IJEMD-CSAI, 4 (1) (2025) https://doi.org/ 10.54938/ijemdcsai.2025.04.1.468

International Journal of Emerging Multidisciplinaries:
Computer Science and Artificial Intelligence

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2791-0164 ISSN (online): 2957-5036

Towards Closing The Major Barrier of Adoption of Blackbox Models in The Medical Arena Based on Human-Centered XAI Design.

Abdullahi Isa[1], Souley Boukari[1] and Muhammad Aliyu[2]

1. Department of Computer Science, Faculty of Computing, Abubakar Tafawa Balewa University Bauchi, Nigeria

2. Department of Computer Science, Federal Polytechnic Bauchi Nigeria

Abstract

The opacity of black-box models presents a significant obstacle to their acceptance in the medical field. To improve their adoption, it is crucial to identify the stakeholders who need explanations of these models and to develop effective methods for providing these explanations. This paper aims to identify the key actors/stakeholders in the medical field who require explanations of black-box models to enhance their adoption. Through a comprehensive literature review, we identify physicians, patients, regulatory bodies, ethicists, and legal professionals etc. as the primary actors with information needs regarding the workings and rationale of black-box models. Physicians require explanations to validate predictions against their clinical expertise, while patients seek transparency to understand the basis of recommendations. Regulatory bodies focus on compliance and ethical considerations, while ethicists and legal professionals evaluate fairness and accountability. By providing tailored explanations to these actors, trust can be fostered, informed decision-making facilitated, ethical concerns addressed, regulatory compliance ensured, and effective communication established. This research highlights the information needs of various stakeholders, proposes two frameworks—Human-Centered XAI Design and a workflow for black-box model research—and emphasizes the importance of explanations in enhancing the adoption of black-box models in the medical field.

Keywords: XAI, Black Box Model, Interpretable Model, Actors/stakeholders in Medical Domain, Medical AI, Explainability, Human Center XAI.

Email Addresses : isaabdullahi2008@gmail.com (Abdullahi Isa) , bsouley2001@yahoo.com (Souley Boukari) , maliyudeba@gmail.com (Muhammad Aliyu)

1 INTRODUCTION

Artificial intelligence framework powered by deep learning techniques covers a range of application in many fields. Recent articles, reveal that such techniques became part of our daily lives, like recognizing and tracking our faces, by cameras in our mobile phone driven by artificial intelligence based on deep learning techniques [1], utilization of computational intelligence techniques to curtail Covid-19 pandemic as shown in [2], [3]. In addition, many essential games like [4], [5][and complex tasks such as [6], [7] had reported to have outperformed humans. Deep learning and artificial intelligence have become indispensable tools for scientific exploration, simulation, and prediction across various domains [8], [9], [10].

However, some critical application that involve human life and safety like (clinical Doman, driverless car) or finance for instance (trading algorithms), are highly concerns on "how and why" artificial intelligent applications, specifically deep learning techniques make decision in such kind of vital applications, because wrong decision can be disastrous. That shows the need to unmask the Blackbox nature associated with artificial intelligence frameworks. Such concern of Blackbox nature of AI framework can be a disqualifying or limiting factor for adapting AI framework in medical domain [11]. This shows lack of explainability of AI framework as one of the major factors that hinders the usability of AI in clinical domains as compared to other fields like entertainment industries.

Recent publications have highlighted the growing attention to explainable AI (XAI) systems, especially since 2020 [12]. This focus is driven by the increasing deployment of AI in critical domains such as healthcare, where explainability is crucial for adoption [13], [14], [15]. AI models must provide transparency to ensure trust from medical professionals, patients, and regulators. In healthcare, explainable AI is essential for clinical decision support systems (CDSS), where decisions made by AI directly impact patient care and must be understandable to end-users [16].

Legal mandates also require that AI systems elucidate how and why specific decisions are reached, especially in high-stakes areas like clinical diagnostics and financial services. In the healthcare domain, the potential for AI models to make biased or erroneous decisions—such as diagnostic inaccuracies or misinterpretations of patient data—has raised growing concerns. The public and regulatory bodies demand transparency to mitigate risks of discrimination [12], [17], [18]. Recent research on explainable AI models for sentiment analysis has shown that when AI systems are implemented without proper transparency, they can lead to biased or incorrect conclusions, which can have serious ethical and legal implications in fields like healthcare [16]. Furthermore, judges and regulators now require insights into the reasoning behind AI-driven decisions to ensure accountability under various legal frameworks [16], [17], [19].

This context raises critical questions for the adoption of AI in healthcare: Who are the actors that need explanations of black-box models in medicine? What information do they require? Which actors should be prioritized in explainable AI models, and why is this explanation critical to them? Addressing these

questions is key to developing reliable, trustable, and widely adopted AI systems in the medical domain. Insights gained from the practical application of XAI in sentiment analysis and bias detection indicate that a robust explanation framework is crucial for ensuring the fair and accurate use of AI in clinical environments [16].

2 METHODOLOGY

This study employs a qualitative approach, primarily focusing on a conceptual analysis of existing literature on explainable AI (XAI) in healthcare. The aim is to identify the key stakeholders, or actors, who require explanations of AI models to support their adoption in medical contexts. A systematic review of recent peer-reviewed literature, industry reports, and regulatory documents was conducted using databases like PubMed, google scholar and IEEE Xplore. Keywords like "explainable AI," "black-box models in healthcare," "AI adoption," and "medical decision support systems" were used to locate relevant studies. To ensure the rigor of the sources, only peer-reviewed journal articles, conference papers, and regulatory documents were included.

The research categorizes key actors in healthcare, including medical professionals, patients, AI developers, and regulators. For each group, the study identifies specific types of information required to build trust in AI systems, such as explanations of how AI decisions are made and the risks involved. Actors were prioritized based on their direct involvement in AI decision-making, with medical professionals, patients and regulators ranked higher due to their critical roles in influencing AI adoption and ensuring patient safety.

Finally, the study introduces a workflow for developing black-box AI models tailored to the medical domain. This workflow includes four sectors: XAI researchers, AI models, key actors, and government agencies. The process aims to improve transparency and accountability in AI systems, facilitating their wider adoption in healthcare. By providing a clear structure, the methodology allows future researchers to replicate and refine the approach for different medical applications.

3 RESULTS AND DISCUSSION

3.1 Status Quo

This section identifies various actors that need explanations of Blackbox models in medical field. Traying to answer first the question raised. Various effort has been made recently, trying to unmask the Blackbox nature of AI algorithms with different level of details and contain of the information based on actor. However, let identify sectors that required an explanation first before identifying actors that are in each sector. The following are sectors that required explanation of Blackbox model as shown in Fig. 1 below.

3.2 Sectors required XAI

This section, presents sectors that requires an explanation of Blackbox model as shown in Figure 1 below. The sectors consist of the following:

I. Hospitals
II. Government Agencies
III. XAI Researcher/Institutions

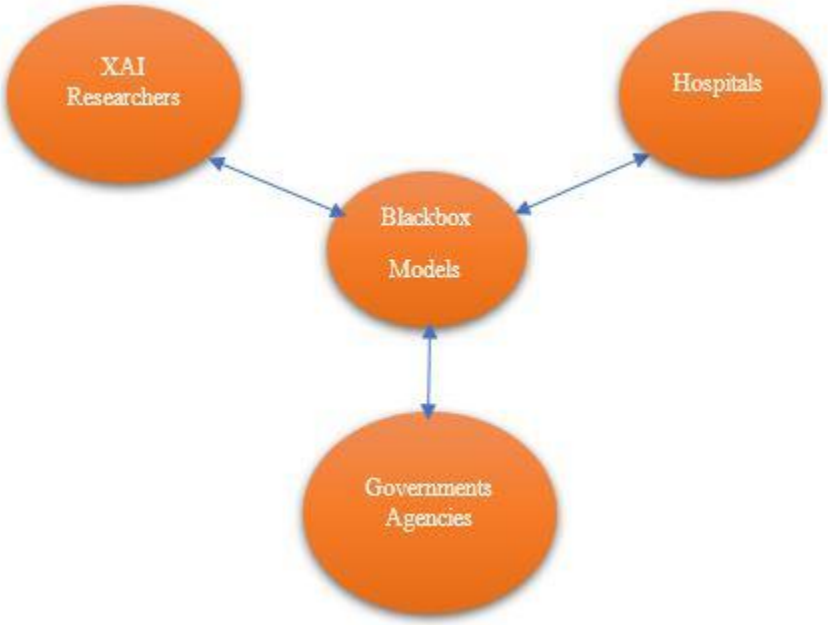

Figure 1 Sectors requires explanation of AI

3.2.1 Hospital

To identify the actors, involved in medical arena, there is need to understand normal daily routines/procedure in hospital, as well as the structure in hospital system. This will guide us in in identifying actors involved in medical field. For smooth operations, and delivering of high-quality services in hospital, the organization was structure vertically as shown in Figure 2 below.

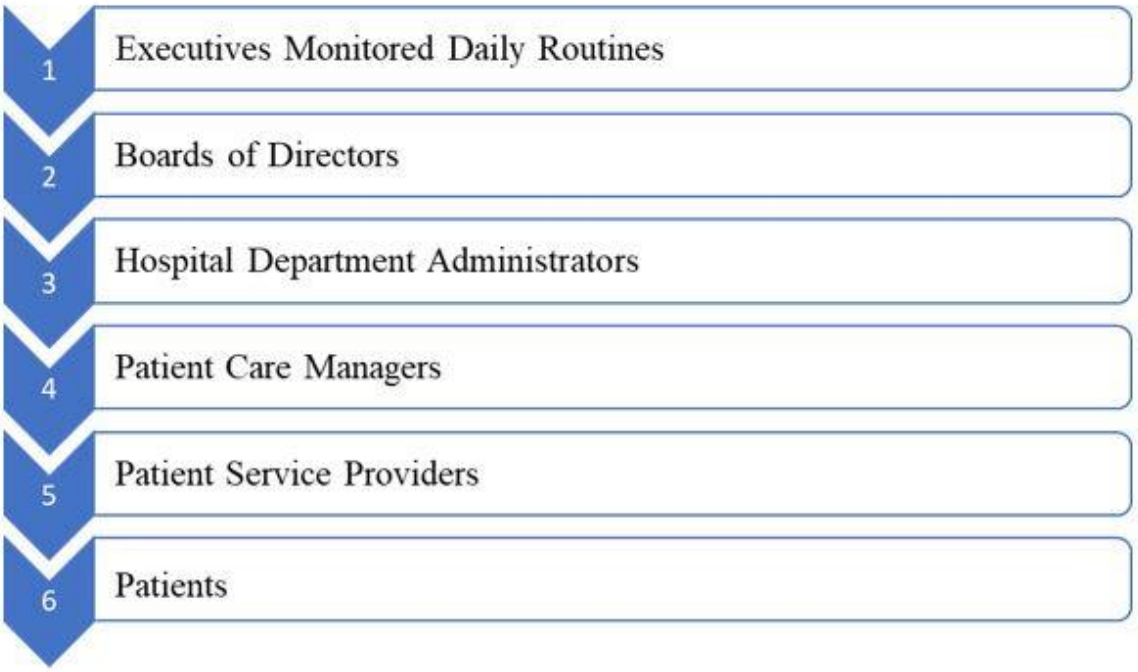

Figure 2 Hospital Organizational Structure vertically

Executives: oversees daily operation in hospital and ensured everything goes successfully. However, most hospital are comprising of chief nursing, medical doctors, information, financial and chief operation officers. This group of individuals create what is called executives.

Board of Directors: this depend on whether the hospital is profit or nonprofit ones. The former oversees by single individual, who control the affairs of the hospitals. However, nonprofit hospitals comprise of important personal in health care domain and leaders of local community. Some include clergies, Imams/Pastors, and congregational leaders, if the hospital was funded by particular religious organizations. Others hospital affiliated to educational institutions, there affairs are managed by top university official.

Hospital Department Administrators: these are managers that reports to board of directors. Because they oversee the clinical or operational service units like the Emergency Unit, Labor and Delivery Unit or Orthopedics Unit etc.

Patient Care Managers: they are directly in charge of patient care to ensure patient are best taking care off, staff work best appropriately, compliance with clinical ethic, rule and regulations of the hospitals, and abiding physician order by nurses and health allied care staffs. They are responsible for scheduling and human resources function of the staff. In addition, if anything goes wrong with clinicians or patient, they are in charge to fixed the issues.

Patient Service Providers: these are staff providing direct service to patients like doctors, nurses, radiologist, lab technician, physical therapist, pharmacist, laundry and cook staff etc. To ensured safety and health of patient restored, it required a lot of hands-on staff to make it possible.

Patient: an individual who requires or receives or are under physician care for illness or waiting for or undergoing medical treatment and care.

3.2.2 Government Agencies

These are initiatives by agencies of governments around the Globe, like UK, House of Loads committee of AI, EU, General Data Protection Regulation (GDPR) and Defense Advanced Research Projects Agency funds in USA, provides laws and regulations towards an attempt to provides AI based technology with ethical standards that characterized by preserving privacy, explainable, trustworthy, transpirable, reliable and fairness ability [20], [21].

3.2.3 XAI Researchers/Institutions

The goals of a various researchers and institutions of research is making contributions towards discovery of issues in the field of Explainable Artificial Intelligence, creation of new or enhancement of existing Blackbox models to satisfy hospital operations based on governments agencies regulations requirements as shown in Figure 1, in Abstract Actors requires explanations of Blackbox models. Here, the researchers create general concepts to provides solutions to the existing problems using prototypes presents their applicability in specific disease, for instance, cancer disease.

Hospital provides researchers with datasets and research questions aligned with given datasets on specific disease to provides an explanation to various groups in hospital as shown in Figure 2. Each group in hospital require different explanations, as individuals subject in group as well requires different explanations as provided by [17]. The prototype produced at this phase by researchers ware not run in medical daily routine, but rather publish their contribution in scientific journal.

3.3 Actors in medical arena

Different actors may need different contents of information as an explanations of AI Blackbox models on why and how they arrived at their decision. For instance, a client may request an explanation for discrimination of Blackbox models on why and how he/she was rejected for a loan in Bank as required by [17]. Coarse explanations may be sufficient for user of Blackbox models, because interpretation is very easy based on that. For deeper insights on working of the model, Blackbox model researchers and developers may certainly request that, to enable them enhance the model. For the case of actors involve in medical field, the focus will be on various groups such as patient care managers, patient service providers and patient. While executives, board of directors and hospital department administrators may require global explanation of the models, which can be obtained when many patients have been analyzed,

by checking the patterns which Blackbox model learned [11]. Below, the list of actors requires explanations of Blackbox models in medical arena to enhance adoption.

3.3.1 Categories of Actors in medical Field

This section outlines two principal categories of Actors/stakeholders in the medical field: major and non-major actors. These actors interact with black-box models, each requiring different levels of explanation to trust, understand, and effectively utilize these models in healthcare decisions, as depicted in Figure 3.

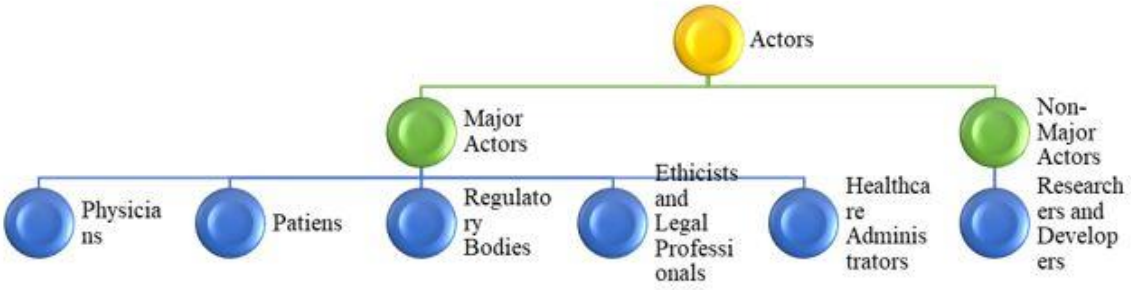

Figure 3 Categories of Actors in Medical Field

A. Major actors

These are key stakeholders who directly influence or are affected by the outcomes of black-box models. They need explanations of how these models operate to make informed decisions in clinical, regulatory, or administrative settings.

1. Physicians: Physicians are at the forefront of patient care and rely heavily on the predictions of black-box models to support clinical decision-making. Nevertheless, the opacity of these models necessitates interpretability to ensure that physicians trust and accurately apply the model's predictions. For example, comprehending the reasoning behind a model's diagnosis or treatment recommendation is essential for incorporating it into their clinical workflow. Research highlights the importance of creating interpretable models in healthcare settings, as physicians must be able to validate a model's output and assess its clinical relevance [22].

2. Patients: Patients play a central role in their own healthcare decisions, and as the recipients of care, they must be informed about how decisions regarding their treatment are made. Providing patients with explanations of black-box models allows them to understand the rationale behind recommendations, fostering trust and improving transparency. This enables patients to actively engage in shared decision-making processes, enhancing their capacity to make informed healthcare decisions. Clear explanations also help patients give informed consent, especially when treatments are based on complex AI models [23].

3. Regulatory Bodies: Regulatory bodies, such as the U.S. Food and Drug Administration (FDA), are responsible for ensuring that medical technologies, including AI-driven tools, comply with safety, ethical, and legal standards. For black-box models, transparency and interpretability are critical to evaluating the model's safety and effectiveness. Regulatory bodies assess whether these models meet regulatory frameworks and ethical guidelines, ensuring they do not pose risks to patient safety. In recent proposals, such as the FDA's regulatory framework for AI/ML-based medical devices, transparent explanations are emphasized to ensure accountability and foster trust [24].

4. Ethicists and Legal Professionals: Ethicists and legal professionals ensure that black-box models conform to ethical standards, including fairness, accountability, and transparency, while mitigating biases. These professionals critically examine the moral implications of using opaque AI models in healthcare. Explanations are necessary for them to assess whether these models comply with legal and ethical frameworks, protecting patients' rights and promoting fairness in healthcare decision-making [25].

5. Healthcare Administrators: Administrators in healthcare organizations are responsible for making strategic decisions about adopting new technologies. They require explanations of black-box models to understand their potential benefits, risks, and cost-effectiveness. Transparent models allow administrators to evaluate the impact on patient outcomes, resource allocation, and operational efficiency. By understanding how these models arrive at predictions, administrators can make well-informed decisions about their implementation in healthcare systems [26]. \

B. Non-Major Actors

Non-major actors encompass researchers and developers who are chiefly responsible for creating, refining, and enhancing black-box models in healthcare. Unlike major actors, their focus is primarily technical, ensuring the accuracy, reliability, and interpretability of models through rigorous methodologies like data collection, model design, and training processes.

Researchers and developers: These professionals work behind the scenes, ensuring that black-box models perform accurately and meet clinical standards. They use techniques such as algorithm optimization and data engineering to improve model performance and transparency. While their primary goal is to refine the models, they also contribute to making them more interpretable for major actors like physicians and regulatory bodies, thus ensuring that the models can be trusted and integrated effectively into clinical practice [22].

3.4 Information actors require

Different actors in the medical field require specific types of information to address their needs when interacting with black-box models. The following details the information necessary for both major and non-major actors.

A. Major Actors

These stakeholders have direct interactions with black-box models in healthcare decision-making and require explanations tailored to their role-specific needs.

I. Physicians:
Physicians rely on black-box models to aid in diagnosis, treatment decisions, and overall patient care. To integrate these models into their practice, they require:

a) Clinical relevance: Physicians need to comprehend how the model's predictions align with established clinical knowledge and guidelines. This understanding aids them in assessing the relevance and reliability of the model's output, ensuring it supports sound clinical judgment.

b) Risk assessment: Information on the model's performance metrics—such as sensitivity, specificity, and predictive values—is critical for physicians to evaluate risks and benefits. They need to gauge how well the model performs under different clinical scenarios and patient conditions.

c) Decision-making process: Physicians require transparency about the factors or features that the model used to arrive at its predictions. Understanding which patient characteristics were key to the model's output helps validate its reasoning, ensuring that it aligns with clinical expertise and supports decision-making [22].

II. Patients:
Patients are increasingly involved in their healthcare decisions, especially as AI-driven models play a role in their diagnosis and treatment. To trust these models, they need:

a) Explanation of predictions: Patients need to know how and why the model made its specific prediction or recommendation. Understanding the underlying factors that contributed to these predictions gives them confidence in the process.

b) Transparency: Patients require clear and comprehensible information on the model's operations, including its limitations and any biases that might influence its decisions. This is crucial for patients to make informed choices about their care.

c) Empowerment: Explanations should be simple and patient-friendly, empowering them to take an active role in their treatment plans. By demystifying the model's decision-making, patients are better positioned to collaborate with their healthcare providers and engage in shared decision-making [23].

III. Healthcare Administrators:

Administrators are responsible for evaluating and adopting black-box models at the institutional level. They require information that helps them assess the model's strategic value and operational feasibility:

a) Value proposition: Administrators need to understand the broader impact of implementing black-box models, including improvements in patient outcomes, cost-effectiveness, and resource allocation. They must justify the integration of these models based on both clinical and financial returns.

b) Risk assessment and management: Administrators require comprehensive performance reports detailing the model's potential risks and the strategies for mitigating them. This information helps them make decisions about adopting the model, including planning for contingencies and understanding liability concerns [26].

IV. Regulatory Bodies:

Regulatory authorities ensure that healthcare technologies meet ethical, legal, and safety standards. For black-box models, they require:

a) Transparency and interpretability: Regulatory bodies require detailed information about the model's architecture, algorithms, and decision-making process to determine its transparency and interpretability. This ensures that the model complies with regulatory standards for transparency and patient safety.

b) Compliance: The model must comply with ethical, legal, and regulatory guidelines. Explanations should demonstrate how the black-box model meets these standards, particularly in the areas of safety, accountability, and fairness [27].

V. Ethicists and Legal Professionals:

These professionals scrutinize the ethical implications and legal compliance of black-box models in healthcare. They need:

a) Fairness and bias: Ethicists and legal experts need to assess how the model manages fairness across various demographic groups and identify any potential biases. They require detailed information on how the model addresses issues of equity and justice in healthcare outcomes [25].

b) Accountability: Explanations should clarify who is responsible for the model's decisions and predictions, especially when these decisions affect patient well-being. Legal professionals need insight into the model's decision-making process to ensure accountability, transparency, and legal responsibility for errors or biases [25].

B. Non-Major Actors

Researchers and developers, while not directly involved in patient care, play a crucial role in designing and improving black-box models. They require specific technical and ethical information.

I. Researchers and developers

Researchers and developers focus on the model's design, performance, and ethical considerations. They need detailed insights into:

a) Data Collection and Preprocessing: Researchers and developers need to comprehend the data used to train the model, including its sources, quality, and how representative it is of the population it serves. Preprocessing steps such as data cleaning, normalization, and feature engineering are crucial for ensuring the model's robustness and preventing biases [22].

b) Model Architecture and Training Process: Researchers and developers require in-depth knowledge of the algorithms or neural networks used in the model, as well as details about the training process. This includes the optimization algorithms, regularization techniques, and hyperparameters that impact model performance [28].

c) Model Performance and Validation: Developers need access to information about the model's performance metrics (e.g., accuracy, precision, recall, F1 score) to assess its effectiveness. They should also be aware of validation techniques, such as cross-validation or using separate test datasets, to ensure the model generalizes well across different patient populations [22].

d) Ethical Considerations: Researchers need to be cognizant of the ethical issues associated with using black-box models in healthcare. These include potential biases, fairness, privacy concerns, and the overall impact of model predictions on patient outcomes [25].

3.5 Which actor will be giving more priority during Explanation?

There is no definitive answer to which actor will be given more priority during explanation of black-box models in the medical field, as it depends on the specific context and objectives of the explanation. However, the needs and perspectives of patients have gained increasing recognition and importance in recent years. Patient-centered care and shared decision-making are key principles in healthcare, emphasizing the involvement of patients in their own healthcare decisions. Therefore, providing explanations that cater to the information needs of patients is crucial to empower them in making informed choices about their treatment options [29]. However, it is important to note that other actors, such as physicians, regulatory bodies, and ethicists, also play significant roles and have specific information needs. Physicians require explanations to validate and trust the predictions of black-box models, while regulatory bodies need transparency and interpretability to ensure compliance with regulations and guidelines. Ethicists require information to assess the ethical implications and potential biases of the models. The prioritization of actors in providing explanations should be based on a balanced approach

that considers the needs and perspectives of all stakeholders involved. Achieving a comprehensive and inclusive explanation framework that addresses the information needs of multiple actors can contribute to a more transparent, accountable, and trusted adoption of black-box models in the medical field.

3.6 How can providing these explanations to actors involved enhance the adoption of such models?

The adoption of black-box models in the medical field is a multifaceted process, as it necessitates the confidence of a diverse group of stakeholders, each with their own concerns, responsibilities, and expectations. Delivering clear and tailored explanations to these stakeholders can significantly enhance the acceptance and implementation of such models in several keyways:

a. Building Trust: Trust is a foundational element in the healthcare environment, particularly when it comes to the integration of new technologies such as AI-driven black-box models. Providing explanations enhances transparency by shedding light on the model's decision-making process [22]. Physicians can trust the model when they understand the logic behind its predictions and how these predictions align with clinical guidelines. This transparency allows them to confidently integrate the model's recommendations into their practice without feeling undermined by a lack of clarity. Patients are more likely to trust black-box models when they are offered clear explanations that demystify the reasoning behind their personalized predictions. Without this, patients may feel alienated by automated systems that influence critical health decisions. Regulatory bodies, too, require comprehensive explanations to trust that black-box models operate safely and ethically. The transparency provided by explanations helps them see that the model complies with established guidelines, making them more likely to approve its use.

b. Informed Decision-Making: One of the most critical benefits of providing explanations is that it enables informed decision-making for both physicians and patients [23]. For Physicians: With clear explanations, physicians can better assess the clinical relevance of the model's predictions. When they understand how a model arrives at certain outcomes, they can weigh the predictions against their own clinical expertise and experience. This synergy between AI and physician judgment enhances the overall quality of medical decision-making, allowing for more accurate diagnoses, personalized treatments, and improved patient outcomes. For Patients: Explanations also play a critical role in empowering patients by enabling them to participate more actively in their healthcare decisions. When patients are provided with transparent and comprehensible explanations of the model's recommendations, they are in a better position to ask informed questions, engage in shared decision-making with their healthcare providers, and feel more in control of their treatment choices. This patient-centric approach improves adherence to treatment plans and boosts satisfaction.

c. Regulatory Compliance: In healthcare, regulatory bodies such as the FDA (Food and Drug Administration) or EMA (European Medicines Agency) are tasked with ensuring that new technologies meet strict standards for safety, effectiveness, and ethics. Explanations are crucial for fulfilling the transparency and accountability requirements set by these regulators. [24]. Regulatory Transparency: Black-box models often face criticism for their lack of interpretability,

complicating the regulatory approval process. Providing detailed explanations of the model's architecture, algorithms, and decision-making processes shows that these models can be understood and scrutinized. This is vital for ensuring that models meet the necessary standards for safety and ethics. Adherence to Standards: Clarifying how the model adheres to legal, ethical, and professional guidelines enables regulators to assess its interpretability and fairness. This minimizes the risk of regulatory pushback, streamlining the approval process and paving the way for quicker implementation.

d. Addressing Ethical Concerns: The implementation of black-box models raises a host of ethical issues—from fairness and bias to accountability and the broader impact on patient care. By offering clear explanations, these ethical concerns can be addressed head-on [25]. For Ethicists: Providing explanations allows ethicists to scrutinize the model for potential biases and evaluate its fairness across different patient demographics. This is particularly important for ensuring that vulnerable populations are not disproportionately impacted by algorithmic decisions. For Legal Professionals: Explanations allow legal experts to evaluate the accountability mechanisms in place for when the model makes a faulty or harmful prediction. Understanding how decisions are made within the black-box model ensures that legal professionals can assess liability, transparency, and ethical responsibility, particularly in the event of adverse outcomes.

3.7 Why explanations important to them?

Explanations are essential to various actors in the medical field—physicians, patients, healthcare administrators, regulatory bodies, ethicists, and legal professionals—for several reasons. Each group relies on explanations to enhance their understanding, facilitate decision-making, and ensure ethical and practical integration of black-box models into healthcare:

a. Trust and Confidence: In healthcare, trust is paramount for the successful adoption of new technologies. Explanations build trust by providing transparency into the inner workings of black-box models, which are often criticized for their opaque nature. Physicians need to feel confident that a model’s predictions are aligned with established medical knowledge and clinical guidelines. By explaining how the model generates its predictions—whether through feature importance, decision paths, or statistical patterns—physicians can validate its accuracy, reliability, and applicability. Without these explanations, they may hesitate to integrate black-box models into their clinical workflow due to concerns over the model's opacity and unpredictability. Patients require a level of transparency that makes them comfortable with AI-based decisions in their care. Explanations demystify complex algorithms and give patients insights into how predictions were made. This transparency fosters trust in the technology, allowing patients to feel reassured that their treatment is based on sound, data-driven reasoning, rather than arbitrary or hidden mechanisms.

b. Validation and Accountability: Explanations provide stakeholders with the tools to validate and assess the performance of black-box models [27]. Physicians and regulatory bodies need explanations to scrutinize the reasoning behind the model’s predictions. For instance, explanations

enable physicians to determine whether the model's output aligns with their clinical expertise, patient history, and medical evidence. If the model's reasoning appears flawed or counterintuitive, physicians can modify their treatment plans or disregard the model's recommendation. For regulatory bodies, explanations are essential for evaluating the model's compliance with healthcare regulations and standards. These explanations help regulators understand how the model meets requirements for accuracy, safety, and fairness, ensuring that it adheres to legal and professional guidelines. This facilitates the assessment of whether the model should be approved for clinical use.

Accountability is also a significant concern, especially when AI-driven decisions impact patient outcomes. Explanations create accountability by making it clear who is responsible for the decision—the model or the human using it. Physicians, legal professionals, and ethicists need to know how decisions are made in order to assign responsibility correctly in case of a medical error or adverse event. Explanations make it possible to pinpoint where errors might have occurred, whether in the model's logic, data input, or human oversight.

c. Ethical Considerations: In healthcare, ethical considerations are paramount, and explanations help uncover potential ethical issues in black-box models [25]. Ethicists and legal professionals rely on explanations to assess issues like fairness, bias, and transparency. For example, explanations can reveal whether the model treats certain demographic groups unfairly, potentially resulting in biased or discriminatory outcomes. Understanding the decision-making process enables these professionals to evaluate whether the model adheres to ethical principles and ensures equitable treatment for all patients. Accountability and responsibility are also ethical concerns. Explanations help ensure that decisions made by black-box models are ethically sound and that those making healthcare decisions—whether human or algorithm—can be held accountable. Legal professionals need explanations to understand the AI's role in healthcare decisions, ensuring liability can be accurately assigned if something goes wrong.
d. Regulatory Compliance: For regulatory bodies like the FDA or EMA, explanations are essential in ensuring that black-box models adhere to strict standards of safety, efficacy, and transparency. Regulatory agencies require that black-box models be interpretable and explainable, as these qualities are essential for verifying that the model's predictions are accurate and fair. Explanations show that the model can be trusted to operate within regulatory frameworks and ensure patient safety. Adherence to ethical and legal standards is also a critical concern. Regulatory bodies need to know how the model was trained, validated, and deployed, and explanations make it easier to audit these processes. Without clear explanations, regulatory approval becomes far more difficult, slowing down the implementation of black-box models in healthcare.
e. Effective Communication Among Stakeholders: Explanations play a key role in bridging communication gaps between different stakeholders in the medical field, such as researchers, developers, administrators, physicians, patients, and regulatory bodies. Researchers and developers need to communicate the capabilities, limitations, and potential risks of black-box models effectively to other stakeholders. Explanations provide a structured way to convey this

information, allowing all parties to understand the benefits and risks of adopting these models. Healthcare administrators require explanations to justify the costs, benefits, and potential return on investment when implementing AI-driven systems. Explanations help them assess whether black-box models are likely to improve patient outcomes and optimize resource allocation.

4 PROPOSE FRAMEWORKS

4.1 Proposed Human-Centered XAI Design

The integration of XAI in healthcare necessitates a human-centered approach that tailors' explanations to the diverse needs of stakeholders, ensuring transparency, trust, and usability as shown in Figure 4. This proposed framework extends existing XAI methods, such as SHAP and LIME, by providing stakeholder-specific explanations that align with their distinct roles and decision-making requirements as shown in section 3.4. For physicians, explanations should be clinically relevant, mapping AI predictions to established medical guidelines and presenting key risk assessment metrics such as sensitivity, specificity, and predictive values. Patients, on the other hand, require non-technical explanations using simplified language, visual aids like heatmaps or decision trees, and personalized insights to help them understand the rationale behind their treatment recommendations. Regulatory bodies need detailed documentation on model transparency, compliance with regulations like GDPR, and mechanisms for auditing AI decision-making processes. Ethicists and legal professionals require fairness assessments, bias analysis across demographic groups, and accountability mechanisms that trace AI decisions and identify potential ethical risks. Meanwhile, healthcare administrators must evaluate the broader value proposition of AI adoption, considering its impact on patient outcomes, cost-effectiveness, and institutional resource allocation. Finally, researchers and developers require in-depth technical explanations covering data preprocessing, model architecture, performance evaluation, and ethical considerations related to bias mitigation and explainability constraints.

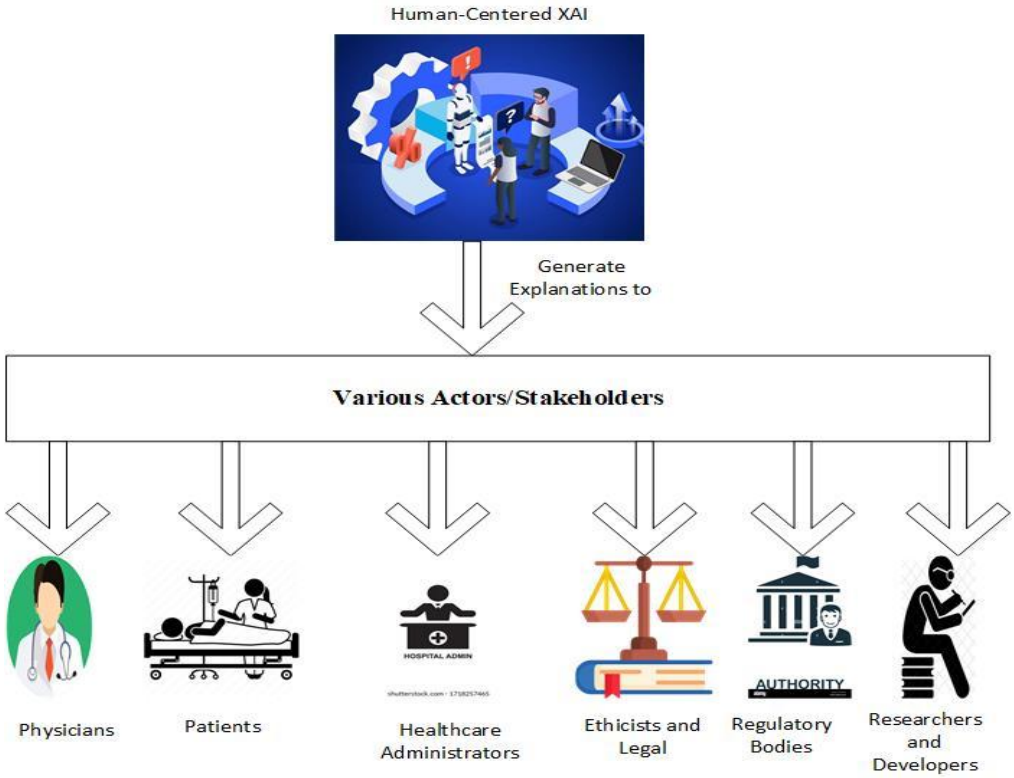

Figure 4: Human-Centered XAI Design

By integrating stakeholder-specific explanations, this framework enhances the interpretability of black-box models, fostering greater adoption of AI in high-stakes medical applications. Ensuring that each stakeholder receives explanations tailored to their expertise and needs bridges the transparency gap, ultimately improving AI-driven decision-making in clinical settings. This approach not only strengthens trust in AI systems but also promotes ethical compliance, fairness, and accountability. Furthermore, addressing scalability challenges and developing robust evaluation metrics will be crucial in refining this framework for real-world implementation. As AI continues to reshape healthcare, a structured, human-centered XAI design will be essential in facilitating responsible and equitable deployment, ensuring that AI-driven innovations contribute meaningfully to medical practice and patient care.

4.2 Proposed Workflow

This section presents the proposed workflow, as shown in Figure 5, which integrates four key components: XAI Researchers, Black Box Models, Actors, and Government Agencies. Each component plays a crucial role in ensuring the ethical, legal, and regulatory adoption of AI systems in the medical field.

1. **XAI Researchers** are responsible for designing and enhancing black-box models, with a focus on making them explainable and compliant with the established guidelines. These researchers work to integrate transparency and interpretability into the AI models while addressing challenges posed by complex machine learning algorithms. Their role is to ensure that the models not only provide

accurate results but also generate understandable explanations for stakeholders. This involves applying advanced explainable AI (XAI) techniques that meet ethical standards, legal obligations, and the technical requirements laid out by regulatory bodies.

2. **Black Box Models** are the AI systems being developed for clinical applications. These models often operate with complex algorithms that make their decision-making processes opaque, which raises concerns in healthcare. The models undergo thorough scrutiny by **Actors**, including medical professionals, patients, and AI developers, who assess their adherence to established standards. The focus is on ensuring that these models not only deliver reliable outcomes but also provide explanations that are accessible and useful to medical practitioners, enabling informed decisions and fostering trust in AI-driven clinical solutions.

3. **Government Agencies** serve a pivotal role in this workflow by establishing regulatory standards that define how AI models should be designed, deployed, and monitored in the healthcare setting. These agencies ensure that the XAI researchers and the actors involved in the healthcare system follow stringent guidelines to ensure safety, ethical use, and compliance with healthcare laws. They also engage in continuous collaboration with researchers and medical professionals to update and refine regulations based on emerging AI technologies and their impact on clinical practice.

In this workflow, there is a continuous feedback loop where actors, such as healthcare professionals, assess AI models in practice to verify compliance with regulatory standards. Any discrepancies are reported back to XAI researchers, who refine the models to meet both technical performance and regulatory requirements, while government agencies update standards as necessary to keep pace with advancements in AI technologies. This cyclical process fosters accountability, reliability, and the successful adoption of AI in the medical field.

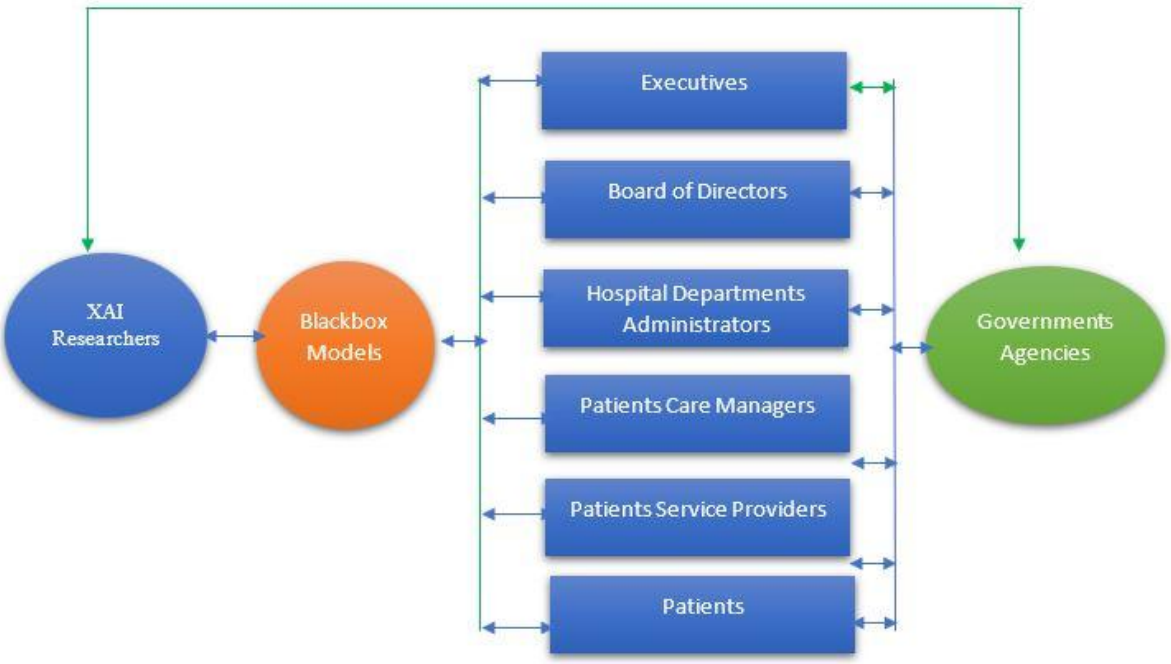

Figure 5 Proposed Workflow: Human-Centred XAI for Healthcare.

Recommendations for an explainable Blackbox models to enhance adoption in medical field
Here are recommendations emphasizing the importance of explanations to enhance the adoption of black-box models in the medical field:

To boost the adoption of explainable black-box models in the medical field, one key recommendation is to develop models that are more transparent and interpretable. While black-box models are powerful, their lack of transparency can lead to skepticism, particularly in high-stakes areas like healthcare. Researchers and developers should focus on techniques such as rule-based models, decision trees, or hybrid models that combine transparency with the power of deep learning. This approach enables healthcare professionals to trace the decision-making process, fostering trust and increasing the likelihood of adoption.

Another critical recommendation is to provide contextual explanations tailored to the specific needs of different physicians, patients, and regulators. Explanations must be relevant to the medical context, highlighting how predictions align with clinical guidelines and the patient's data. By delivering explanations that match the expertise of the user, such as detailed technical insights for physicians or simplified reasoning for patients, black-box models can become more accessible and accepted across the medical field.

Incorporating uncertainty estimation in explanations is crucial for earning the trust of medical professionals. Models should convey the level of confidence or uncertainty in their predictions, aiding healthcare providers in making informed decisions by understanding the risks and reliability of the

model's output. This feature is especially important in healthcare, where the stakes are high, and understanding the margin of error is vital for effective treatment planning.

Additionally, ensuring that explanations are Human-Centered based as demonstrated in 4.1 and user-friendly and accessible is crucial for widespread adoption. Explanations should use clear, non-technical language, and incorporate visual aids such as charts or heatmaps to make complex predictions easier to understand. For patients and non-technical actors, user-friendly designs can enhance comprehension, while interactive tools can engage users by allowing them to explore how different factors impact predictions [30].

Lastly, fostering collaboration and feedback as demonstrated in section 4.2 proposed workflow, where actors, like healthcare professionals, assess AI models in practice to verify compliance with regulatory standards. Any discrepancies are reported back to XAI researchers, who refine the models to meet both technical performance and regulatory requirements, while government agencies update standards as necessary to keep pace with advancements in AI technologies. This cyclical process fosters accountability, reliability, and the successful adoption of AI in the medical field

5 CONCLUSION

In conclusion, this research highlights a critical challenge in the adoption of black-box models in the medical field by identifying the key stakeholders who require explanations for these models. The study underscores the vital roles of various stakeholders, including physicians, patients, regulators, ethicists, and legal professionals, in the successful integration of black-box models into clinical practice. Through a comprehensive analysis of explainable AI (XAI) within the healthcare context, this work identifies the specific information each stakeholder requires and highlights the importance of tailored explanations to foster trust and usability.

The practical implications of this research are clear in the proposed workflow, which involves collaboration among XAI researchers, healthcare professionals, and regulatory bodies to ensure that black-box models meet ethical, legal, and technical standards. This workflow, coupled with recommendations such as developing transparent models, providing contextual explanations, and incorporating uncertainty estimates, establishes a foundation for enhancing the reliability and acceptance of AI systems in the medical field. The findings serve as a roadmap for developers and policymakers to create AI models that not only perform well but are also understandable and trusted by all stakeholders.

Theoretically, this work deepens the understanding of how explainable AI (XAI) can bridge the gap between complex AI algorithms and practical healthcare applications. By concentrating on the needs of various stakeholders and their interactions with AI models, this research enhances the growing body of knowledge on Human-Centered XAI, especially in high-stakes domains like medicine. Future research can implement the proposed frameworks and build upon them to further refine XAI techniques and examine their impact on healthcare outcomes.

Competing Interests

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We would like to extend our sincere gratitude to the Department of Computer Science, Faculty of Computing, Abubakar Tafawa Balewa University Bauchi, Nigeria, for providing the essential resources and facilities that supported this research. We also acknowledge the valuable feedback and suggestions from our colleagues and peers in the Department of Computer Science, University of Maiduguri, which significantly contributed to the development of this work. Furthermore, we express our heartfelt thanks to our family and friends for their unwavering support and understanding throughout the research process. It is important to note that this research did not receive any grants or financial support from any funding agencies, public or private.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[2] A. Isa, "Computational Intelligence Methods in Medical Image-Based Diagnosis of COVID-19 Infections," in *Studies in Computational Intelligence*, vol. 923, Springer Science and Business Media Deutschland GmbH, 2021, pp. 251–270. doi: 10.1007/978-981-15-8534-0_13.

[3] T. Goel, R. Murugan, S. Mirjalili, and D. K. Chakrabartty, "Automatic Screening of COVID-19 Using an Optimized Generative Adversarial Network," *Cognit Comput*, vol. 1, p. 3, Jan. 2021, doi: 10.1007/s12559-020-09785-7.

[4] D. Silver *et al.*, "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, Oct. 2017, doi: 10.1038/nature24270.

[5] M. Moravčík *et al.*, "DeepStack: Expert-level artificial intelligence in heads-up no-limit poker," *Science (1979)*, vol. 356, no. 6337, pp. 508–513, May 2017, doi: 10.1126/science.aam6960.

[6] C. Lu and X. Tang, "Surpassing human-level face verification performance on LFW with GaussianFace," in *Proceedings of the National Conference on Artificial Intelligence*, AI Access Foundation, Jun. 2015, pp. 3811–3819.

[7] D. Cireşan, U. Meier, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *Proceedings of the International Joint Conference on Neural Networks*, 2011, pp. 1918–1921. doi: 10.1109/IJCNN.2011.6033458.

[8] S. Chmiela, H. E. Sauceda, K. R. Müller, and A. Tkatchenko, "Towards exact molecular dynamics simulations with machine-learned force fields," *Nat Commun*, vol. 9, no. 1, pp. 1–10, Dec. 2018, doi: 10.1038/s41467-018-06169-2.

[9] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nat Commun*, vol. 8, no. 1, pp. 1–8, Jan. 2017, doi: 10.1038/ncomms13890.

[10] A. W. Thomas, H. R. Heekeren, K.-R. Müller, and W. Samek, "Analyzing Neuroimaging Data Through Recurrent Deep Learning Models," *Front Neurosci*, vol. 13, p. 1321, Dec. 2019, doi: 10.3389/fnins.2019.01321.

[11] W. Samek and K. R. Müller, "Towards Explainable Artificial Intelligence," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11700 LNCS, Springer Verlag, 2019, pp. 5–22. doi: 10.1007/978-3-030-28954-6_1.

[12] R. Confalonieri, L. Coba, B. Wagner, and T. R. Besold, "A historical perspective of explainable Artificial Intelligence," *WIREs Data Mining and Knowledge Discovery*, vol. 11, no. 1, p. e1391, Jan. 2021, doi: 10.1002/widm.1391.

[13] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning."

[14] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.," *Queue*, vol. 16, no. 3, pp. 31–57, May 2018, doi: 10.1145/3236386.3241340.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, "Nothing Else Matters: Model-Agnostic Explanations By Identifying Prediction Invariance," *arXiv:1611.05817 [stat.ML]*, Nov. 2016, Accessed: Nov. 16, 2023. [Online]. Available: https://arxiv.org/abs/1611.05817v1

[16] S. Velampalli, C. Muniyappa, and A. Saxena, "Performance Evaluation of Sentiment Analysis on Text and Emoji Data Using End-to-End, Transfer Learning, Distributed and Explainable AI Models," *Journal of Advances in Information Technology*, vol. 13, no. 2, pp. 167–172, Apr. 2022, doi: 10.12720/JAIT.13.2.167-172.

[17] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (United Kingdom General Data Protection Regul," *https://webarchive.nationalarchives.gov.uk/eu-exit/https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504*.

[18] B. Waltl, "Explainable Artificial Intelligence : the new frontier in legal informatics," *Datenschutz - LegalTech : Tagungsband des 21. Internationalen Rechtsinformatik Symposions IRIS 2018 = Data Protection - LegalTech : Proceedings of the 21st International Legal Informatics Symposium*, 2018.

[19] D. Ashley, "The Judicial Demand for Explainable Artificial Intelligence," *Virginia Public Law and Legal Theory Research Paper*, 2019, Accessed: Nov. 16, 2023. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3440723#paper-citations-widget

[20] G. Yang, Q. Ye, and J. Xia, "Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion : A mini-review , two showcases and beyond," *Information Fusion*, vol. 77, no. January 2021, pp. 29–52, 2022, doi: 10.1016/j.inffus.2021.07.016.

[21] M. Van Gerven and P. Haselager, “Explanation Methods in Deep Learning : Users , Values , Concerns and Challenges ∗,” no. February 2021, 2018, doi: 10.1007/978-3-319-98131-4.

[22] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, “Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 2015-August, pp. 1721–1730, Aug. 2015, doi: 10.1145/2783258.2788613.

[23] F. Cabitza, R. Rasoini, and G. F. Gensini, “Unintended Consequences of Machine Learning in Medicine,” *JAMA*, vol. 318, no. 6, pp. 517–518, Aug. 2017, doi: 10.1001/JAMA.2017.7797.

[24] J. A. Smith, R. E. Abhari, Z. Hussain, C. Heneghan, G. S. Collins, and A. J. Carr, “Industry ties and evidence in public comments on the FDA framework for modifications to artificial intelligence/machine learning-based medical devices: A cross sectional study,” *BMJ Open*, vol. 10, no. 10, 2020, doi: 10.1136/bmjopen-2020-039969.

[25] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, “The ethics of algorithms: Mapping the debate,” *Big Data Soc*, vol. 3, no. 2, Dec. 2016, doi: 10.1177/2053951716679679/ASSET/IMAGES/LARGE/10.1177_2053951716679679-FIG1.JPEG.

[26] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nature Medicine 2019 25:1*, vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: 10.1038/s41591-018-0300-7.

[27] “Artificial Intelligence and Machine Learning in Software as a Medical Device | FDA.” Accessed: May 28, 2023. [Online]. Available: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

[28] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, “Explainable artificial intelligence (XAI) in deep learning-based medical image analysis,” *Med Image Anal*, vol. 79, p. 102470, Jul. 2022, doi: 10.1016/J.MEDIA.2022.102470.

[29] C. Charles, A. Gafni, and T. Whelan, “Shared decision-making in the medical encounter: What does it mean? (Or it takes, at least two to tango),” *Soc Sci Med*, vol. 44, no. 5, pp. 681–692, Mar. 1997, doi: 10.1016/S0277-9536(96)00221-3.

[30] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, “Causability and explainability of artificial intelligence in medicine,” *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 9, no. 4, Jul. 2019, doi: 10.1002/WIDM.1312.