the AI's applicability in clinical settings. Furthermore, prioritizing ethical considerations, such as informed consent and data privacy, fosters trust and transparency in AI development, while focusing on data quality ensures the clinical relevance and reliability of the algorithms. The inclusion of contextual data, such as patient health history and lifestyle factors, allows for the development of more nuanced and precise AI models, capable of making informed predictions and recommendations in healthcare.

Incorporating HCAI principles in the Data Acquisition phase requires adherence to legal and regulatory frameworks, ethical data sourcing, transparency in data acquisition processes, diverse and representative data collection, emphasis on data quality and relevance, responsible data partnerships, and attention to sociodemographic factors. Compliance with regulations like HIPAA, GDPR, and CCPA ensures the protection of patient data and privacy, while ethical sourcing respects patient rights and societal norms. Transparency in data acquisition builds trust and ensures ethical practices, while diverse data collection enhances the representativeness of AI algorithms. Responsible data partnerships emphasize the importance of ethical practices in data exchange, and addressing sociodemographic factors ensures the inclusion of diverse health determinants in AI development. For example, in developing a diagnostic tool for skin conditions, including a variety of skin tones is essential to ensure the tool's effectiveness across different ethnicities, while ensuring the data is current and relevant to the specific healthcare context.

The Model Development phase integrates HCAI principles through clear problem definition, ethical data pre-processing and labeling, transparent and explainable AI design, incorporation of human factors, robustness and generalizability testing, iterative development with feedback integration, compliance with regulatory standards, and addressing societal and ethical concerns. Defining the AI model's purpose and target user group ensures its relevance and applicability in clinical settings. Ethical handling of data in pre-processing and labeling prevents the introduction of biases. Developing AI models with transparency and explainability enhances user trust and understanding. Considering human factors in model design improves usability and clinical integration. Testing the model across diverse populations ensures its generalizability and unbiased performance. Iterative development allows for continuous refinement based on end-user feedback. Compliance with healthcare regulations ensures the model's safety and efficacy. Addressing broader societal and ethical concerns ensures that the AI tool does not exacerbate healthcare disparities and is aligned with societal values. For example, in developing an AI model for diabetic retinopathy, involving multiple ophthalmologists in the labeling process can minimize bias, while pilot testing in clinical settings can provide valuable feedback for model refinement.

The Model Evaluation phase in the development of clinical AI algorithms, under the umbrella of Human-Centered Artificial Intelligence (HCAI), mandates comprehensive, ethical, and effective evaluation strategies. These encompass performance evaluation across diverse populations, clinical workflow integration assessment, fairness and equity analysis, user feedback incorporation, explainability and transparency assessment, robustness and reliability testing, regulatory compliance verification, and longitudinal performance monitoring. For instance, an AI model for diagnosing skin diseases must be rigorously tested on a varied set of skin tones to avoid biases and ensure accuracy for all patient groups. Clinical workflow integration is assessed through trials in real clinical settings, observing clinicians' interactions with the model. Fairness analysis ensures that the AI does not perpetuate healthcare disparities, necessitating an analysis of the model's predictions for biases based on sociodemographic factors. User feedback, particularly from clinicians, is used in evaluating the model's practicality and utility. Testing under various clinical scenarios assesses the model's robustness and reliability. Compliance with medical device and patient safety standards is a key aspect of regulatory verification. Longitudinal monitoring ensures sustained accuracy and effectiveness, adapting to changes in treatment practices and patient demographics.

Strategies include In the Model Deployment phase monitoring real-world performance and impact, facilitating clinician and patient understanding, ensuring ethical use and preventing misuse, adapting to diverse clinical environments, maintaining data security and patient privacy, responding to feedback for iterative improvements, assessing impact on clinical practice and patient outcomes, and compliance with evolving regulations and standards. Continuous monitoring of the AI model's performance post-deployment is essential to track accuracy and unintended consequences. Both clinicians and patients must understand the AI model's functions and limitations, necessitating training and clear explanations. Ethical use guidelines and safeguards are critical to prevent misuse. The AI model should be adaptable to various clinical environments and operational workflows. Maintaining high standards of data security and patient privacy is crucial even post-deployment. Iterative improvements based on user feedback enhance the AI system's utility. Assessing the impact of AI on clinical practice and patient outcomes is essential for evaluating its effectiveness. Staying updated with and compliant to evolving healthcare technology regulations ensures that the AI model remains relevant and safe for clinical use. For example, in deploying an AI system for patient risk stratification in hospitals, regular evaluation of its impact on patient outcomes and adjustments based on clinician feedback would be imperative for its ongoing effectiveness and ethical application.

references

Carlson, G. *et al.* (2020) "General Data Protection Regulation and California Consumer Privacy Act: Background," *Currents: J. Int'l Econ. L.* HeinOnline. Available at: https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/curritlj24&section=12.

Dignum, F. and Dignum, V. (2020) "How to center AI on humans," in *NeHuAI 2020, 1st International Workshop on New Foundations for Human-Centered AI, Santiago de Compostella, Spain, September 4, 2020*, pp. 59–62.

Holzinger, A. *et al.* (2022) "Digital Transformation in Smart Farm and Forest Operations Needs Human-Centered AI: Challenges and Future Directions," *Sensors* . mdpi.com, 22(8). doi: 10.3390/s22083043.

Khanna, S. and Srivastava, S. (2020) "Patient-Centric Ethical Frameworks for Privacy, Transparency, and Bias Awareness in Deep Learning-Based Medical Systems," *Applied Research in Artificial Intelligence and Cloud Computing*, 3(1), pp. 16–35.

Mämmelä, A. *et al.* (2018) "Multidisciplinary and Historical Perspectives for Developing Intelligent and Resource-Efficient Systems," *IEEE Access*. IEEE, 6, pp. 17464–17499.

Riedl, M. O. (2019) "Human-centered artificial intelligence and machine learning," *Human behavior and emerging technologies*. Hindawi Limited, 1(1), pp. 33–36.

Roth, W. *et al.* (2020) "Resource-Efficient Neural Networks for Embedded Systems," *arXiv [stat.ML]*. Available at: http://arxiv.org/abs/2001.03048.

Saxena, A. K. (2020) "Balancing Privacy, Personalization, and Human Rights in the Digital Age," *Eigenpub Review of Science and Technology*, 4(1), pp. 24–37.

Shneiderman, Ben (2020) "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy," *International Journal of Human–Computer Interaction*. Taylor & Francis, 36(6), pp. 495–504.

Shneiderman, B. (2020) "Human-centered artificial intelligence: Three fresh ideas," *AIS Transactions on Human-Computer Interaction*. aisel.aisnet.org. Available at: https://aisel.aisnet.org/thci/vol12/iss3/1/.

Xu, W. (2019) "Toward human-centered AI: a perspective from human-computer interaction," *Interactions*. New York, NY, USA: Association for Computing Machinery, 26(4), pp. 42–46.