IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 13, 2021, accepted November 6, 2021, date of publication November 12, 2021,
date of current version November 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3127881

A Systematic Review of Human–Computer Interaction and Explainable Artificial Intelligence in Healthcare With Artificial Intelligence Techniques

MOBEEN NAZAR[1,2], MUHAMMAD MANSOOR ALAM[1,3], EIAD YAFI[4], (Senior Member, IEEE), AND MAZLIHAM MOHD SU'UD[1,5]

[1]Malaysian Institute of Information Technology, Universiti Kuala Lumpur (UniKL MIIT), Kuala Lumpur 50250, Malaysia
[2]Department of Software Engineering, Bahria University, Karachi Campus, Karachi 75260, Pakistan
[3]Faculty of Computing, Riphah International University, Rawalpindi 46000, Islamabad
[4]Faculty of Information and Communication Technologies, Institute of Business, Dili, Timor-Leste
[5]Univeristi Kuala Lumpur-Malaysia France Institute (UmiKL-MFI), Bandar Baru Bangi 43650, Malaysia

Corresponding author: Mazliham Mohd Su'ud (mazliham@unikl.edu.my)

ABSTRACT Artificial intelligence (AI) is one of the emerging technologies. In recent decades, artificial intelligence (AI) has gained widespread acceptance in a variety of fields, including virtual support, healthcare, and security. Human-Computer Interaction (HCI) is a field that has been combining AI and human-computer engagement over the past several years in order to create an interactive intelligent system for user interaction. AI, in conjunction with HCI, is being used in a variety of fields by employing various algorithms and employing HCI to provide transparency to the user, allowing them to trust the machine. The comprehensive examination of both the areas of AI and HCI, as well as their subfields, has been explored in this work. The main goal of this article was to discover a point of intersection between the two fields. The understanding of Explainable Artificial Intelligence (XAI), which is a linking point of HCI and XAI, was gained through a literature review conducted in this research. The literature survey encompassed themes identified in the literature (such as XAI and its areas, major XAI aims, and XAI problems and challenges). The study's other major focus was on the use of AI, HCI, and XAI in healthcare. The poll also addressed the shortcomings in XAI in healthcare, as well as the field's future potential. As a result, the literature indicates that XAI in healthcare is still a novel subject that has to be explored more in the future.

INDEX TERMS Artificial intelligence, deep learning, explainable artificial intelligence, healthcare, human-computer interaction, human-centered design, machine learning, usability, user-centered design.

I. INTRODUCTION

Nowadays, digital technologies have been adopted for interaction in the modern world. Computing has become one of the most essential and integral parts of all industries and disciplines. Among all the advanced technologies, mobile computing has become one of the dominant factors in today's era [1]. Technological interaction has been given importance in many advanced areas. Interaction factors also have significant importance in the technical perspective for being easily used and managed by the person using it [2]. Artificial intelligence playing a vital role in making interaction more flexible and intelligent by integrating itself within the systems with the help of different technology acceptance theories [3]. Human-Computer Interaction is the field that is mainly used for making technological interaction easy for the user [4]. Artificial intelligence, in other words is known for making interactions intelligent [5]. A new era has created a bridge between human-computer interaction and artificial intelligence by introducing Explainable artificial intelligence. The main focus of XAI is mainly to explain the interaction to the end-user in order to create a trustworthy environment [6]. Explainability in AI is a field that is active in a variety of domains, including medical healthcare, business processes, security, financial and legal decisions, autonomous vehicles, smartphones, and AI for designers [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno Garcia.

The research mainly focuses on the emerging field of Explainable Artificial Intelligence and its challenges and problems in health care. The major goals of this research paper are:

(a) To identify problems in Human-Computer Interaction.
(b) To review Machine Learning Characteristics.
(c) To identify different techniques for Explainable Artificial Intelligence.
(d) To identify problems and challenges in Explainable Artificial Intelligence.
(e) To review Explainable Artificial Intelligence in healthcare.
(f) To identify challenges of Explainable Artificial Intelligence in healthcare.

II. METHODOLOGY OF RESEARCH

This section will elaborate on the research design and set of research papers explored in the literature, with additional data sources and explanation criteria.

A. RESEARCH QUESTIONS

The research questions addressed in the research are:

(a) What are the problems in HCI?
(b) What is the Significance of ML Characteristics?
(c) What are the Various Explainability techniques?
(d) What are the problems and challenges in XAI?
(e) What are the challenges of XAI in healthcare and how to overcome it?
(f) What is the research gap existing in XAI in healthcare?
(g) What is the future of XAI in healthcare?

The above-mentioned questions are answered below by the effective information gathered through research related to HCI, AI, and XAI. The responses are described below:

1) WHAT ARE THE PROBLEMS IN HCI?

The research has shown that HCI has further divided into two of its subfields, which include usability and Human-Centered Design (HCD). Usability is mainly concerned with issues related to interfaces and the only problem they can encounter is a deficiency in the interfaces. For that purpose, the user can check the usability of the interfaces given in section III (A) below. The HCD problems that may be encountered are described in [8], which includes requirements that are not clear, the design solution is not correct, and the context of use is not clear. These problems can be neglected by taking clear requirements from the client in the beginning and involving the client in it. For the design purpose, usability should be considered so that the user level can be encountered, and for the last problem, concerning the HCD context of the system, it must be understandable to achieve satisfaction from the user.

2) WHAT IS THE SIGNIFICANCE OF ML CHARACTERISTICS?

Machine learning is the subfield of AI, which is performing in many domains. Machine Learning (ML) has many of the characteristics that have huge importance and can provide easy adaptability to the users working in machine learning. The accountable factor is directly concerned with consequences related to any event if occurs [9]. The autonomous concern of machine learning is related to autonomous sensing and decision making in a dynamic environment. It usually works with software agents and autonomous vehicles [10]. The fairness of the ML model ensures individuals will not be treated based on their race, gender, and disabilities [11]. Transparency of the ML model provides the end-user with information on how the model works [12]. The other factor in the explainability of ML is associated with providing the user with information on why the model is working this way [12]. A trustworthy model provides safety to the user, which makes them use the system with no concern [13]. The last thing covered by the ML model is privacy, which will not lead to the unauthorized use of services [14], [15].

3) WHAT ARE THE VARIOUS EXPLAINABILITY TECHNIQUES?

The focus of the paper is on XAI, which is the combinational field of Artificial Intelligence and Human-Computer Interaction. The primary concern of XAI is explainability, so we have picked the factor of explainability and described 11 techniques of explainability. Every technique has its importance, which discussed explainability concerns in different areas or fields. The first technique is text-based explainability, which is used to bring the explainability model for generating text explanation of the results [7]. The local explanation is used to express the individual decision of the classifier [16]. A global explanation is another explainability technique that is used as a whole for explaining the entire function of algorithms [12]. Visual explanation, on the other hand, provides visual explanation and the behavior of the model [7]. Another model explanation by example is concerned for the data extraction and better understanding of the model itself [17]. Explanation by simplification is a type of explanation in which a whole new system is rebuilt based on the trained model to be explained [7]. Feature relevance is another technique related to the post-hoc explainability of the internal function of the system for defining how it affects the decisions [7]. Provenance-based explainability technique is used for illustrating explainability and is said to be one of the effective usability techniques. The surrogate model technique on the other hand, uses other models as a proxy for explainability. The declarative induction technique sets rules, trees, and programs that are said to be a human-readable representation. Last explainability technique rule-based methods are good for learning the classification of data [18]. The techniques discussed above prove the importance of explainability in different domains.

4) WHAT ARE THE PROBLEMS AND CHALLENGES IN XAI AND HOW TO OVERCOME IT?

The survey disclosed many of the challenges related to the field of XAI. It is emerging in many domains and has many issues and problems identified in the literature. The major challenges identified and discussed in the papers from the literature are security, performance, vocabulary, evaluation of explanation, generalization of XAI. For the challenge of vocabulary, the expertise of the audience should be involved in the XAI model to determine what explanation they expect from the XAI model [7]. Evaluation of explanation can be explained by Ad-hoc experiments or the KPI method or in the general case, other proxy measures can be used, such as the number of rules, nodes, or input variables considered in an explanation or explainable model for evaluation. The improvement of XAI performance by creating hybrid AI combining different approaches. Use of Machine Learning algorithms to work from existing expert-built symbolic representations of physical models to leverage existing knowledge [19]. The generalization problem of XAI can be resolved through a focus on domain-level explanation and work towards local generalization from a domain standpoint. For the security purposes of XAI perturbations have been given importance as input data to learning models, bias, and fairness are the key enablers for the security of AI models [20]. Hence, domain adaption in XAI remains a challenge along with the XAI twin and the performance factor of XAI needs to be worked further for future directions.

5) WHAT ARE THE CHALLENGES OF XAI IN HEALTHCARE AND HOW TO OVERCOME IT?

The research in this paper has reviewed many articles which identify the gaps in XAI. The domain gaps which have been viewed in this article are related to healthcare. The paper has identified many of the challenges related to XAI in the healthcare domain, which include System Evaluation, Organizational, Legal, socio-relational, and Communicational issues, XAI. A few papers from the literature have identified some solutions for the identified issues but, others have suggested it as a research gap that needs to be filled in the future. Systematic plans for AI implementation management can improve the Organizational challenge of XAI in healthcare. Communicational issues can be resolved by the doctor's awareness on how the patients will perceive the system and by double-check of health information with patients. The socio-organizational issue or challenge can be resolved by patient education that will be helpful in AI usage [21]. The paper has identified the need to work on the improvement of abstraction and lack of explainability of models by including feature importance, which will help in prediction or classification of ML models [22]. Another study proposed developing more transparent models for major diseases such as diabetes and cancer, as well as doctors and health professionals with basic AI knowledge, in order to achieve the goal of XAI [23].

6) WHAT IS THE RESEARCH GAP EXISTING IN XAI IN HEALTHCARE?

Regarding the advancement in the field of XAI in healthcare, some gaps still exist in the literature. Gaps identified in the literature include the development of the XAI model capable of XAI methods that should be useful for the end-users and clinical expertise in the medical domain or normal users or individuals. Interface development for XAI for medical domains is still a challenge. The model agnostic AI model is still an open area of research [24]. Despite XAI models advancement and working in the healthcare domain, transparency remains an issue that needs to be worked on in the future for improvement in the models [25]. The study in [23] has given importance to focusing more emphasis on studying uncommon diseases for etiologies in predictive analytics to prevent extensive and expensive workups, among other things. The researchers have given the focus on using XAI to assist medical professionals to overcome their medical knowledge biases.

7) WHAT IS THE FUTURE OF XAI IN HEALTHCARE?

XAI is a new field that was started in 2017 by DARPA [26]. The field is now emerging and is integrating into many of the new fields. The research related to XAI in healthcare has many challenges and gaps described in (e) and (f) above that can be worked in the domain of healthcare. The other thing is that XAI is a relatively new field that can be worked on and can provide many future pieces of research.

B. RESEARCH CRITERIA

Well-known researchers from the past have performed s systematic analysis of HCI, AI, XAI, and their different aspects. The following research string was used when the literature survey was performed for this paper Human-Computer Interaction, HCI, Human-Centered Design, Usability, HCI in healthcare, Artificial Intelligence, AI, Deep Learning, Machine Learning, AI In healthcare, Explainable Artificial Intelligence, XAI, XAI IN HEALTHCARE, and Challenges of XAI in healthcare.

C. DATA SOURCES

Several different data sources were used for searching the literature. Figure 1 above shows the percentage of papers taken from each data source. The research papers that were found include journal papers and conference papers in Google scholar, books, Scopus, and blogs. The databases that were used in the search are mentioned in Table 1. Below.

D. EXPLORATION CRITERIA

Research for this paper was performed from 2016 to 2021. The paper that was explored was evaluated first according to the criteria and keywords to be included.

Figure 2 below shows the percentage of papers read between 2016 and 2021.

E. EXCLUSION & INCLUSION CRITERIA

The articles that were selected include the HCI field and its information on subfields and the problems that are being encountered in these fields. The key terms of HCI, Usability and Human-Centered Design were used for selecting articles in the HCI domain. For the Field of Artificial Intelligence key

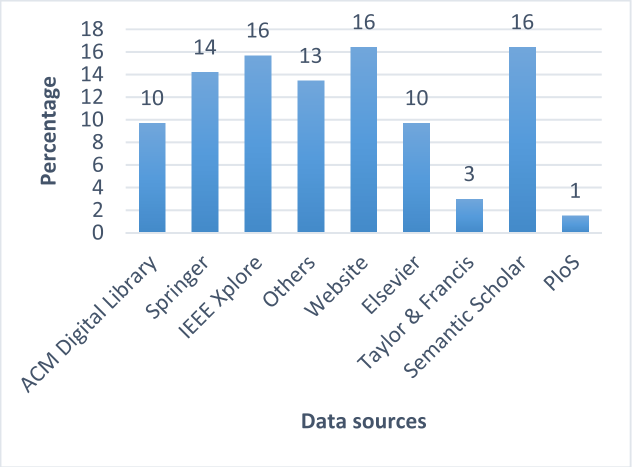

FIGURE 1. Percentage of Research Papers from Data Sources.

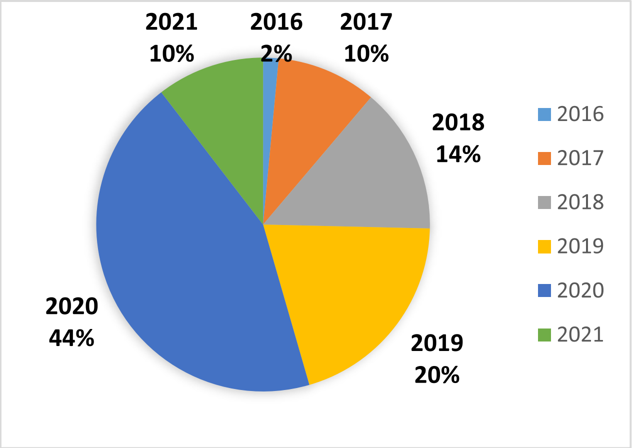

FIGURE 2. Percentage of Paper included between 2016 and 2021.

TABLE 1. Database engines.

Database Engines	Sources Address
IEEE Xplore	https://ieeexplore.ieee.org
ACM Digital Library	https://www.acm.org/
Semantic Scholar	https://www.semanticscholar.org
Taylor & Francis	https://taylorandfrancis.com
Springer	https://www.springer.com
Elsevier	https://www.elsevier.com

terms of AI, Deep Learning, and ML were used for defining the introductory part of this field. Explainable Artificial intelligence domain papers were selected by the key term XAI and by including the keyword of health care along with it and the challenges and gaps of this field. Prisma model below in Figure 3. Shows the overall paper reviewed for this paper, the paper included, and provides the number. of papers according to each domain.

III. BACKGROUND OF RESEARCH

In this section, HCI and Artificial Intelligence will be explored in terms of the fields in which they are working. This section will first give the basic understanding of HCI, and after that, AI and its related fields will be discussed in this section.

A. HUMAN COMPUTER INTERACTION

Human-Computer Interaction (HCI) is one of the fields that is emerged and has been successful in both the fields of computer science, and psychology & cognitive sciences [27]. HCI is also contributing to other fields of ergonomics, sociology, graphic design, and business. HCI helps human beings to understand and interact with and through technology by providing a good means of communication [28]. Figure 4 shows the detailed illustration of HCI and its fields, along with their subfields and functions of each subfield, which is the aim of completing them. HCI aims to improve the interaction between users and computing devices. The main aim supported by HCI is to provide interaction following the needs and capabilities of users [29]. The easy structure of communication is mainly supported by technology. Psychology's role in HCI includes a general framework for the interaction of human beings with systems and software, and it includes verifying the usability of the system and software after it is developed [30].

Problems that are covered by HCI include better describing design and development work for understanding. The other thing is to better describe the role that psychology, in particular, social and behavioral science broadly plays in HCI [31].

One of the factors for which HCI has been given importance in every field is improving the visual design for the interaction of users and the cognitive abilities of users. The visual design of the computer system and applications can be improved by applying usability [32]. Although ISO does not directly provide a standard for HCI, it does define HCI in terms of usability and human-centered design [33].

1) USABILITY

Usability is considered to be one of the factors that impacts the user's decision to use your system [34]. If your interactive system does not provide the required product or service to the users of the system, the users will search for alternative options for better results and services which will provide them usability [35]. In the past, many evaluation methods have been developed to determine the usability of an interactive system. The methods that were developed previously only focus on the usability problems and some of them provide numerical value about the usability of a software product. There are two methods in which the usability of the system can be justified includes qualitative assessment and quantitative assessment. Qualitative assessment includes heuristic evaluation and cognitive walkthroughs, which mainly allow the identification of issues in interface design. In quantitative assessment, software metrics are used to determine the

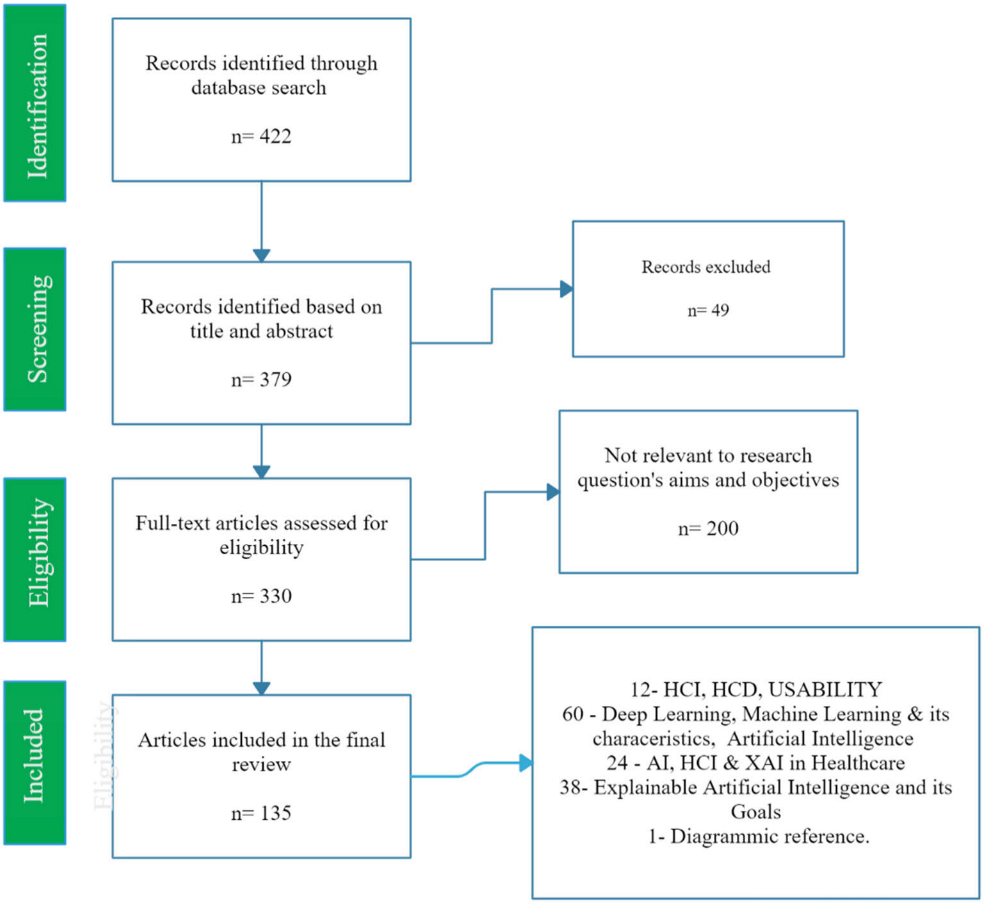
Identification
Screening
Eligibility
Included
Records identified through database search
n= 422
Records identified based on title and abstract
n= 379
Records excluded
n= 49
Full-text articles assessed for eligibility
n= 330
Not relevant to research question's aims and objectives
n= 200
Articles included in the final review
n= 135
12- HCI, HCD, USABILITY
60 - Deep Learning, Machine Learning & its characeristics, Artificial Intelligence
24 - AI, HCI & XAI in Healthcare
38- Explainable Artificial Intelligence and its Goals
1- Diagrammic reference.

FIGURE 3. Prisma Model Depicts no. of record included and excluded.

usability of a software product by identification of the aspects in which software needs to be more intuitive and functional [35].

Software metrics are used as the primary technique to quantify the usability of the system by the user's task performance. The user is performing the task, and the observer is taking notes on the satisfaction level of the user [35]. ISO has set many standards for the evaluation of software products and usability metrics calculation of the software system. ISO 9126 is used for the inspection or evaluation of software products in a specific domain. While the other standard of ISO 9241 is used for providing the usability metrics, it provides some factors that need to be checked for calculating the usability of the system. Some of the factors that need to be focused on the usability metrics include effectiveness, efficiency, and satisfaction [36].

a) **The effectiveness** of the system can be measured through the user performance of the system. In other words, effectiveness can be defined as the level at which the user can properly and fully achieve the stated goal [37].

$$\textit{Effectiveness} = \frac{\text{No. of Tasks accomplished}}{\text{Total No. of Tasks}} \times 100$$

b) **Efficiency** can be defined as the time taken by the user to achieve the specified objective [37].

$$\textit{Efficiency} = \frac{\sum_{j=1}^{R}\sum_{i=1}^{N}\frac{n_{ij}}{t_{ij}}}{NR}$$

where,
N = Total No. of Goals
R = No. of Users
nij = Task result i by user j
tij = Time to complete the task i by user j (in seconds)

c) **Satisfaction** or success rate can be defined as the extent to which user requirements are fulfilled by using the system for a particular task [37]. It defines the success rate of the system.

$$\textit{Success Rate} = \frac{S + (Tu \times 0.5)}{Tp} \times 100$$

where,
S = Successful attempts
T_u = Total User
T_p = Total attempts for performing the task.

These are some of the ways discussed above to calculate the usability of the system, from the user's perspective.

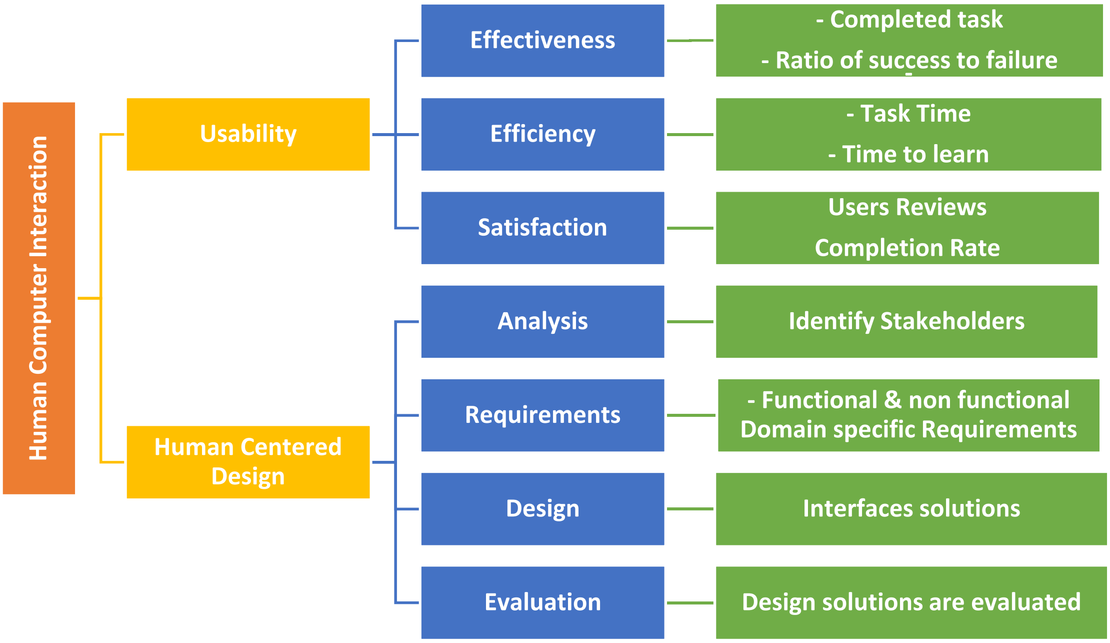

FIGURE 4. Human-Computer Interaction.

For evaluating the usability of the system these things can be calculated manually or with the help of integrating them into the application being used by the end-users.

2) HUMAN CENTERED DESIGN

Human-centered design (HCD) is a field of HCI in which methods have been developed to understand people, culture, and co-evolution of these factors in technology. It is a field that focuses on the development of an interactive system on making the system usable. In other words, we can define HCD as a process in which systems understand the perspective of how people think to design an effective system [38].

The term Human-Centered Design (HCD) is named or grown as user-centered design (UCD) due to the intersection of psychology and artificial intelligence. HCD, also known as Design Thinking (DT), and UCD can be classified as the same thing [33].

Human-Centered Design or HCD is divided into four phases. In the first phase, stakeholders for the system or product that is being developed are identified with the help of the context of use. In the second phase, after the analysis has been done, the functional & non-functional requirements are assembled, which can also include domain-specific requirements. Design solutions or interfaces are collected in the third step of HCD. When the design solutions are finalized, those design solutions are evaluated in the last phase [39], [40].

When we talk about the HCD phases, certain problems may occur and require certain steps to be performed. The possible continuation includes an analysis of context to be done for a second time if serious problems are occurring or tend to occur. If the analysis step has been done correctly but the requirements do not seem to be according to the domain or not specified according to functional & non-functional requirements, the step needs to be performed again. The third possibility that may occur is improving the design solutions [8].

When we talk about usability and human-centered design, both fields are subfields of HCI and have many similarities. The focus of both of the fields is to provide ease to the user. The difference between the fields is that usability is performed on interfaces to check whether the particular interface is efficient and effective to achieve the satisfaction level of the user [35]. On the other hand when we talk about Human-Centered Design, its main focus is to go through some of the steps to achieve the design that will be made according to the user's needs or expectations [41].

B. ARTIFICIAL INTELLIGENCE

The main concept or idea behind Artificial Intelligence (AI) is to understand intelligent entities. Many definitions address AI in different terminologies and frameworks. In 1990, Kurzweil defined AI as "the art of creating machines that perform functions that require intelligence when performed by people". The other idea of AI was proposed by Winston in 1992, who defined AI as "the Study of the compilations that make it possible to perceive reason and act". In 1993, Luger defined AI as "The branch of computer science that is concerned with the automation of intelligent behavior" [42].

Artificial Intelligence means the study of intelligent agents, which means a device that perceives its environment and takes action, which is the reason for maximizing its success in the goal. In other words, we can say that AI works as an intelligent agent that takes the best possible action in a situation [43]. AI is working in many fields, like network

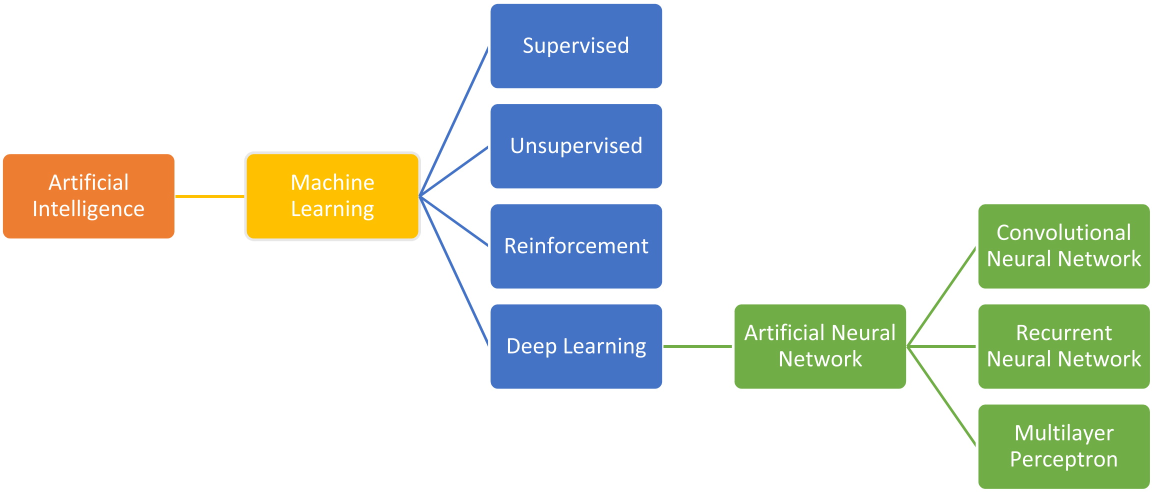

FIGURE 5. Artificial Intelligence.

load balancing, smart agriculture strategies, security, livestock, inventory management, and manufacturing and production [44]–[46]. Figure 5 briefly summarizes artificial intelligence its subfields, and their further division into other subfields. Some of the fields related to AI are discussed below:

1) MACHINE LEARNING

Machine Learning (ML) is the subfield of AI that discovers and constructs algorithms that can learn from and make predictions on data. It is also said that ML gives the computer the ability to learn without being explicitly programmed. ML is used in those areas where designing and programming explicit algorithms with good performance is difficult or unfeasible. ML includes applications like email filtering and the detection of network intruders. ML also deals with computational statistics which focuses on prediction making through the use of computers [47]. ML includes the four widely used methods discussed below [48]:

Supervised: Supervised learning algorithms are used where the desired output is known and is trained using labelled data.

Unsupervised: In unsupervised learning, no historical labels are used and the algorithm has to figure out what is being shown.

Semi-Supervised: Semi-supervised algorithms are used for the same applications as supervised learning. It includes both labelled and unlabeled data for training.

Reinforcement Learning: The reinforcement learning algorithm works on the mechanism of the best policy. This algorithm includes three primary components the agent, the environment, and actions. The objective is for the agent to choose actions that make the most of the expected reward over a given amount of time. The agent will reach the goal much faster by following a good policy.

a: DEEP LEARNING

Deep learning is a field of ML, known as deep machine learning, deep structured learning, and hierarchical learning. In deep learning, features are extracted by deep learning itself without human intervention. This technique is inspired by the structure of the human brain known as an artificial neural network [49].

Artificial Neural Network (ANN): Solves the problems that would prove impossible or difficult by human or statistical standards. ANN can be said to as a piece of computing system design intended to simulate the way the human brain analyzes and processes information. ANN is further divided into three classes known as Multilayer Perceptrons (MLPs), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) [43].

Multilayer Perceptrons (MLPs): They are used to distort the input space to make the classes of data linearly separable. They are known as feed-forward neural networks, with a small set of requirements for data set consisting of three layers of the dataset with inner, hidden, and outer layers [50].

Convolutional Neural Networks (CNNs) or ConvNet: Are designed in the form of multiple arrays. An example can be taken as a simple image containing three 2D arrays. CNN's has many applications in the fields of image, video processing, natural language processing, and recommender systems [43], [50].

Recurrent Neural Networks (RNNs): Tasks that involve sequential inputs in the form of speech or language, in such cases, recurrent neural networks are best used. It is used in applications for speech recognition, machine translation, and language modeling [50], [51].

IV. CHARACTERISTICS OF MACHINE LEARNING

The section mainly covers the main characteristics of machine learning and provides a literature review of the fields in which machine learning and its characteristics are

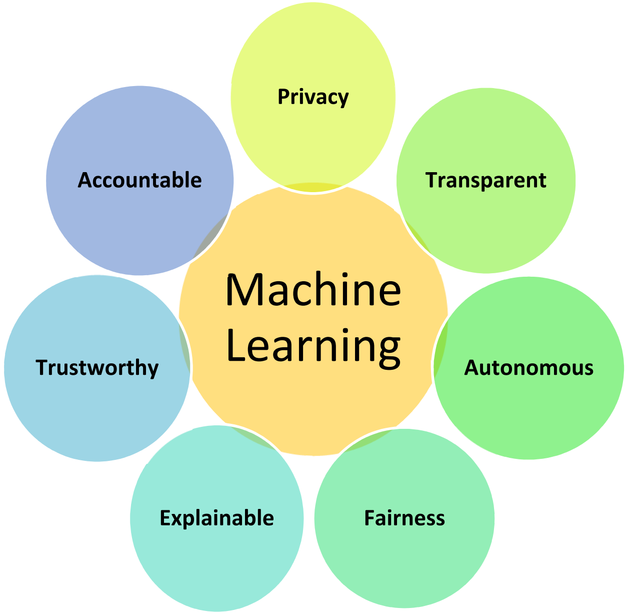

FIGURE 6. Machine Learning & its Characteristics.

being instigated. Figure 6 summarizes the features of machine learning in the form of a diagram.

The characteristics are those entities that are the main concern or said to be the things needed to be completed by the specified model. Some of the main characteristics that are predictable from ML Includes privacy, accountability, transparency, fairness, trustworthiness, autonomous and explainability. A detailed discussion of these fields and their expected characteristic consumption is discussed below. Table 2 at the end of this section summarizes the characteristics of machine learning-focused in the literature in different fields, along with the algorithms and techniques of artificial intelligence that have been used.

A. ACCOUNTABLE

In terms of AI or ML, accountability can be defined as terminology which is directly connected to the consequences. In some factors, the system cannot be held accountable for all the actions, and at every stage, AI or ML can be involved in taking any accountable action [9]. For achieving accountability in AI systems, many pieces of research have been done, which are being discussed below. The MapReduce Model is an existing model which is used as a powerful parallel data processing model. It is used for solving large-scale computing problems. The paper [52] proposes an Accountable MapReduce Model in place of the existing model which employs an auditor group to conduct an A-test on every worker in the system. The model is also able to detect malicious behavior if it occurs. To improve the performance, P-Accountability is introduced as a means of trading the degree of accountability with efficiency.

The study in [53] was done for the ethical accountability of applications. The basic idea behind the study was to identify, analyze and explain the ethical consequences that can result from the datafication of services. The study used a midrange conceiving approach of the presently disconnected perspectives on technology-enabled services, data-driven business models, data ethics, and business ethics to introduce an innovative logical framework centered on data-driven business models as the general metatheoretical unit of analysis. The resulting midrange theory offers new insights into how using machine learning, AI and big data sets can lead to unethical implications. – Future research based on the framework can help guide practitioners to implement and use advanced analytics more effectively and ethically with the declaration of accountability.

The paper [54] focused on four aspects of machine learning, which include fairness, accountability, trust, and privacy. Federated learning allows the development of a machine learning model among collaborating agents without requiring them to share their fundamental data. The study has used federated learning to close the gaps in trust and accountability by developing a Block Flow system. The main theme or idea behind developing this system was providing a federated learning platform. The future work of the system is recommended by providing improved performance.

Research conducted in [55] has proposed a system called BlockFLA based on the blockchain framework. The main purpose of the proposed model is to achieve decentralization and transparency by integrating public and private blockchains. The other purpose of this is to provide accountability to the responsible parties for discouraging backdoor attacks on federated learning algorithms. The proposed framework facilitates the adoption of any federated learning and it also provided the implementation of Federated Averaging and SignSGD our general blockchain framework and the empirical results conclude that the proposed framework maintains the communication-efficient nature of algorithms. For future purposes, the researchers have decided to develop s Trojan detection algorithm for SignSGD.

As machine learning spreads in an open world, system accountability has become a top priority. In the paper [56], Pandora, a hybrid component-based human-machine method and tools, have been presented for describing and explaining system failures. The evaluation of the system is done with an image captioning system. Results show that Pandora provides information about the failures hidden in the statistics of traditional metrics. For future purposes, the research proposed the conceptualization of views that can jointly cluster different types of failure.

B. AUTONOMOUS

The Factor of autonomous is an intelligent class of ML which works on Autonomous sensing, decision making and work in a dynamic environment. It works in the field of intelligent software agents and autonomous vehicles [10].

Many pieces of research have been done in the field of Autonomous vehicles with machine learning. One of the research studies [57] has proposed simulation-based

TABLE 2. Characteristics of machine learning and algorithms/techniques.

Characteristics References	Accountable	Autonomous	Fairness	Transparent	Explainable	Trustworthy	Privacy	Algorithm/ Techniques
[52]	✓							Random Forest
[9]	✓		✓	✓	✓			Bootstrap Methods
[68]						✓	✓	Convolutional Neural Network, Pattern Refinement Algorithm
[63]					✓			Brain Imaging technique
[64]					✓			Genetic Algorithm, Regression Algorithms
[12]				✓	✓			Genetic Algorithm, Logistic Regression, Support Vector machine
[65]					✓			Reinforcement Learning Algorithms, Convolutional Neural Network, Artificial Neural Network
[53]	✓							Topic Modeling Technique
[54]	✓		✓			✓	✓	Differential Privacy, Multi-party computation, Blockchain-based Accountability.
[70]							✓	Deep Learning Models
[14]							✓	Cryptographic Approaches, Homomorphic Encryption Techniques, etc.
[60]			✓					Risk Assessment Technique
[61]			✓					Adversarial bias framework
[57]		✓						Optimization based Algorithm
[58]		✓						Signal Temporal Logic
[59]		✓						Deep Neural Networks
[69]					✓	✓		Blockchain-Based Framework
[62]				✓				Active Learning Algorithm
[55]	✓			✓				Blockchain-Based Framework
[56]	✓							Decision Tree
[66]					✓			Convolutional Neural Network
[67]					✓			Random Forest
[71]							✓	Neural Networks, Logistic Regression, Linear regression.

adversarial test generation to be used for the testing of autonomous vehicles. The paper has provided the idea of testing for autonomous vehicles due to some of the challenges that exist related to testing, debugging and certifying the performance of the autonomous driving system. The study has proposed a testing framework that is compatible with test case generation and automatic falsification methods for evaluating cyber-physical systems. The framework proposed evaluated closed-loop and test case generation methods. The study concludes that the framework could be used in the future to increase the reliability of autonomous driving systems.

Study in [58] has focused on providing a testing framework for signal temporal logic (STL) for component level and system-level behavior. The key component of the research is that ML components are being supported in the system design, such as deep learning. The framework given in the research is used for the evaluation of test cases and also automatically discovers the test cases that have failed the requirements. The study demonstrated a simulation-based adversarial test generation framework for autonomous vehicles. The study is useful for finding new ways to find the critical behavior of vehicles.

The literature review [59] on the emerging field of AI and autonomous surgeries was conducted, with a focus on the legal, regulatory, and ethical aspects of AI and autonomous robotic surgery. The paper also provides discussion on the responsibilities and classification which includes Accountability, Liability, and Culpability. Accountability has been discussed in the current reflection as Blackbox. Liability has been discussed as what has already been done for surgical robots. Culpability in the context of literature is less clear because its capabilities are far beyond the current technologies. The study showed that in the future, robots will learn from the humans and will perform the daily operative tasks.

C. FAIRNESS

Fairness is a field of ML that studies how to ensure that partialities in the data and model imprecisions do not lead to unfavorable models. Unfavorable models are those that treat individuals based on characteristics such as e.g. race, gender, and disabilities [11].

In [60], researchers have shown the existing definitions and criteria of fairness which include classification, anti-classification, and calibration. These suffers from some significant limitations which may cause harm to the purpose for which they have been designed. The research shows that in past literature, many people have used one of these criteria for the evaluation of their systems or used one of them as a constraint when developing new algorithms. The researchers have provided the consequences of the system for future fairness, which include randomized control trials and decoupling the statistical problem of risk assessment. The study has provided a risk assessment to help researchers and practitioners productively work in the area of fairness.

In other research [61], researchers have proposed an adversarial bias framework for data positioning attacks against fair machine learning. Adversarial bias adds a fairness gap to the test data, and these are said to be attacks in which something is added intentionally into the model to make a mistake. The experiment conducted in the research concluded that adding a small adversarial sample/data point can reduce the model accuracy which can easily be achieved in an unconstrained model.

D. TRANSPARENCY

Transparent or transparency of a system is enhanced by explainable systems. When we talk about transparency, its main aim is to provide the end-user with information on how a model works. The form of providing information to the end-user can include publishing the algorithmic code and disclosing the properties of the data set and training procedure [12]. Many of the studies conducted in the past have communicated explainability and transparency together. Some of the studies have been discussed in this section.

In [62], a robot was built that could leverage transparency and help the teacher to provide better instructions. The study suggested that active learning is characteristically a part of the transparent machine learning approach. The study provided an active learning approach implemented on Simon Robot. The pilot study firstly indicates the potential for transparency in active learning that will help improve the accuracy and efficiency of the learning process. The study also shows some of the undesirable effects which need to improve by applying control strategies for social learning interaction.

E. EXPLAINABLE

Explainable or explainability is the term that is nowadays referred to as being used as a technique that helps users of machines understand why the model is working or behaving the way it does [12]. The term can be used in many forms and many fields. This section will focus or provide the knowledge of some of the recent studies that provide the knowledge of explainability or explainable systems and the fields in which they have been used in the past and can be used for future purposes.

Counterfactuals, defined in [63], is a concept that is being used for prediction to explain past outcomes and for prediction of what could happen in the future. It is used in AI applications and is now also used in Explainable Artificial Intelligence (XAI). The paper discussed counterfactuals and their causes, as well as fault management. The study suggested incorporating psychological experiments in XAI about the knowledge of what people create and comprehend counter-culturally.

The research in [64] provided a systematic literature review on the generation of contrastive and counterfactual explanations. Based on the survey, a state-of-the-art computational framework was stated. The fields of contrastive and counterfactual are found to be in great demand across different fields of AI mainly in Computer Vision and Natural Language Processing. The survey first examined the theoretical foundation of contrastive, computational, and contrastive-counterfactual frameworks. The survey provided the various properties of areas of study and also provided the shortcomings of the fields that can be improved in the future.

Another study in [12] explores the use of explainability for the consumption of stakeholders by organizations. The study shows that there is a gap between explainability and the goal of transparency. In a study conducted, the mechanism has been observed to be different among different organizations, deployment explainability algorithms. The interviews were conducted in two groups to assess the explainability. The study has provided the limitations based on the results of the current techniques, which include feature importance, Adversarial Perturbation, and counterfactual explanations that prevent the model from performing its functions to end-users. The study has suggested overcoming the limitations of a domain expert to evaluate explanations, the risk of spurious correlations reflected in model explanations, the lack of causal intuition, and the latency in computing and showing explanations in real-time.

A study in [65] has presented a framework named explAIner for interactive and explainable machine learning, grasping the theoretical and practical state-of-the-art notions in the field. The study has focused on the XAI field as its core concept or framework. For explaining the XAI framework, the XAI pipeline was described in this study. Which includes an interactive workflow of three stages: understanding of the model, diagnosis, and refinement. The XAI pipeline, along with the framework global monitoring and steering mechanisms by providing provenance tracking, was implemented in the framework presented in this study. The user study with nine participants was performed to check the usability and usefulness of the tool presented. The presented framework was found intuitive by the users and considered to be combined in their daily workflow.

The paper in [66] has demonstrated that the subfield of artificial neural network, CNN, can serve as an ultra-fast electromagnetic simulator for predicting and explaining the optical properties of nanophotonic structures with remarkable precision. The algorithm that has been demonstrated accordingly is Deep Shapely Additive Explanations, or SHAP which identifies the contributions of individual image features. CNN Predicted unknown structures with an accuracy of 95%. The offered explainable artificial intelligence method shows that the patterns and principles encoded within the ML model can be extracted to originate valuable intuitions into the nanophotonic structure behavior, even in complex freeform structures whose behavior is typically not easy to understand.

A study conducted [67] is used to compare different ML algorithms on a dataset of 43,420 oxide glass compositions and their respective glass transition temperature. The use of the ML Algorithm in this research is an investigation of different ML algorithms for predicting the glass transition based on their chemical composition. Six different algorithms were used for this research. The result of the study indicates that Random Forest (RF) provides the best prediction performance along with the best visual explainable model as compared to other algorithms. RF provides an explainable model which provides individual importance of chemical elements for developing glass with very low and very high glass transition. The future of the study proposed that this study can be further used for the prediction of other composition property combinations.

F. TRUSTWORTHY

A trustworthy system is said to be the system in which the user feels safe and secure while using it. It is said in one of the papers that some of the technologies need to be considered while ensuring the trustworthiness of the system. It includes fairness, explainability, auditability and safety [13].

Research in [68] has proposed a new framework named PriModChain which is used for trustworthy ML and the Industrial Internet of Things (IIOT). For assuring the aspects of privacy and trustworthiness. Federated Learning is used as an ML model and differential privacy imposes privacy on ML models. In this paper, five pillars of trustworthiness (security, privacy, reliability, safety, and resilience) have been given importance for making it a feasible solution.

The study [69] has proposed a blockchain-based framework for achieving more trustworthiness and XAI by leveraging features of blockchain, smart contracts, trusted oracles, and decentralized storage that can reduce biases and adversarial attacks. The study has focused on the lack of explanation of the AI algorithms that are critical for decision making. The proposed model in the research can be used to develop more trustworthy decentralized AI & XAI systems and applications.

G. PRIVACY

Privacy is about protecting the information of users against unintended information outflow. Such as unauthorized use of services, data leaks, and privacy compromised due to unpredictable providers and some accidents [14], [15].

In [70] Chiron, a new system design, implementation, and evaluation has been done for privacy-preserving outsourced machine learning (ML) has been presented. Chiron is implemented through Software Guard Extensions (SGX) enclaves that enable data holders to use ML-as-a-service without disclosing their data to the service providers. The Chiron platform is different from the existing ML platform since it does preserve the privacy of the service provider by securing their data. The evaluation of the model was done through deep learning models shows that its training performance and accuracy of the resulting models are practical for common uses of ML-as-a-service.

The research conducted in [14] has provided the intersection between machine learning and privacy by providing techniques that can be used for data protection. The paper addresses some of the issues and techniques related to privacy preservation. The techniques discussed in the study include Cryptographic Approaches, Homomorphic Encryption, Garbled Circuits, Secret sharing, secure processors, Perturbation Approaches, Differential Privacy, Local Differential Privacy, and Dimensionality Reduction. Although privacy techniques can be helpful in privacy-preserving, some of the issues are there, which include flexibility, scalability, security, and privacy policies.

SecureML was introduced in [71] for the privacy-preserving concern for Machine Learning. According to their research, the privacy of the data becomes a challenge when the amount of data is large. For distributing the data and training it accurately, the system was implemented in C++ with a new technique for the support of secure arithmetic operations. A new and efficient protocol was introduced in the paper for preserving ML for linear regression, logistic regression, and neural network training using the stochastic linear descendent method. The working of protocols takes place in a way that protocols fall between two servers. The server data is distributed between two non-colluding servers by the owner. The models are then trained on data using secure two-party computation (2 PC). Experiments conducted recently show that protocols introduced in research are several orders of magnitude faster than state-of-the-art implementations for privacy-preserving linear and logistic regressions, and can scale to millions of data samples with thousands of features.

V. HUMAN COMPUTER INTERACTION & ARTIFICIAL INTELLIGENCE

HCI and AI work hand in hand in such a way that AI mimics human beings to build intelligent systems, and HCI attempts to understand human beings to adapt the machine to improve safety, efficiency, and user experience. AI focuses on the internal mechanism of intelligent systems and HCI focuses on the fundamental phenomenon of interaction among people and tools. AI fieldwork to create intelligent interfaces that provide the capability to perceive, act and learn by themselves. On the other hand, HCI is focused on usability, creativity, and innovation of the system. This section addresses some of the papers that have focused on the fields of HCI and AI together as both of the fields overlap each other in various domains and the technologies that are developed using both of these.

The paper [72] provides the gap identified between HCI and ML. The Gap mainly focuses on how to improve the design in ML for user experience (UX) Using HCI. 2494 research publications related to HCI were analyzed for this, and 3 under-explored areas were identified through this paper that can help to prepare for design innovation. The main thing that was identified during the research was that UX and ML are not addressed together. 9 papers have mentioned machine learning, 3 papers have described or given importance to the design in ML system. Two of the groups were created in this research in which the first group identified the well-established ML topics. The second group identified the under-explored ML topics. The study, after analysis, provided seven clusters of HCI concerning ML: 1) intelligent UI and usability; 2) intelligent environment; 3) recommenders and user modeling; 4) social network and sensor framework; 5) AI and knowledge systems; 6) search and deep learning) and 7) sentiment analysis and affective computing. The analysis identified two clusters of ML technical advances that have not yet been destined to particular conveniences, interactions, or user experiences. The first is sentiment analysis, the second is social network mining.

In [73], an AI-infused orchestration system named Formative Assessment Computing Technologies (FACT) was developed for teacher management of classroom workflow that mixes small groups, individual, and whole-class activities. FACT technology can provide a user experience for students that they can use on a desktop, laptop, or tablet with the help of a web browser. Students are also provided with the right to edit, draw, type, erase, or move the activities they are performing, similar to other online collaborative editors like Google Docs. The usability of the system is also being evaluated in the testing method by working on two of the factors, including time consumed in FACT and the amount of time wasted in the paper, what are the failures that are occurring in FACT and how frequently do they happen? Management on the teacher's side can be done with the help of a tablet. AI intelligence is involved in FACT in such a way that it monitors the students and edits and updates teacher dashboards in real-time to show progress and alerts.

Another paper published in [74] addresses the association of HCI and AI in different systems. The paper proposes that an intelligent system cannot perform its functions properly without a concept and design based on solid HCI principles. This paper reviews the following domains: intelligent user interfaces, and more specifically, conversational animated effective agents; capitalization, formulation, and use of HCI ergonomic knowledge for the design and evaluation of interactive systems; and synergy between visualization and data mining.

In [75], Based on a proposed psychological model in the past and improvement in the intelligent system, an Intelligent interactive computing model based on the human-computer cooperation mental model is proposed. For designing the proposed model, the first module was natural interaction behavior characteristics, as the design principles are used for the interface module. The second module includes the perceptual module, which was based on the human-computer cooperation mental model and includes sensations, attention, and perceptions. It provides the acquisition of multi-modal interactive data and, based on this, the user task intention is extracted and forwarded to the cognitive computing module to solve the task, which was the third module of this model. The cognitive computing module receives the task and solves the interaction task. Task solving is dependent on knowledge experience that is based on ML models for self-learning and knowledge updates. The results are obtained and handed over to the action response evaluation module. which is the fourth and last module of the proposed model. It provides feedback of the task processing results to the user in a natural and suitable form of representation.

The paper presented in [76] proposes that AI and HCI are supposed to be the era of intelligent information. Many commercial applications have an intelligent user interface that is presented in this paper. HCI and AI as the core fields include applications like natural gesture interaction, emotional computing, and voice dialogue systems. The human-machine dialogue-based technologies that are also named in this paper. including Apple Siri, Microsoft xiaoic, Google home, and amazon are solving human-machine dialogue. The paper also includes a part that focuses on some of the research publications done in the past that focused on combining research in HCI and AI and their future scope. This paper summarizes the ways in which we anticipate that AI and HCI research will continue and even increase, and is a useful area of research.

The paper in [77] explains the history of HCI and some of the core issues addressed through the use of HCI in the interface. New thinking or ideas regarding HCI are given by describing that HCI is not said to belong only to the perspective of interface or GUI but also to its transit to the natural user interface. The interface should be more humanized that will be easy and comfortable from the user's point of view. It is also discussed from a research perspective that increased use of HCI in commercial products is also increasing for use with AI. as discussed in some names in [76].

The paper [78] also provides the knowledge and use of HCI with AI technologies. The purpose that was covered in this paper is to provide an understanding and assessment of how HCI knowledge can be used in studying new interaction modalities in intelligent built environments to make advancements in the AI field. In this direction, it is intended to review the applicable research methodologies that can be derived from the HCI research community to envision the gradual change in human experiences with and within buildings alongside the advancements in information, communication, sensor, and actuation technologies.

DARPA's Explainable AI system proposed in [26] is a program that endeavors to create AI systems whose learned models and decisions can be understood and appropriately understood by end-users. The main idea behind XAI technology is to provide a variety of new ML techniques. With the advancement in modified DL techniques that learn explainable features; methods that learn more structured, interpretable, and causal models. Results indicate that these three broad strategies merit further investigation and will provide future developers with design options covering the performance versus explain-ability trade space.

VI. EXPLAINABLE ARTIFICIAL INTELLIGENCE

The boundary between HCI and AI is narrowing these days, and new technologies are being developed to bridge the gap between the two. To close the gap, a concept known as Third-wave AI was introduced in several trends, including Human-Centered AI (HAI), Useful and Usable AI, and Explainable Artificial Intelligence (XAI). In HAI, AI is used to enhance humans rather than replace them. It focuses on technically reflecting the depth characterized by human intelligence to improve human capabilities rather than replacing them and focusing on AI's impact on humans. Usable AI can be defined as an AI solution that is easy to learn and use via optimal UX created by effective HCI design. XAI enables the users to understand the algorithm and parameters used. The main concern with it is to address and resolve the AI-black box problem. A black box problem or phenomenon causes users to question the decision made by the system. 1) Why did you do this? 2) Why is this the result? 3) When have you succeeded and when have you failed? 4) When can I put trust in you? The main things that concern XAI are human trust and decision-making efficiency. DARPA's XAI program was introduced in 2017. Its main concern was to develop new or improved XAI algorithms. The program also investigated user interfaces (UIS) with advanced HCI techniques (i.e. UI visualization and Conversational UI)[78]. Figure 7. Below demonstrates the overall view of XAI.

XAI has gained popularity because it was launched by the Defense Advanced Research Projects Agency (DARPA), which aims to create new techniques to understand, appropriately trust, and effectively manage these artificially intelligent partners [26]. XAI is divided into three stages. Explainable construction process, explainable decisions, and explainable decision-making process. In the first stage, explainability is key to improving the adoption and performance of AI applications. In the second stage explainability of AI, decisions have been made to build trust with users and supervisors. The third stage is regarding the interoperability of AI systems with each other and other software [19].

A. STEPS OF XAI

The main purpose behind XAI, given by DARPA was to create a set of ML techniques that produce more explainable models for humans to understand and trust emerging AI systems. XAI impact is characterized by three stages, starting from the higher level, which has a great impact, to the lower level, which is the lightest [19].

The paper in [19] focused on systems that are built based on ML, but the proposed three stages are relevant to any class of AI. The three stages are presented and are discussed below, along with an illustration in Figure 8.

1) EXPLAINABLE BUILDING PROCESS

In the first step of explainable artificial intelligence, the process for the improvable implementation and performance of AI applications. In this step, the main focus is creating a team of experts who can provide the knowledge related to the field of artificial intelligence they are working on. The other focus is given to the visualization techniques, as they provide a scheme of the output against the context variable to get a feel for how an artificial intelligence performs over the target domain. In this step, debugging tools or techniques are also discussed. Tensor Flow Playground, ConventJS, and Seq2Seq are examples of tools originally designed for data scientists.

2) EXPLAINABLE DECISION

The second step explains every AI decision taken by the system to build trust with users and supervisors. Trust is a key priority of a system. In the explainability decision, it has

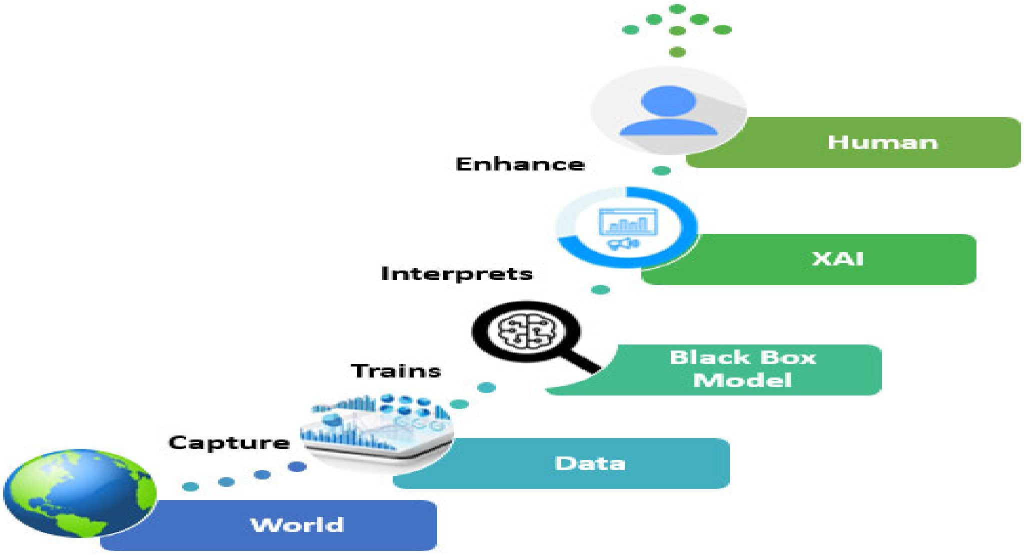

FIGURE 7. Explainable Artificial Intelligence (XAI) [79].

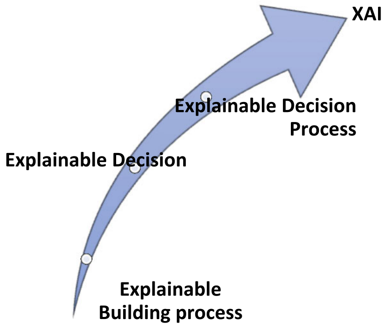

FIGURE 8. Steps of XAI.

been focused that the system is not surprising and it behaves according to the mental model of users. The ability to explain AI decisions is an active field of innovation with methods such as Tree Interpreter, LIME, and SHAP. The main purpose of this stage is explanations needed for the business value of AI.

3) EXPLAINABLE DECISION PROCESS

The third step of XAI is used to enable the interoperability of AI systems with each other and other software using business logic.

B. XAI GOALS

Explainable artificial intelligence is working in many fields, like HCI, security, and business processes. As the field is explored and used in many fields, its goals have started to expand in many fields. XAI also has goals that somewhere touch the areas of AI and ML classification and the expectation of users in these areas. The goals of XAI are described as the expectations of the intended audience from the system which is achieving providing XAI capability [80]. The main goals of XAI are given in Figure 9. Below.

1) TRUSTWORTHINESS

It is considered to be one of the major purposes to be considered for explainable AI. For achieving explainable AI, trustworthiness is one of the main concepts that is measured. Trust of the system will be considered to be the intended behavior expected from the system as the user performs the function or task on the system. A trustworthy system is important for individuals who may be affected by the decisions of the model or domain experts [7], [81].

2) CAUSALITY

The impact of causality involves correlation. Its basic necessity in explainable AI is to find a causal relationship among involved variables. For witnessing the causal relationship, prior knowledge is mandatory for observing and acknowledging that the effects are causal. The main targeted audience for checking the causality among variables is domain experts [7], [82].

3) TRANSFERABILITY

The main aim of transferability means making users reuse the knowledge in another problem domain. This may cause a problem when the assumption of the user is not correct according to the domain or some of the constraints may not allow the transferability of the model. Explainable AI helps explain the boundaries for better understanding & implementation. ML models can be provided with the training-testing technique for checking their transferability, but not all the models can be explainable [7], [83].

4) INFORMATIVENESS

In this factor, an explainable model can provide the necessary information for solving the problem encountered by the user. The main goal behind this is to provide the information needed for decision-making to the user

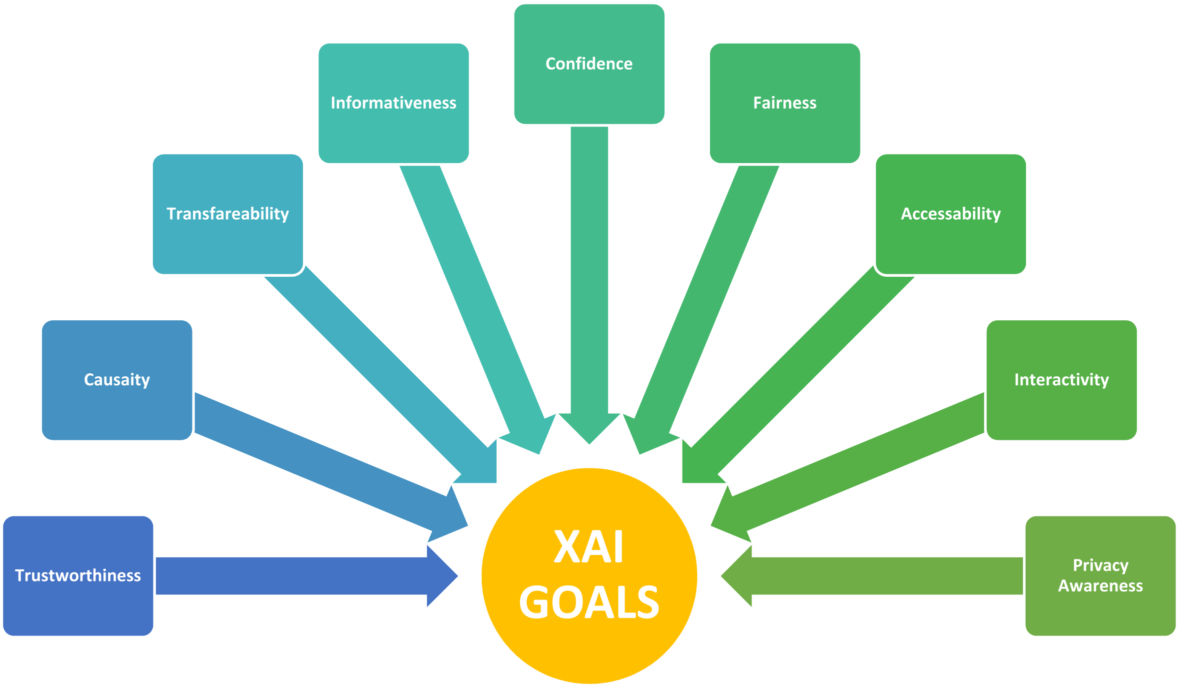

FIGURE 9. XAI Goals.

and to avoid any delusion. The main factor that needs to be measured is explaining the internal functioning or relations of the model by providing all the knowledge needed [7], [84].

5) CONFIDENCE

The explainability of a model will be considered confident enough if it has confidence in its working rule. If the model is enough to provide the factor of reliability to the user, then it can be considered stable and robust to provide confidence to the user while using the system. For checking the confidence of the system, factors might be different for different models [7], [85].

6) FAIRNESS

From the fairness perspective, the system should provide a clear visualization or moral analysis of the system's internal functionalities which might affect the results. As models are developed, they involve human life, so interpretability should provide confidence to eliminate unethical and unfair use of algorithms [7], [86].

7) ACCESSIBILITY

The literature has provided accessibility among the third goal to be considered for explainability. Accessibility in XAI can be pondered and comprised with the help of interfaces. We can think about using visual explanations to provide accessibility in explainable AI. Visual explainable provides access to information to the user with the help of explaining internal functions to the user [7], [87].

8) INTERACTIVITY

The other goal of the explainable AI model is to be interactive with the end-user through proper design of contact between the human and technology. The system should understandably provide information to the end-user. This needs to be ensured where the user interaction with the system is mandatory and the user success will be only valid when the user can perform an interaction with the system [7], [88], [89].

9) PRIVACY AWARENESS

Assessing the privacy of the system is the factor or goal of explainable AI. Some of the ML models provide a complex representation of their learned patterns. The complex representation might be the source of difficulty in the system's understandability, which may cause privacy breaches. The literature suggests including privacy as one factor to be considered during the system development life cycle [7], [90].

C. PROPERTIES OF EXPLAINABILITY

The key properties that are included in the explanation model [91] should have the following properties:

- **Social** in being able to model the explanations of the interpreter.
- **Selective** in selecting explanations related to the different competing hypotheses.
- **Contrastive** in being able to differentiate between the properties of three hypotheses.
- **Local** use for a particular decision.
- **Global** use for a complete model.

TABLE 3. Explainability techniques.

Techniques References	*Text*	*Local*	*Global*	*Visual*	*Provenance based*	*Surrogate Model*	*Explanation by Example*	*Explanation by Simplification*	*Rule-based Method*	*Feature Relevance*	*Declarative induction*
[12]			✓								
[92]	✓										
[16]		✓									
[7]				✓							
[18]					✓						
[7]							✓				
[7]								✓			
[96]		✓									
[100]				✓							
[18]						✓					
[7]										✓	
[110]									✓		
[18]											✓
[95]		✓									
[101]				✓							
[17]						✓	✓			✓	
[98]		✓	✓								
[97]		✓	✓								
[99]		✓	✓								
[102]							✓		✓		
[103]								✓			
[111]									✓		
[108]						✓					
[104]									✓		
[105]										✓	
[106]					✓						
[107]						✓					
[109]											✓

- **Abstraction** is used when the model is complex and a simplified model can provide useful feedback.

D. TYPES OF EXPLAINABILITY

Explanation of a model has different types [91] to be used, including:

- **Plan-based** is used to explain the plans as a set of observations.
- **Model-based** is used when the user of the model requires the explanation of the system to understand why the system performs the way it does. In which the user can be engaged in the criteria of the explanation and the user can use a mental model for designing the system.
- **Algorithm-based** is used when the model explanation is based on a specific algorithm.

E. EXPLAINABILITY TECHNIQUES

Behind many of the characteristics of machine learning discussed in Section A of this background study, the main characteristic that combines the concepts of HCI and AI is explainability or XAI. The summary of explainability techniques in the literature is shown in Table 3. Which shows the literature review of various papers and provides the future of these fields together. Keeping in mind the adherence literature review, we show explainable machine learning techniques which are listed down below:

1) TEXT

The technique of text explanation is used to increase the explainability of a model, which will help generate text for the explanatory results of the model. The text model will also include method-generating symbols that will help represent the functioning of the model [7]. Text explanations work in the natural language processing and computer vision domain and generate the rationales/explanations derived from input data [92].

Text-based explanation techniques have been used in many domains, like neural networks, legal lawsuits, natural language processing, and the medical domain. The researchers in [93] concentrated on the explainable text-driven neural network for stock prediction. The paper discussed that the previous work done in the field has not focused on the explanation perspective but only on stock prediction. The paper proposed a dual-layer attention-based neural network to address the issue. The workflow was based on the initial extraction of financial news, then the most influential news is given input attention. In the last stage, an output attention mechanism is then used to weigh different days in terms of contribution. The presented model in the last produced remarkable explanation results as compared to the traditional way. The researchers in [94] have focused on the attention of legal communities towards AI. The researchers have reviewed the literature and proposed that it is possible to develop Explainable systems which have the potential for completing the demands of legal systems related to document review matters. For the document review process, predictive coding is used as one of the best ways to enhance the quality and speed of the document review. Predictive coding is said to be one of the best methods that provides confidence to lawyers with results similar to supervised learning tasks. Hence, the results provide an open space for future research in explainable predictive coding.

2) LOCAL

The Local explanation is used to express the individual feature attributions of a single instance. It is also said to be the level of explaining individual decisions made by the classifier [16].

The study conducted in [95] has shown the effect of confidence score and local explanation on a particular problem. The confidence score in the experiment is shown with the help of two human experiments that help build people's trust in the AI model. The study shows some of the problems related to local explanations and suggests working on them in the future. Another piece of research conducted in [96] has provided a way of evaluating local explanations. The proposed methodology in the paper uses synthetic ground truth explanation to assess the level of explanations returned by the local explanation methods. The proposed methodology generates a synthetic ground explanation for the decision taken. The results show that the proposed methodology provides easy evaluation for a local explanation.

3) GLOBAL

Global explainability refers to techniques that attempt to explain the model as a whole [12]. It refers to explaining all of the algorithm's functions.

The paper presented in [97] has provided a complete framework for defining the right level of explainability by combining the technical, legal, and economic aspects of explainability. The main idea behind this was to achieve trust and accountability, as designers and operators of machine learning must know the internal workings of the algorithms, its results, and failures of algorithms. For defining the framework, three logical steps are proposed. The first is to define contextual factors, the second operational context, and the third is to provide a legal/regulatory framework. The first step will mainly define the audience involved. The second step will identify the technical tools available, which will also include post hoc approaches and hybrid AI approaches. The third step as a function of the first two steps will involve choosing the right level of global and local explainability. The paper provided seven types of cost and emphasis. It also concludes that explainability is both a functional and an ethical requirement, and that it will become part of the performance over time. Another study published in [98] developed a new performance explainability analytical framework for evaluating and benchmarking machine learning methods. The framework mainly provides a set of characteristics that will systemize the performance-explainability assessment. The main idea was to identify ways for improving current machine learning methods and for designing new ones. The main focus in the paper was given to local and global explainability criteria. To illustrate the framework used, it is applied to the current state-of-the-art multivariate time series classifiers. The future work includes extensively applying the framework to different machine learning methods.

The research conducted in [99] proposed a risk management framework for the implementation of AI in banking. The main consideration was given to the explainability and outlining of implementation requirements. Financial institutions and the customers and markets they assist can achieve a positive outcome through AI. The work presented in the paper evaluates three algorithmic approaches for nine banking use cases. The algorithms involved include Neural Networks, Type 2 Fuzzy logic, and Logistic Regression. The results of the research show that Type 2 fuzzy logic can deliver good performance in terms of precision, accuracy, and recall and has more advantages over other algorithms on both global and local levels. The future work for the paper suggested including more popular algorithms for evaluation purposes.

4) VISUAL

Visualization or Visual technique, is a post-hoc explainability technique that is used for visualizing model behavior. Visualization is coupled with other techniques for improving understanding and is said to be one of the most suitable ways for introducing complex interaction [7].

Visual explanation is one of the new fields in AI which is mainly working in the field of deep learning and is providing a new way to introduce and add explanations related to the given tasks. One of the research conducted in [100] has provided a review in academia and industry, including basic toolkits, advanced computational techniques, and intuitive and interactive visual analytics systems. In this research, new gaps and opportunities such as a human in the loop, visual analytics integrating human knowledge, and a data-driven learning approach related to visual analytics have been provided for future research. Deep learning will help achieve accurate, interpretable, efficient, and secure artificial intelligence. Another survey conducted in [101] has provided a state of artwork for providing a way of enhancing trust in ML models with the use of visualization. The research includes 200 articles that are categorized to make it accessible to the public online. An interactive survey browser, TrustMLVis Browser, is implemented and made available online. The browser supports the reader's exploration of the rich information provided and the enhancement of trustworthiness with the help of the ML model for interactive visualization. The future of the survey includes the extension of the data set, categorization, and corresponding analysis.

5) EXPLANATION BY EXAMPLE

It considers the explanation for the data extraction that can relate to the result generated by a certain model for a better understanding of the model itself or returning data instances as examples for explaining the model behavior [7], [17].

In one of the research studies, the researchers provided a review of various explanation techniques developed in AI and law. In the common-law tradition of the United States, explanation by example is defined as taking examples from the past cases and trying to solve the case by matching them with the cases that need to be solved. In the past, this was said to be a form of contrastive explanation. Examples can be said to be the use of both negative and positive. Explanation by example, in the case of law, takes place in three-step dialogues. First, the plaintiff cites a case similar to the current case. Second, the defendant replies with the help of offering counterexamples, and third and final step is the plaintiff's attempt at contradiction, which means distinguishing the counter examples for which the reason will be offered [102].

6) EXPLANATION BY SIMPLIFICATION

It is a type of explanation in which a whole new system is rebuilt based on the trained model to be explained [7].

The study conducted in [103] has provided an extensive survey on text simplification. Text simplification is one of the fields which reduces the chances of complexity and improves the readability and understandability of the text. Text simplification is mainly focused on providing knowledge to those non-native learners and those struggling with reading problems. The survey has provided text simplification with covering resources, corpora, and evaluation methods being used in this field. The study provided the text simplification approaches along with machine translation techniques and languages in which text simplification has been applied so far. The study also addressed the recent work of monolingual machine translation which covert the original text translation into a simple one. The study concludes that automatic text simplification is far from perfect and needs further improvement. For improving automatic simplification reverse engineering has been proposed for the next breakthrough in automatic text simplification. Another paper presented in [104] proposed two solutions for convolutional neural network simplification named objective pruning with progressive retraining which eliminates kernels of a given convolutional layer based on objective relevance criterion and subjective pruning with progressive retraining works similarly as the other except kernel relevance criterion. The proposed model produced or achieved more network simplification than the original model. The conclusion of the study is the two solutions objective pruning and subjective pruning are relevant contributions to the literature of kernel pruning methods. The future work of the study proposed methods to construct CNNs in a layer-by-layer fashion.

7) FEATURE RELEVANCE EXPLANATION

It is used for post-hoc explainability which provides clarification for the internal functioning of the system. It is also said to be pointing out how each feature affects the decision [7], [17].

The research published in [105] proposed a novel architecture for XAI based on semantic technologies and AI. The scenario was tailored to the scenario of forecasting and it was validated with the real-world case study. Explanation combines the concept of feature relevance for a particular forecast, related media events, and metadata regarding an external dataset of interests. The proposed model also used a surrogate model for the prediction of the sample. A knowledge graph was used in the proposed architecture for conveying feature information at a high abstraction level, domain knowledge forecasted values, forecast explanation, and for keeping the sensitive details safe. The ontology and the data set for the research conducted for the use case are available publically and can be used for future research.

8) PROVENANCE BASED

It is used for explaining by illustrating some or all of the prediction derivative process. It is said to be an intuitive and effective usability technique [18].

The paper presented in [106] has focused on the importance of machine learning in many fields, like healthcare, security, investment, and many other critical applications. Hence, machine learning can be manipulated and used in many applications. Machine learning can be used in manipulating data models. One way is introduced, which was through a poisoning or causative attack in which the adversary feeds carefully crafted poisonous data points into the training set. Taking advantage of recently developed tamper-free provenance frameworks, the researchers in the paper proposed a methodology that uses contextual information about the origin and transformation of data points in the training set to identify the poisonous data. The proposed methodology is said to be the first defense strategy for preventing poisonous attacks. The study presents two variations of the provenance defense for both partially trusted and fully untrusted datasets.

9) SURROGATE MODEL

The Surrogate model is a model in which another model works as a proxy for them to work as a more explainable model. In other words, they are said to be interpretable using other explanation models like feature attribution and example-based. An example of a surrogate model is LIME, which learns surrogates using an operation called input perturbation [18], [17].

Surrogate explanation in the field of machine learning is said to be one of the fields which can be applied to any type of data like images, text, and tabular form. The surrogate model is often unified with LIME and is considered that LIME is the solution to surrogate explainability. In this research, the researchers have proposed a principle algorithmic framework for building a custom local surrogate explainer of the black box model itself, including LIME itself. The paper also discussed the danger associated with algorithmic choices and for avoiding common pitfalls. In future investigations of the behavior, surrogate models and the quality and stability will be investigated [107].

The review conducted in [108] has proposed that a framework for selecting an appropriate surrogate model for a given function or problem is lacking. The framework for such a model will help industry personnel get information about themselves before applying it to any of the problems. The researchers in the paper worked on the three main parameters' including size, accuracy, and computational time to get the gap that existed in the existing model and to create practical guidance for use in the future, this will help save time. The results provided 6 different quantitative categories for the surrogate model. These categories provide a framework for selecting an efficient surrogate modeling process for the assistance of selecting an appropriate surrogate model.

10) DECLARATIVE INDUCTION

They are said to be human-readable representations which can be said as rules, trees, and programs are induced as an explanation [18].

The paper in [109] provides an overview of the key points and important developments in the area of computational sense-making. It is used for developing methods and systems for making sense of complex data and information. The main goal of it is to provide insights to enhance the understandability of subsequent intelligent actions. The declarative induction will help to include guiding knowledge into the process, and explication will provide interpretability, transparency, and explainability in the process, which are also said to be the key elements.

11) RULE-BASED METHOD

The rule-based explainability method is considered to have good interoperability. The rules that are made are good for learning and classification of data. There is a large ecosystem of software tools established which work with the help of rules [110].

Rule-based models have been given importance in many of the domains. The research discussed here provides the rule-based models in the paper in terms of the law. It is a rule-based technique that mainly provides the rules that needs to be matched with the result. Law cases have a strong feature of common law traditions and laws are found in statutes that define the particular concept in different case areas. The research proposed that some of the traditional expert systems have used production rules and rules based on knowledge from the domain experts which provide all the standard explanations of how, why, and what-if. One of the things that were suggested in the study is that more insights from AI and law should be provided with an explanation as to the main concern for future explainable AI [102].

Another research conducted in [111] discussed the significance of rule interpretability and how it should be taken seriously. Five experiments were conducted for insight into the plausibility of rule learning results. The main focus in the paper was that a longer explanation is more significant than a shorter one, which makes the learning of models easier. Users in the study have been confronted with pairs of learned rules. The gap that was observed in the study was the issue of background knowledge by picking a domain in which the participants may have clear domain knowledge. The results of the research reveal that simple rules have a strong preference and longer rules in some domains have been given a weak preference.

VII. CHALLENGES IN XAI

When we talk about the advancements related to the field of XAI. As this is a new field that came into existence, it is somehow facing some of the challenges in it. Table 4. Below, we provide a summary of challenges and problems related to the field of XAI.

While there are many progressions in the field of XAI, the literature identifies gaps that have been identified in this field that need to be worked on for improvement. One of the papers in [19] has identified the challenges that have been recognized while reviewing the literature that includes evaluating explanations and performance of XAI. Evaluating explanation means providing the best explainability techniques for explaining the transparency of the model. The performance of XAI will affect the performance of the system. This challenge is said to be used as one of the primary goals to overcome for the future.

Another paper in [7] also identified the performance of XAI as a challenge which means, if we include more explainability in the model, it will produce unmatched results when solving complex computational tasks. The paper also pointed

TABLE 4. Summary of XAI challenges.

References	Challenges	Description
[19]	Evaluating Explanation	The techniques to be used for the best explainability [19].
	Performance of XAI	More explainability in the model can cause an effect on the performance of the system [19].
[7]	Performance	Unmatched results produced by the model when solving complex computational tasks [7].
	Vocabulary	Lack of agreement on vocabulary and definitions concerning XAI [7].
	Evaluating Explainability	Concepts and Researches for the explainability of the model [7].
	Deep Learning	Challenges in obtaining explainability in deep learning models [7].
[16]	Black-Box	Black-Box nature of deep neural networks raising ethical and legal concerns including the absence of trust [16].
[112]	Human Machine (Brain) Interface	Developing intuitive interfaces that can fit into existing workflows and processes [112].
	XAI Twin	Developing an explainable twin system to work in parallel with a deep learning system for optimization performance [112].
	Defense Against Attacks	Developing a defense mechanism that can recognize targeted attacks against DL and XAI engines [112].
[113]	XAI System Evaluation	What explanation means from a sociological viewpoint? Evaluating the quality of explanation [113].
	XAI Interpretation	Interpretation of the model sometimes becomes difficult for the users [113].
	Legal Challenges	Working with sensitive information might create legal challenges in XAI [113].
	Practical Challenges	Fewer practical applications for ensuring AI factor in XAI [113].
[114]	Manual Evaluation	The manual system involves human assistance for the judgment of XAI method results which is time-consuming [114].
	Automated Evaluation with High Computation Time	Automated evaluation can cause distribution shifts in the test data and can violate assumptions in training data [114].
	XAI Methods are not Stable and could be Attacked	Literature shows that XAI can show similar results with normally trained models and modern trained models [114].
[20]	Generalization of XAI	XAI is vastly environment and domain-dependent thus the generalization of XAI might take time [20].
	Adaption of XAI	The applicability of XAI in different domains still needs effort and research to be adapted [20].
	Security of XAI	By the use of adversarial machine learning thus the effect on ML and DL is a vast topic to be used for its security in research [20].
	XAI and Responsible AI	The real-world application should be investigated in terms of responsible AI where the effect of XAI on usability should also be investigated [20].
	Performance of XAI	The performance of XAI should be checked by adding additional features in the baseline of AI models [20].
[115]	Interpretability	A major thread of AI in explanation explore techniques and limitation of interpretability [115].
	Tailor explanation	Explanation of AI tailored according to the user and should be made interactive by adding humans in the loop [115].
	Abstraction	Systematizing the discovery of abstraction has been a challenge [115].
	Explaining Decisions	The new situation should have the ability to help end-user to understand and explaining the abilities of AI systems [115].

towards another challenge that was vocabulary. According to the results, there was a lack of vocabulary and definitions related to XAI, which needed to be increased to improve performance in this field. This paper also provided concern, as in [19], for the evaluation of the explainability of choosing the best explainability techniques to be used. The last concern or challenge that was provided in the paper was related to deep learning explainability. The paper is trying to convey the work being done on providing explainability models for deep learning.

The paper in [16] was concerned about the issue of the deep neural network black box, which is creating trust issues for people. The main concern with this is to provide transparency to the people, which may improve the trust issues for them.

Another paper in [112] came up with the challenge of human brain interfaces that needed to be created spontaneously in the existing workflows. Another challenge was the XAI twin, which means an issue with the deep learning twin system, if created, will work in parallel with deep learning, which will improve the explainability issue. Another issue was security issues that can arise from deep learning and explainable artificial intelligence. The study proposed working on defense security mechanisms to fight against such attacks.

The paper in [113] also provided the challenge concerning XAI system evaluation, which was also discussed in [19] and [7], which mainly means that choosing the best technique for explainability is still a challenge that needs to be improved by providing evaluation criteria for explainability. The research also pointed toward the challenge of interpretation, which means that users sometimes are not able to clarify the things concerning the explainability model. The legal and practical issue are also one of the factors that needs to be worked on, which means that, from a legal perspective, using sensitive information or data might create some issues, and for practical purposes, there are only a few applications that can provide reassurance of the AI factor in XAI.

The research in [114] has pointed out some of the challenges in which one of the challenges is related to the manual evaluation of the explainability model, which will be time-consuming and may produce bias and inaccurate evaluations. The research also provided the challenge concerning XAI automated data and identified that it may also cause distribution shift which results in the assumption of training data. The other concern was the security issues which mean that XAI methods are not suitable and can be attacked.

In [20], researchers have provided a list of challenges which include XAI being massively independent and it being strict about the environment and domain. The paper also pointed toward XAI security, as in the previous papers. The paper also pointed toward the challenge of XAI adaptability, which is still vague and will require some time for people to trust it and work with it. Another concern was the usability and performance of the XAI application, which still needed to be tested on real-world applications for AI.

The research in [115] has proposed to keep humans in the loop while taking their concerns and making the XAI applications tailored according to them. The research also provided the challenge of interpretability in terms of AI. Other challenges include abstraction and explainability decisions. Abstraction of explainability remains a challenge, as does explaining your decision in relation to the system that you are using and what the benefits of the system are.

VIII. HEALTHCARE

Many advances have been made in the field of healthcare in the past few years by improving healthcare and shifting to digital health technology, which includes many areas like artificial intelligence, big data, wearable's, and medical technology and devices. In this section, a literature survey of articles from the field of HCI, AI, and XAI are discussed to explore the focus of these fields in healthcare. This section will provide the survey results of healthcare that have been discussed in the past and the main focuses. Table 5. Brief about the summary of HCI in the field of healthcare along with the topic addressed, solution, achievement or results, and weakness or limitations.

A. HUMAN COMPUTER INTERACTION IN HEALTHCARE

The growing demand for healthcare technology is increasing day by day. There are many challenges and gaps that need to be closed between industry and academia to improve acceptance of technology, ensure compliance, good ergonomics, and high-performance design for all users and contexts of use. Many applications have been named for improving HCI in different research in the past regarding healthcare. Some of them include a natural user interface, child computer interaction, and interpretation for people with disabilities, and human factors for healthcare [116]. This section is regarding some of the other research that has been done in the past to talk about HCI and healthcare.

The paper is presented in [117] it gives human-robot interaction details in general. Human-robot interaction is the leading field when we talk about AI & HCI. In the past, robots have been used for industrial purposes. But nowadays, robots are also being used for social purposes. The important purpose that social robots fulfill is to fit into the human environment and socializing with humans. Healthcare sector robots are being deployed to overcome the shortage of healthcare professionals, rising costs in healthcare, and growth in vulnerable populations like the sick, aged, and children with disabilities. The challenges faced in robots are safety, usefulness, acceptability, and appropriateness. At the interaction end, the usability issues include privacy, trust, safety, users' attitude, culture, robot morphology, as well as emotions and deception.

Another article in [118] has also discussed social robots as a new viewpoint in the healthcare sector. The paper discussed various robots developed in the past. Some of them include Nao Robot (2006), which mimics human behavior like a toy, Paro (2009), developed for supporting therapy and

TABLE 5. Summary of HCI in healthcare.

Reference	Topic Addressed	Solution	Achievement/ Results	Weakness/Limitation
[117]	Shortage of health professionals	Robot Technology	• Social use	• Safety, privacy • Usefulness • Acceptability • Appropriateness • usability issues • trust
[118]	Care of elders in hospital	Social Robots	• Use of social robots in healthcare	• Sensing emotions in care and use of implanted sensors.
[119]	Review for checking the most used HCI categories	Review of available HCI healthcare categories	Most used HCI categories in Healthcare • Usability • Security, privacy & trust • Automation • Training & simulation • Information/patient records • Human Factors/ machine Interaction • At-Home healthcare	• Healthcare automation Safety • Home health and security
[120]	Review of HCI theories in healthcare	Activity Theory, Actor-Network Theory, Distributed Cognition, Structuration Theory, and Situated Action	• The active theory is considered to be the most useful theory	• Some contradictions need to be worked on for smoother system implementation.
[121]	Review for checking sufficiently rules of HCI and user-centered design	Review on Existing internet-based applications for mental illness and depression	• Computer interaction importance in e-health	• Poor understanding of safety. • Effectiveness • Dependable • Credible • Reliable • trustworthy implementation of the interventions
[123]	Cross functioning of HCI and Visualization tool	A literature review was conducted	• For big data problems, users can use visualization tools and HCI principles	health services have • huge amount of complex • unstructured • high dimensional data
[122]	Need for medical care for adults	Successful active aging and e-health.	• Factors for improvement of HCI in Healthcare	• Trust • personal integrity • technological acceptance • e-health literacy • Accessibility • Information Communication Technology

TABLE 5. ***(Continued.)*** **Summary of HCI in healthcare.**

[124]	Healthcare improvement for clinicians and managers	Narrative review	• Address the identified gaps. • The narrative review summarizes and provides a pragmatic LHS framework and context 'road map' for clinicians is warranted.	• HCI, embedded within the clinical setting, may lack academic rigor and potential efficacy. • Clinician engagement and leadership in improvement needs aptitude with both technical (clinical) and change management skills
[125]	Challenges and Opportunities in healthcare	Digital Health technologies review	• New strategies are needed for designing and deploying interactive healthcare systems.	• Improved usability. • Safety and efficiency. • A new model of digital-enabled technologies.

care of elderly patients in hospitals, Robear (2015) provide support for lifting a patient out of bed and the most popular, Sophia (2017), which learns and adapts to human behavior. The main theme behind this paper was to make it look like social robots have gained popularity in the healthcare sector and can somehow fulfill the requirements of the sector.

The article presented in [119] has presents a review of papers from 2010-2017 in which the most popular topics of HCI in healthcare are presented, including information and patient records. Some of the main important topics that were covered in the reviewed articles related to HCI include usability, security/privacy/trust, automation, training & simulation, information/ patient records, human factor/machine interaction, and safety. These are some of the factors that need to be covered when implementing applications relating to the medical healthcare system.

The literature review [120] presented here discussed five of the HCI theories that are considered to be the most suitable for use in the healthcare context. Theories are selected based on their popularity and include activity theory, actor-network theory, distributed cognition, structuration theory, and situated action. The results of the study show that activity theory is less popular than structuration theory, but with regard to HCI, it is by far the most applied theory in research. Actor-network theory and situated action are the least popular theories. Distributed cognition is not as regularly applied as activity theory, but still has a reason for its existence, as several case studies have shown. The results declared that activity theory is considered to be a valuable theory for facilitating a better understanding of technologies in a healthcare context. Some of the contradiction layers that have been proposed in the paper will help to reveal key issues that, when resolved, can lead to smoother system implementation and more streamlined processes.

The literature review conducted in [121] shows that many health professionals are creating internet-based treatments for mental health conditions like depression and anxiety. The main reason for conducting this literature review is to find out how sufficiently the rules of HCI and user-centered design is being incorporated. Some of the negligence that is still made in developing those applications include poor understanding of safety, effectiveness, dependable, credible, reliable, and trustworthy implementation of the interventions. The endorsement arising from this review is that HCI should be carefully considered when mental health nurses and other practitioners adopt e-mental health interventions for therapeutic purposes to assure the quality and safety of e-mental health interventions on offer to patients.

The systematic literature review in [122] pointed out the main concern about the need for medical care, which has been increasing for adults around the globe. The main theme behind this is successful active aging and e-health. The key research question behind the study conducted was to identify the factors for improved HCI in technology-enhanced healthcare systems for older adults. The findings of the studies show several factors needed to be considered for improvement, which include trust, personal integrity, technological acceptance, e-health literacy, and accessibility of Information and Communication Technology (ICT) as the most determinant. The following challenges need to be addressed and improved to make independent living easier for older people.

The literature review in [123] pointed out the cross functioning of HCI and visualization tools in detail. The main theme of the review was to check how we use the HCI and visualization tool to get the users desired information from a large amount of data. Healthcare services have crucial and high-dimensional data that needed to be analyzed and integrated. The effective analysis and management of large amount of data in healthcare has become a priority. To overcome the integration problem, a new approach has been utilized in the study for effective analysis. Based on the sample research, the study focused on the capabilities and opportunities of working together with these two fields on healthcare services to get accurate and effective results. The researchers investigated and used the most popular D3.js, Welkin, and Gephi visualization tools in this study for their research.

The narrative review conducted in [124] provides a review of the previous studies conducted in the field of HCI in healthcare and how HCI can help improve clinicians' engagement and leadership. The literature also addresses the issue of not providing necessary information focused on frontline practice. The issues are addressed in the literature for the improvement in the complex health system to assist clinicians. The review integrated the field of HCI into the health system by highlighting the key role of clinicians. The review also proposed a clinicians and manager's guide for the planning, enacting, sustaining, and scaling of HCI.

The paper in [125] has discussed a variety of digital health technologies for professional patients and analysts that support health management and discovery. The review also highlighted the benefits of digital health technologies. The survey also provided ways of improving healthcare by providing an integrated development cycle in which step-by-step work will help create a good healthcare technology.

B. ARTIFICIAL INTELLIGENCE IN HEALTHCARE

AI is the field that is providing opportunities for advancement in many fields. In healthcare, AI is providing insights to improve patient and clinical team outcomes, reduce costs, and influence population health [126]. The other role AI is performing in healthcare includes early detection and diagnosis [127]. AI is said to be applied in various types of healthcare data, which may include structured and unstructured data. Popular AI techniques of include machine learning, vector machines for structured data, and natural language processing for unstructured data. The most common disease areas in which AI is working are the fields of neurology, cancer, and cardiology [128]. It is said in many articles that AI will add capabilities that will lead to more efficient and effective care by healthcare providers [129]. AI in healthcare is said to be The hope, The hype, and The promise [127]. In this section, some of the paper from the literature have been reviewed from the AI perspective in healthcare, the challenges, and what could be the future of AI in healthcare. Table 6. summarizes the literature that has been done focusing on the importance of AI in healthcare in the past.

A paper published in [128] reviewed the importance of AI in healthcare and provided a literature survey of the applications of AI in healthcare in three major areas of early detection, diagnosis, and treatment as well as outcome prediction and prognosis evaluation. The motivation of AI systems is to reduce diagnostic and therapeutic errors that are unable to be overcome in normal human clinical practice. The paper provided the analysis of ten diseases in which the AI working fields include neoplasms, nervous, cardiovascular, urogenital, pregnancy, digestive, respiratory, skin, endocrine, and nutritional diseases. The most common areas among these diseases include cancer, neurology, and cardiology due to the severity of these diseases. This is the reason that early diagnosis is an important factor in these fields. The review provided the importance of AI based on the fact that AI will unlock massive hidden information from the field of healthcare that will assist physicians in making better clinical decisions in the near future. The paper provided machine learning algorithms in the medical literature which are used for searching the data within healthcare. The most common algorithms that have been used in ML include Support Vector Machine (SVM) and Neural Network (NN). Input to the ML algorithm includes patient traits and some medical outcomes. Some of the baseline traits of patients include age, gender, and disease history, and disease-based data such as diagnostic imaging, gene expression, EP tests, physical examination results, clinical symptoms, and medication. Patient medical outcomes that may be needed for ML inputs include indicators, patient survival time, and quantitative disease levels.

Another paper published in [129] is a review of ML in healthcare, traditional, clinical, and public health applications with an important role in privacy, data sharing, and genetic information. This paper focuses on the implementation of AI in the healthcare fields in disease diagnosis & prognosis, treatment optimization and outcome prediction, drug development, and public health. In ML, AI is used for making automated clinical decision systems. The main concern with technological advances is that they require collecting and sharing a massive amount of data, which will generate privacy concerns. The paper points selected areas in which ML has high potential in clinical translation and public health, which include clinical disease prediction and diagnosis, drug discovery and repurposing, and public health (epidemic outbreak prediction). The study pointed out that the research done in the past has focused on cancer, the nervous system, and cardiovascular disease. Studies show that many of the drug discoveries done in the past are the reason for combining different domains accidentally. ML is used in drug discovery for making cross-domain linkage due to the high cost of drug

TABLE 6. Summary of AI in healthcare.

Ref.	Year	Topic Addressed	Technique/ Algorithm	Diseases	Weakness/Limitation/ Future work
[25]	2021	Review of common healthcare applications and Projects	Rule-based Decision support system	• General	• Unstable Rule-based model • Poor Abstraction
[128]	2017	Review of Common AI diseases	Support Vector/ Neural Network	• Cancer • Cardiology • Neurology	• Safety • Efficacy • Performance • Adaptive Design • Data Exchange
[127]	2019	Review of AI in applications in India	SWOT Analysis	• Radiology	• AI development • Strategic positioning • ethical considerations • joint public-private sector collaborations
[129] [130]	2019 2019	Review of ML in healthcare traditional, clinical & public health applications	Not Identified	• Cancer • nervous system • cardiovascular disease	• Privacy • Interoperability • Trust
[131]	2019	Implementing privacy-preserving classification	Random Forest Decision tree	• General healthcare system	• More research is to be done for making the field more common.

development. In public health epidemic outbreaks predictions such as peak and duration of infection can be easily made possible if model parameters are partially known, as done in [130]. Some of the challenges regarding healthcare that have been identified in this study include privacy, interoperability, and the issue of trust.

The paper published in [127] reviews the status of AI in developing nations like India and highlights some of the factors that can be beneficial for providing new directions and opportunities for AI in healthcare. The field that is pointed out in this review is the field of Radiology. Radiology is said to be one of the most evolving fields of medicine. AI-infused systems have already been developed in the past for the field of radiology, including the Missouri Automated Radiology System (MARS), a computer-based expert system being developed for radiologists (ICON), Pheonix, and MARS II. The ethical framework for AI in radiology includes autonomy, beneficence, justice, explicability, and transparency. The study proposed that AI and radiology can be combinable in the form of Augmented Intelligence, which will make the future of healthcare lively.

Reviews for AI in healthcare are also presented in one of the papers in [25], which surveys the present state of healthcare applications and the projects related to them. The medical literature shows that sophisticated algorithms can be developed using AI to read features from vast datasets of healthcare data and can use the knowledge learned to help clinical practice. The review shows that AI systems can help reduce medical & therapeutic mistakes. The Clinical Decision Support System (CDSS) in medicine was famous in the mid-twentieth century. Rule-based models in the field of healthcare are found to be unstable and can require clear expressions of decision rules. Hence, the paper proposes that with the involvement of AI in healthcare, patients' satisfaction needs to be maximized.

The paper in [131] presents the first secure multiparty computation (SMC) enabled cryptographic protocols for private classification with tree ensembles, random forest, and boosted decision trees. The SMC system was also integrated with the KenSci healthcare analytics problem and is also supposed to be one of the first privacy-preserving machine learning protocols implemented in a real-world scenario. The review also shows that there are still gaps in security protocols to be implemented in healthcare and can be worked on for future research.

C. EXPLAINABLE ARTIFICIAL INTELLIGENCE IN HEALTHCARE

Artificial intelligence is being implemented and used in different fields of healthcare and medicine. The function of AI in healthcare includes diagnosis, treatment (identification), health management/patient engagement, and health system simulation, etc. Thus, the field of artificial intelligence will be reached to the level of superiority when the field of XAI will be given the functionality of being able to be communicative with human users in a realistic manner for answering questions that make the users lack trust in AI [21]. Table 7 summarizes previous studies, including diseases, techniques used, and limitations. There are many studies conducted in the past related to AI's importance in healthcare. One of the studies conducted in [21] has provided some of the dysfunctional items in healthcare related to AI, including organizational issues, communicational issues, and socio-relational issues. The organizational issue is mainly concerned with AI applicability in healthcare which may struggle to adapt to the timing, procedures, and organization boundaries of healthcare. Another concern was a communicational issue which may lead to confusion of tongue and the information provided by the doctors may be lost by AI or transformed, which will confuse the patients. The third issue is relational ambiguity, which means involving AI may lead to confusion about patient and doctor roles, which will cause socio-relational issues. The study also provided the solution that needed to be considered for improving the aforementioned issues. For the improvement of organizational issues, systematic plans related to AI implementation management should be included. For the communicational issue, doctor awareness is necessary along with the double-check of health information with the patient, and for the socio-relational issue, the education of the patient and ad hoc formative resource usage should be considered. One of the aspects that have been said to be looked into is XAI, which aims to make the communication between humans and AI easy and trustworthy.

Based on the increasing trend of AI in healthcare. In the field of providing trust and explainability to the users, XAI gains importance. For understandability and interoperability of AI systems, the researchers in [132] have presented different interoperability techniques. According to the paper presented, the AI system provides a variety of available techniques which can be adventurous in the healthcare domain. The paper presents datasets related to healthcare diseases and explains the advantages of using explainability techniques on them. The techniques that were discussed in the paper include example-based techniques and feature-based techniques. The important benefit learned from the study of these techniques is that the approaches all speak about how various features are responsible for the model's outcomes and assist in the process of learning along with the ability to explain the behavior of the model.

The research conducted in [24] has given importance to using the XAI system for explaining the predictions made by AI systems. The paper has discussed XAI as a system for the analysis and diagnosis of health data through the use of AI systems and has proposed an approach for achieving transparency, accountability, result tracking, and model improvement in the field of healthcare. The researchers discussed some of the studies conducted in the past and their methods of working in healthcare in which LIME and IF-THEN rules were the prominent methods that were discussed. According to the proposed model in the paper, using some of the existing XAI models in combination with clinical knowledge can be used to obtain more benefits in AI-based systems. According to the proposed model, smart healthcare applications will be used to capture the health information of individuals, and a trained AI model will be used to predict the abnormalities or disease. The predictions will be used by XAI models for the explanation. The explanations will be matched with the clinical knowledge for validation purposes if find correct then, valuable insights and recommendations will be generated through it. If find incorrectly then, inaccurate predictions between clinical knowledge and explanation will be taken into account for improvement. Thus, the model has given importance to XAI models for use in healthcare systems.

Research conducted in [25] has firstly discussed the survey of the current progress made in XAI and its advances in healthcare applications. The survey provided a mini-review of XAI methods being used in digital healthcare and medicine which promotes the concept of XAI globally. The review was conducted on the advanced use of explainability in the healthcare sector. The mini-review conducted on the research is focused on the research related to XAI in healthcare and medicine, which is categorized into five categories, including XAI via dimension reduction, XAI via feature importance, XAI via attention mechanism, XAI via knowledge distillation, and XAI via surrogate representations. The researchers then proposed solutions for XAI for multi-modal and multi-center data fusion and the validation of both of them was done using real clinical scenarios. Research conducted has demonstrated two typical but important applications of using XAI, which have been developed for classification and segmentation. Two of the most widely discussed problems in medical image analysis and AI-powered digital healthcare. The developed XAI techniques have been manifested using CT image classification for COVID-19 patients and segmentation for hydrocephalus patients using CT and MRI datasets. The results of the study have shown promising XAI results. Other research conducted in [133] has focused on the recent investigation into the interpretability and explainability of artificial intelligence and the discussion of its impact on medicine and healthcare. The applications of AI in healthcare are also discussed in the paper, which includes diagnosis and prognosis, drug development, population health, healthcare organization, and patient-facing applications. The paper also discusses some of the limitations of AI in which one of the drawbacks is difficulty in validating the output of AI. For that purpose, the field of XAI came into existence, which makes AI results more understandable to humans. Several explanation methods are also discussed in the paper, which include complexity-related methods, scoop-related methods, and model-related methods. The paper has discussed some of the key characteristics that explainable healthcare should include, including adaptability, context-awareness, consistency, generalizability, and fidelity. Some of the future research opportunities that have been discussed in the paper, which include human computing interaction, human in the loop, explanation evaluation, and other explainable intelligent systems, should also be developed. As a result, the paper discussed XAI as an emerging field that was said to be the field that could increase AI acceptance in the healthcare sector. The research conducted in [134] was focused on deep learning explanations related to skin cancer. Deep learning decision support in medical applications is concentrated on and discussed in relation to skin cancer diagnosis using cancer, dermoscopic, and histopathological images. The research was mainly focused on checking the explainability perspective in medical healthcare. The results show that histopathological skin images have received little attention. The review shows some of the existing variety of taxonomies for XAI methods in the literature, which are grouped into four categories: Visual Relevance Localization, Dermoscopic Feature Prediction & Localization, Similarity Retrieval, and Intervention. The review suggests that future work should focus on meeting the stakeholders' cognitive concepts, providing exhaustive explanations that combine global and local approaches, and leveraging diverse modalities. The paper [22] proposed a powerful solution for the increasing explainability of AI-based solutions for the individuals such as medical practitioners. The proposed solution will provide an explainability solution for ML methods and underlying workflows to be integrated into standard ML workflow. The paper in [135] has discussed a trustworthy framework on how to create a trustworthy explainable AI in healthcare. The framework was created from the previous studies conducted in the past. The framework was developed with the help of using two components explanation characteristic and human-machine trust, which is further divided into two types' cognitive-based trust and affect-based trust. The framework provided in this study can be used for building trustworthy explainable healthcare systems in the future

IX. CHALLENGES OF XAI IN HEALTHCARE

The advancement of XAI is integrating different AI technologies. It includes business processes, security, and AI for designers. Healthcare is one of the fields which is now improving itself by integrating technological advances in it. In the past, different ML algorithms were used in the healthcare domain for the diagnosis of diseases and new drug production for patients. XAI is one of the factors which is now being used along with healthcare for making the healthcare sector more advanced. Table 8. Below, we provide a summary of the challenges of XAI in healthcare.

Research in [23] has been conducted for the review of existing XAI systems which are using electronic health records and for providing the techniques that have been used in these studies along with the research gaps and challenges to cater to future perspectives. The study shows that further research needs to be done in XAI for the medical field, along with the review that such research will be helpful for medical professionals, and many opportunities exist for working in this domain.

Another research conducted in [24] expressed concern about the difficulties of restricting the AI model to specific domains in order to achieve greater accuracy. Another issue raised in the paper was the need to explain XAI techniques so that they are easily understood by model users and the medical healthcare system. Lastly, the issue of the development of appropriate user interfaces was suggested for effectively displaying information related to XAI.

The researchers in [23] have provided different challenges related to XAI concerning healthcare. The first issue was that visualization does not always provide good explainability to health professionals. The robustness of the system should be increased by adding more features related to XAI. More features added to the XAI will help to increase the accuracy of the XAI model. The other challenge was that the predictive analysis provided by the model will raise false causation which may not be true related to the said disease and insufficient explainability is one of the causes of not adapting XAI in healthcare.

Other research in [25] has focused on the challenge of explainability and transparency of the XAI model, which may cause inadaptability of XAI models in the healthcare sector. Another study in [22] raised the concern of a lack of explainability and operations of XAI in clinical expertise, which could be a problem for adaptability. The other main issue is integrating XAI with existing ML workflows and the lack of high-level explainability related to ML models.

X. OUR CONTRIBUTION

The paper provides an overview of HCI, AI, and XAI based on the literature, which would be beneficial to anybody interested in these topics. Because all of these disciplines have fundamental information in the literature. The characteristics of ML were also examined in many domains. In the realm of AI, ML characteristics are extremely important to incorporate into different models. The field of Explainable Artificial Intelligence, as well as its techniques, was another focus of the review. The relevance of various XAI explanation approaches in various domains has been established through a review in which visual explainability is used to describe visual explanation of model behavior. The text explanation approach is another strategy that is used to provide text for describing the model. The healthcare area was the primary focus of this XAI evaluation. The review also discussed the relevance of XAI in healthcare and previous research, as well as the problems associated with each review, which are included in Table 8. Interpretability, inadequate explainability, and model correctness are among the most significant problems addressed in various studies.

TABLE 7. Summary of XAI in healthcare.

Reference	Year	Area of Research	Technique /Algorithm/ Tools	Disease/Domain of healthcare	Weakness/Limitation/Future work
[21]	2020	Problems and solutions to dysfunctional items in AI	• Decision Making Process	• medical practice and consultation	• Trustworthy communication between humans and AI
[132]	2020	Interoperability of AI system	• Feature-based technique • Example based Technique	• Heart Disease	• Explain the process of learning in the black-box model. • Explain the behavior of the model.
[24]	2020	Smart Healthcare application	• model-agnostic methods • Local Interpretable • Model-Agnostic Explanation • Recurrent Neural Network	• Asthma • Diabetes • lung cancer	• Better improvements related to its adoption and usage of AI-based Systems. • Assumption-based operation. • Computational cost.
[25]	2021	Survey of XAI and its advances in healthcare application	• dimension reduction • feature importance • attention mechanism • knowledge distillation • surrogate representations	• Clinical Scenarios • Hydrocephalus • Covid-19	• development of XAI in medicine and digital healthcare
[133]	2020	Interpretability and explainability of AI	• Complexity related methods • Scoop related method • Model-related methods.	• Diagnosis and Prognosis • Drug development • Population Health • Healthcare organization • patient-facing application	• Validating output of AI • Human-Computer Interaction • Human in the loop
[134]	2021	Deep learning explanation	• Visual Relevance Localization • Dermoscopic Feature Prediction & Localization • Similarity Retrieval • Intervention	• Skin cancer	• meeting the stakeholder's cognitive concepts
[22]	2020	Feature Importance scores can be integrated into an ML workflow is adopted	• Feature importance	• Automated medical diagnostic	• developing a framework that automates the training and validation of models appropriate
[135]	2020	Framework for interpreting explicability and trust in healthcare	• General framework	• Healthcare	• qualitative and quantitative study for the investigation of explanation and trust in healthcare

TABLE 8. XAI challenges in healthcare.

Reference	Challenges	Description
[21]	Organizational issues	It will cause decision paralysis which will cause uncertainties, delays, and confusion in AI implementation in practice [21].
	Communicational issues	The communication issue can be said a confusion of tongues in which information is lost when the conversion of information for the patient is communicated to AI [21].
	Socio-Relational issues	Role ambiguity acts on socio-relational aspects in healthcare contexts. It refers to the unwanted effects on trust and quality of relationship related to the addition of an artificial converser within the context [21].
[25]	Transparency	Lack of accessibility of models is also one of the reasons for being uncommon in clinical practices [25].
[24]	Restriction of AI models	Restriction for choice of using AI models for achieving greater accuracy [24].
	Incorporation of XAI techniques	The explanation should be made in such a way that it will be understandable by normal individuals or experts in the medical domain [24].
	XAI explanation	Development of appropriate user interfaces should be included for effectively displaying explanation [24].
[23]	Interpretability	All visualization does not provide better interpretability to health professionals [23].
	Robustness	More longitudinal features needed to be added in XAI [23].
	Explainability	The absence of the definition of explainability for different use cases [23].
	Accuracy	More features needed to be added for the improvement of model accuracy [23].
	False causation	Predictive analysis on uncommon diseases might result in some causation that is not known [23].
	Insufficient Explainability	Insufficient explainability is one of the reasons for the maladaptation of XAI in healthcare [23]
[22]	Lack of explainability	Related to feature engineering processes there is a lack of explainability to incorporate clinical expertise [22].
	Complexity	In the integration of XAI approaches with existing ML workflow [22].
	Lack of model operations	In different medical settings, there is a lack of explainability of model operations [22].
	Lack of high-level explainability	High-level explainability is lacking related to data for ML models [22].

A. LIMITATIONS OF THIS REVIEW

For the literature review, only a small number of databases, journals, and conferences were evaluated. Before 2016, no articles were listed. The number of strings and keywords that could be used to search the literature was restricted. Finally, the study focused on the fundamentals of HCI, AI, and its developing fields of XAI in healthcare, as well as the problems that come with them.

XI. CONCLUSION

This study presents a comprehensive review of an emerging XAI area that combines Human-Computer Interaction with Artificial Intelligence. The papers chosen for the evaluation were divided into areas such as HCI, AI, XAI & Impacts, and application of these domains in healthcare. The article highlighted the relevance of the domains of HCI and AI, as well as the newly found area of Explainable Artificial Intelligence, which was created by merging the two areas. Along with the contributions and good XAI in many areas, this article also discusses the goals of XAI and the work of XAI in healthcare, as well as the problems of XAI in the healthcare sector. Because machine learning (ML) is a popular topic in AI due to its algorithmic contributions and features, this study focuses on ML's key properties and explainability approaches. The problems of healthcare can be used in this study to better focus on the study's shortcomings and how to close them.

XII. FUTURE DIRECTIONS

XAI is a new field that improves user trust by bringing people into the loop and eliminating transparency concerns. Future research should be conducted on the XAI algorithms in order to assess their effectiveness and make them more beneficial for application in many areas. The difficulties indicated by the review, which include enhancing and adding new explainability approaches linked to certain domains, as well as improving model interpretability and accuracy concerns, are another area that should be given priority in the future.

REFERENCES

[1] J. Grudin, "Anticipating the future of HCI by understanding its past and present," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 5–8, doi: 10.1145/3290607.3298806.

[2] Z. Zeng, P. J. Chen, and A. A. Lew, "From high-touch to high-tech: COVID-19 drives robotics adoption," *Tour. Geogr*, vol. 22, no. 3, pp. 724–734, 2020, doi: 10.1080/14616688.2020.1762118.

[3] K. Sohn and O. Kwon, "Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products," *Telemat. Inform.*, vol. 47, no. Dec. 2019, pp. 1–14, 2020, doi: 10.1016/j.tele.2019.101324.

[4] Y. Yun, D. Ma, and M. Yang, "Human–computer interaction-based decision support system with applications in data mining," *Future Gener. Comput. Syst.*, vol. 114, pp. 285–289, Jan. 2021, doi: 10.1016/j.future.2020.07.048.

[5] E. Bryndin, "Development of artificial intelligence by ensembles of virtual agents with mobile interaction," *Autom., Control Intell. Syst.*, vol. 8, no. 1, p. 1, 2020, doi: 10.11648/j.acis.20200801.11.

[6] B. Shneiderman, "Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems," *ACM Trans. Interact. Intell. Syst.*, vol. 10, no. 4, pp. 1–31, Dec. 2020, doi: 10.1145/3419764.

[7] B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[8] P. Forbrig, "Continuous software engineering with special emphasis on continuous business-process modeling and human-centered design," in *Proc. 8th Int. Conf. Subject-Oriented Bus. Process Manage.*, Apr. 2016, pp. 1–4, doi: 10.1145/2882879.2882895.

[9] D. Shin, "User perceptions of algorithmic decisions in the personalized AI system:Perceptual evaluation of fairness, accountability, transparency, and explainability," *J. Broadcast. Electron. Media*, vol. 64, no. 4, pp. 541–565, Oct. 2020, doi: 10.1080/08838151.2020.1843357.

[10] U. M. Gidado, H. Chiroma, N. Aljojo, S. Abubakar, S. I. Popoola, and M. A. Al-Garadi, "A survey on deep learning for steering angle prediction in autonomous vehicles," *IEEE Access*, vol. 8, pp. 163797–163817, 2020, doi: 10.1109/access.2020.3017883.

[11] L. Oneto and S. Chiappa, "Fairness in machine learning," *Stud. Comput. Intell.*, vol. 896, pp. 155–196, Oct. 2020, doi: 10.1007/978-3-030-43883-8_7.

[12] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. F. Moura, and P. Eckersley, "Explainable machine learning in deployment," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2020, pp. 648–657, doi: 10.1145/3351095.3375624.

[13] E. Toreini, M. Aitken, K. Coopamootoo, K. Elliott, C. G. Zelaya, and A. van Moorsel, "The relationship between trust in AI and trustworthy machine learning technologies," 2019, *arXiv:1912.00782*.

[14] M. Al-Rubaie and J. M. Chang, "Privacy preserving machine learning: Threats and solutions," 2018, *arXiv:1804.11238*.

[15] E. De Cristofaro, "An overview of privacy in machine learning," 2020, *arXiv:2005.08679*.

[16] A. Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (XAI): A survey," 2020, *arXiv:2006.11371*.

[17] J. Schneider and J. P. Handali, "Personalized explanation for machine learning: A conceptualization," 2019, *arXiv:1901.00770*.

[18] M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, and P. Sen, "A survey of the state of explainable AI for natural language processing," 2020, *arXiv:2010.00711*.

[19] C. Mars, R. Dés, and M. Boussard, "The three stages of explainable AI: How explainability facilitates real world deployment of AI How XAI makes a difference," Tech. Rep., 2019.

[20] F. Hussain, R. Hussain, and E. Hossain, "Explainable artificial intelligence (XAI): An engineering perspective," 2021, *arXiv:2101.03613*.

[21] S. Triberti, I. Durosini, and G. Pravettoni, "A 'third wheel' effect in health decision making involving artificial entities: A psychological perspective," *Frontiers Public Heal.*, vol. 8, pp. 1–9, Apr. 2020, doi: 10.3389/fpubh.2020.00117.

[22] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Incorporating explainable artificial intelligence (XAI) to aid the understanding of machine learning in the healthcare domain," in *Proc. CEUR Workshop*, vol. 2771, Dec. 2020, pp. 169–180.

[23] S. N. Payrovnaziri, Z. Chen, P. Rengifo-Moreno, T. Miller, J. Bian, J. H. Chen, X. Liu, and Z. He, "Explainable artificial intelligence models using real-world electronic health record data: A systematic scoping review," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 7, pp. 1173–1185, Jul. 2020, doi: 10.1093/jamia/ocaa053.

[24] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Explainable AI in healthcare," in *Proc. Int. Conf. Cyber Situational Awareness, Data Anal. Assessment (CyberSA)*, Jun. 2020, pp. 1–5, doi: 10.1109/CyberSA49311.2020.9139655.

[25] 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," 2021, *arXiv:2102.01998*.

[26] D. Gunning and D. W. Aha, "DARPA's explainable artificial intelligence program," *AI Mag.*, vol. 40, no. 2, pp. 44–58, 2019, doi: 10.1609/aimag.v40i2.2850.

[27] M. A. Kamal, M. M. Alam, H. Khawar, and M. S. Mazliham, "Play and learn case study on learning abilities through effective computing in games," in *Proc. 13th Int. Conf. Math., Actuarial Sci., Comput. Sci. Statist. (MACS)*, Dec. 2019, pp. 1–6, doi: 10.1109/MACS48846.2019.9024771.

[28] F. Gurcan, N. E. Cagiltay, and K. Cagiltay, "Mapping human–computer interaction research themes and trends from its existence to today: A topic modeling-based review of past 60 years," *Int. J. Hum. Comput. Interact.*, vol. 37, no. 3, pp. 267–280, Feb. 2021, doi: 10.1080/10447318.2020.1819668.

[29] E. Kurilovas and S. Kubilinskiene, "Lithuanian case study on evaluating suitability, acceptance and use of IT tools by students—An example of applying technology enhanced learning research methods in higher education," *Comput. Hum. Behav.*, vol. 107, Jun. 2020, Art. no. 106274, doi: 10.1016/j.chb.2020.106274.

[30] K. Sagar and A. Saha, "A systematic review of software usability studies," *Int. J. Inf. Technol.*, to be published, doi: 10.1007/s41870-017-0048-1.

[31] S. R. Hong, J. Hullman, and E. Bertini, "Human factors in model interpretability: Industry practices, challenges, and needs," *Proc. ACM Hum. Comput. Interact.*, vol. 4, pp. 1–26, May 2020, doi: 10.1145/3392878.

[32] L. Punchoojit and N. Hongwarittorrn, "Usability studies on mobile user interface design patterns: A systematic literature review," *Adv. Hum. Comput. Interact.*, vol. 2017, pp. 1–22, Nov. 2017, doi: 10.1155/2017/6787504.

[33] A. N. Bazzano, J. Martin, E. Hicks, M. Faughnan, and L. Murphy, "Human-centred design in global health: A scoping review of applications and contexts," *PLoS ONE*, vol. 12, no. 11, pp. 1–24, 2017, doi: 10.1371/journal.pone.0186744.

[34] M. L. Tan, R. Prasanna, K. Stock, E. E. H. Doyle, G. Leonard, and D. Johnston, "Understanding end-users' perspectives: Towards developing usability guidelines for disaster apps," *Prog. Disaster Sci.*, vol. 7, Oct. 2020, Art. no. 100118, doi: 10.1016/j.pdisas.2020.100118.

[35] T. Radüntz, T. Mühlhausen, N. Fürstenau, E. Cheladze, and B. Meffert, *Application of the Usability Metrics of the ISO 9126 Standard in the E-Commerce Domain: A Case Study*, vol. 903. Cham, Switzerland: Springer, 2019.

[36] M. G. Siavvas, K. C. Chatzidimitriou, and A. L. Symeonidis, "QATCH—An adaptive framework for software product quality assessment," *Expert Syst. Appl.*, vol. 86, pp. 350–366, Nov. 2017, doi: 10.1016/j.eswa.2017.05.060.

[37] J. M. Ferreira, S. T. Acuña, O. Dieste, S. Vegas, A. Santos, F. Rodríguez, and N. Juristo, "Impact of usability mechanisms: An experiment on efficiency, effectiveness and user satisfaction," *Inf. Softw. Technol.*, vol. 117, Jan. 2020, Art. no. 106195, doi: 10.1016/j.infsof.2019.106195.

[38] G. A. Boy, "Human-centered design of complex systems: An experience-based approach," *Des. Sci.*, vol. 3, pp. 1–23, Jan. 2017, doi: 10.1017/dsj.2017.8.

[39] T. Farooqui, T. Rana, and F. Jafari, "Impact of human-centered design process (HCDP) on software development process," in *Proc. 2nd Int. Conf. Commun., Comput. Digit. Syst. (C-CODE)*, Mar. 2019, pp. 110–114, doi: 10.1109/C-CODE.2019.8680978.

[40] M. Arrivillaga, P. C. Bermúdez, J. P. García-Cifuentes, and J. Botero, "Innovative prototypes for cervical cancer prevention in low-income primary care settings: A human-centered design approach," *PLoS ONE*, vol. 15, no. 8, pp. 1–22, Aug. 2020, doi: 10.1371/journal.pone.0238099.

[41] T. F. P. Silva and T. F. P. Marques, "Human-centered design for collaborative innovation in knowledge-based economies," *Technol. Innov. Manage. Rev.*, vol. 10, no. 9, pp. 5–15, Sep. 2020, doi: 10.22215/TIMREVIEW/1385.

[42] F. K. Dosilovic, M. Brcic, and N. Hlupic, "Explainable artificial intelligence: A survey," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2018, pp. 210–215, doi: 10.23919/MIPRO.2018.8400040.

[43] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," in *Proc. 15th Int. Conf. ICT Knowl. Eng. (ICT&KE)*, Nov. 2017, pp. 1–6, doi: 10.1109/ICTKE.2017.8259629.

[44] M. A. Shahid, N. Islam, M. M. Alam, M. M. Su'ud, and S. Musa, "A comprehensive study of load balancing approaches in the cloud computing environment and a novel fault tolerance approach," *IEEE Access*, vol. 8, pp. 130500–130526, 2020, doi: 10.1109/ACCESS.2020.3009184.

[45] M. A. Shahid, N. Islam, M. M. Alam, M. S. Mazliham, and S. Musa, "Towards resilient method: An exhaustive survey of fault tolerance methods in the cloud computing environment," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100398, doi: 10.1016/j.cosrev.2021.100398.

[46] S. I. Hassan, M. M. Alam, U. Illahi, M. A. Al Ghamdi, S. H. Almotiri, and M. M. Su'ud, "A systematic review on monitoring and advanced control strategies in smart agriculture," *IEEE Access*, vol. 9, pp. 32517–32548, 2021, doi: 10.1109/ACCESS.2021.3057865.

[47] V. Dunjko and H. J. Briegel, "Machine learning & artificial intelligence in the quantum domain: A review of recent progress," *Rep. Prog. Phys.*, vol. 81, no. 7, Jul. 2018, Art. no. 074001, doi: 10.1088/1361-6633/aab406.

[48] H. Anandakumar and A. Ramu, "Business intelligence for enterprise Internet of Things," Tech. Rep., Jul. 2020.

[49] K. G. Kim, "Deep learning book review," *Nature*, vol. 29, no. 7553, pp. 1–73, 2019.

[50] Y. Jing, Y. Bian, Z. Hu, L. Wang, and X.-Q.-S. Xie, "Deep learning for drug design: An artificial intelligence paradigm for drug discovery in the big data era," *AAPS J.*, vol. 20, no. 3, pp. 1–10, May 2018, doi: 10.1208/s12248-018-0210-0.

[51] L. Arras, G. Montavon, K.-R. Müller, and W. Samek, "Explaining recurrent neural network predictions in sentiment analysis," 2017, *arXiv:1706.07206*.

[52] W. Xu and V. T. Hoang, "MapReduce-based improved random forest model for massive educational data processing and classification," *Mobile Netw. Appl.*, vol. 26, no. 1, pp. 191–199, Feb. 2021, doi: 10.1007/s11036-020-01699-w.

[53] C. F. Breidbach and P. Maglio, "Accountable algorithms? The ethical implications of data-driven business models," *J. Service Manage.*, vol. 31, no. 2, pp. 163–185, May 2020, doi: 10.1108/JOSM-03-2019-0073.

[54] V. Mugunthan, R. Rahman, and L. Kagal, "BlockFLow: An accountable and privacy-preserving solution for federated learning," 2020, *arXiv:2007.03856*.

[55] H. B. Desai, M. S. Ozdayi, and M. Kantarcioglu, *BlockFLA: Accountable Federated Learning via Hybrid Blockchain Architecture*, vol. 1, no. 1. New York, NY, USA: Association for Computing Machinery, 2020.

[56] B. Nushi, E. Kamar, and E. Horvitz, "Towards accountable AI: Hybrid human-machine analyses for characterizing system failure," 2018, *arXiv:1809.07424*.

[57] C. E. Tuncali, G. Fainekos, H. Ito, and J. Kapinski, "Simulation-based adversarial test generation for autonomous vehicles with machine learning components," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1555–1562, doi: 10.1109/IVS.2018.8500421.

[58] C. E. Tuncali, G. Fainekos, D. Prokhorov, H. Ito, and J. Kapinski, "Requirements-driven test generation for autonomous vehicles with machine learning components," *IEEE Trans. Intell. Vehicles*, vol. 5, no. 2, pp. 265–280, Jun. 2020, doi: 10.1109/TIV.2019.2955903.

[59] S. O'Sullivan, N. Nevejans, C. Allen, A. Blyth, S. Leonard, U. Pagallo, K. Holzinger, A. Holzinger, M. I. Sajid, and H. Ashrafian, "Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 15, no. 1, pp. 1–12, 2019, doi: 10.1002/rcs.1968.

[60] S. Corbett-Davies and S. Goel, "The measure and mismeasure of fairness: A critical review of fair machine learning," 2018, *arXiv:1808.00023*.

[61] H. Chang, T. D. Nguyen, S. K. Murakonda, E. Kazemi, and R. Shokri, "On adversarial bias and the robustness of fair machine learning," 2020, *arXiv:2006.08669*.

[62] N. Chen, A. Klushyn, A. Paraschos, D. Benbouzid, and P. Van der Smagt, "Active learning based on data uncertainty and model sensitivity," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1547–1554, doi: 10.1109/IROS.2018.8593552.

[63] R. M. J. Byrne, "Counterfactuals in explainable artificial intelligence (XAI): Evidence from human reasoning," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 6276–6282, doi: 10.24963/ijcai.2019/876.

[64] I. Stepin, J. M. Alonso, A. Catala, and M. Pereira-Farina, "A survey of contrastive and counterfactual explanation generation methods for explainable artificial intelligence," *IEEE Access*, vol. 9, pp. 11974–12001, 2021, doi: 10.1109/ACCESS.2021.3051315.

[65] T. Spinner, U. Schlegel, and M. El-assady, "ExplAIner: A visual analytics framework for interactive and explainable machine learning," Tech. Rep., 2019.

[66] C. Yeung, J. M. Tsai, Y. Kawagoe, B. King, D. Ho, and A. P. Raman, "Elucidating the design and behavior of nanophotonic structures through explainable convolutional neural networks," Aug. 2020, *arXiv:2003.06075*.

[67] E. Alcobaça, S. M. Mastelini, T. Botari, B. A. Pimentel, D. R. Cassar, A. C. P. D. L. F. de Carvalho, and E. D. Zanotto, "Explainable machine learning algorithms for predicting glass transition temperatures," *Acta Mater.*, vol. 188, pp. 92–100, Apr. 2020, doi: 10.1016/j.actamat.2020.01.047.

[68] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "A trustworthy privacy preserving framework for machine learning in industrial IoT systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6092–6102, Sep. 2020, doi: 10.1109/TII.2020.2974555.

[69] M. Nassar, K. Salah, M. H. ur Rehman, and D. Svetinovic, "Blockchain for explainable and trustworthy artificial intelligence," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 1, Jan. 2020, Art. no. e1340, doi: 10.1002/widm.1340.

[70] T. Hunt, C. Song, R. Shokri, V. Shmatikov, and E. Witchel, "Chiron: Privacy-preserving machine learning as a service," 2018, *arXiv:1803.05961*.

[71] P. Mohassel and Y. Zhang, "SecureML: A system for scalable privacy-preserving machine learning," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 19–38, doi: 10.1109/SP.2017.12.

[72] Q. Yang, N. Banovic, and J. Zimmerman, "Mapping machine learning advances from HCI research to reveal starting places for design innovation," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–11, doi: 10.1145/3173574.3173704.

[73] J. Wetzel, H. Burkhardt, S. Cheema, S. Kang, D. Pead, A. Schoenfeld, and K. VanLehn, "A preliminary evaluation of the usability of an ai-infused orchestration system," in *Artificial Intelligence in Education*, vol. 10948, 2018, pp. 379–383, doi: 10.1007/978-3-319-93846-2_71.

[74] C. Kolski, "Interaction and artificial intelligence to cite this version: HAL Id: HAL-02424944 cross-fertilisation between human-computer interaction and artificial intelligence," Tech. Rep., 2020.

[75] Y. Liu, Y. Wang, Y. Bian, L. Ren, and Y. Xuan, "A psychological model of human-computer cooperation for the era of artificial intelligence," *SCIENTIA SINICA Inf.*, vol. 48, no. 4, pp. 376–389, Apr. 2018, doi: 10.1360/n112017-00225.

[76] X. Fan, J. Fan, F. Tian, and G. Dai, "Human-computer interaction and artificial intelligence: From competition to integration," *SCIENTIA SINICA Inf.*, vol. 49, no. 3, pp. 361–368, Mar. 2019, doi: 10.1360/n112018-00181.

[77] F. Tian, J. Fan, G. Dai, Y. Du, and Z. Liu, "Thoughts on human-computer interaction in the age of artificial intelligence," *SCIENTIA SINICA Inf.*, vol. 48, no. 4, pp. 361–375, Apr. 2018, doi: 10.1360/n112017-00221.

[78] F. Topak and M. K. Pekeriçli, "Towards using human-computer interaction research for advancing intelligent built environments: A review," in *Proc. 6th Int. Project ConstrUction Manage. Conf.*, 2020, p. 835.

[79] *Holy Grail of AI for Enterprise—Explainable AI.* [Online]. Available: https://www.kdnuggets.com/2018/10/enterprise-explainable-ai.html

[80] W. Xu and I. Corporation, "A perspective from human-computer interaction," Tech. Rep., 2019.

[81] M. Eiband, D. Buschek, A. Kremer, and H. Hussmann, "The impact of placebic explanations on trust in intelligent systems," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 97–105, doi: 10.1145/3290607.3312787.

[82] D. Shin, "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI," *Int. J. Hum.-Comput. Stud.*, vol. 146, Feb. 2021, Art. no. 102551, doi: 10.1016/j.ijhcs.2020.102551.

[83] P. Linardatos, V. S. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, pp. 1–45, 2021, doi: 10.3390/e23010018.

[84] L. Chazette and K. Schneider, "Explainability as a non-functional requirement: Challenges and recommendations," *Requirements Eng.*, vol. 25, no. 4, pp. 493–514, Dec. 2020, doi: 10.1007/s00766-020-00333-1.

[85] J. Wanner, L.-V. Herm, K. Heinrich, C. Janiesch, and P. Zschech, "White, grey, black: Effects of XAI augmentation on the confidence in AI-based decision support systems," in *Proc. 41st Int. Conf. Inf. Syst.*, Sep. 2020, pp. 1–9.

[86] J. J. Ferreira and M. de Souza Monteiro, "Evidence-based explanation to promote fairness in AI systems," 2020, *arXiv:2003.01525*.

[87] *Designing Accessible XAI*.

[88] Y. Nakao, K. Ohori, and H. Anai, "Interactive recommendation AI to support transparent human decision making," *Fujitsu Sci. Tech. J.*, to be published. [Online]. Available: https://www.fujitsu.com/global/documents/about/resources/publications/technicalreview/2020-01/article03.pdf.

[89] P. Madumal, L. Sonenberg, T. Miller, and F. Vetere, "Explainable AI through rule-based interactive conversation Christian," in *Proc. Int. Jt. Conf. Auton. Agents Multiagent Syst. (AAMAS)*, vol. 2, 2018, pp. 1033–1041.

[90] L. Sanneman and J. A. Shah, "A situation awareness-based framework for design and evaluation of explainable AI," in *Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, 2019, doi: 10.1007/978-3-030-51924-7_6.

[91] T. Chakraborti, S. Sreedharan, and S. Kambhampati, "The emerging landscape of explainable automated planning & decision making," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 4803–4811.

[92] V. Arya, R. K. E. Bellamy, P.-Y. Chen, A. Dhurandhar, M. Hind, S. C. Hoffman, S. Houde, Q. V. Liao, R. Luss, A. Mojsilović, S. Mourad, P. Pedemonte, R. Raghavendra, J. Richards, P. Sattigeri, K. Shanmugam, M. Singh, K. R. Varshney, D. Wei, and Y. Zhang, "One explanation does not fit all: A toolkit and taxonomy of AI explainability techniques," 2019, *arXiv:1909.03012*.

[93] L. Yang, Z. Zhang, S. Xiong, L. Wei, J. Ng, L. Xu, and R. Dong, "Explainable text-driven neural network for stock prediction," 2019, *arXiv:1902.04994*.

[94] R. Chhatwal, P. Gronvall, N. Huber-Fliflet, R. Keeling, J. Zhang, and H. Zhao, "Explainable text classification in legal document review a case study of explainable predictive coding," 2019, *arXiv:1904.01721*.

[95] Y. Zhang, Q. V. Liao, and R. K. E. Bellamy, "Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2020, pp. 295–305, doi: 10.1145/3351095.3372852.

[96] R. Guidotti, "Evaluating local explanation methods on ground truth," *Artif. Intell.*, vol. 291, Feb. 2021, Art. no. 103428, doi: 10.1016/j.artint.2020.103428.

[97] V. Beaudouin, I. Bloch, D. Bounie, S. Clémençon, F. d'Alché-Buc, J. Eagan, W. Maxwell, P. Mozharovskyi, and J. Parekh, "Flexible and context-specific AI explainability: A multidisciplinary approach," 2020, *arXiv:2003.07703*.

[98] K. Fauvel, V. Masson, and É. Fromont, "A performance-explainability framework to benchmark machine learning methods: Application to multivariate time series classifiers," 2020, *arXiv:2005.14501*.

[99] J. Adams and H. Hagras, "A type-2 fuzzy logic approach to explainable AI for regulatory compliance, fair customer outcomes and market stability in the global financial sector," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2020, pp. 1–8.

[100] J. Choo, "Subject index," Tech. Rep., 2000, pp. 273–279, doi: 10.1016/s0065-2296(00)33046-4.

[101] A. Chatzimparmpas, R. M. Martins, I. Jusufi, K. Kucher, F. Rossi, and A. Kerren, "The state of the art in enhancing trust in machine learning models with the use of visualizations," *Comput. Graph. Forum*, vol. 39, no. 3, pp. 713–756, Jun. 2020, doi: 10.1111/cgf.14034.

[102] K. Atkinson, T. Bench-Capon, and D. Bollegala, "Explanation in AI and law: Past, present and future," *Artif. Intell.*, vol. 289, no. 1920, pp. 1–39, 2020, doi: 10.1016/j.artint.2020.103387.

[103] M. Shardlow, "A survey of automated text simplification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 1, pp. 58–70, 2014, doi: 10.14569/specialissue.2014.040109.

[104] D. Osaku, J. F. Gomes, and A. X. Falcão, "Convolutional neural network simplification with progressive retraining," 2021, *arXiv:2101.04699*.

[105] J. M. Rožanec and D. Mladenić, "Semantic XAI for contextualized demand forecasting explanations," 2021, *arXiv:2104.00452*.

[106] N. Baracaldo, B. Chen, H. Ludwig, and J. A. Safavi, "Mitigating poisoning attacks on machine learning models: A data provenance based approach," in *Proc. 10th ACM Workshop Artif. Intell. Secur.*, Nov. 2017, pp. 103–110, doi: 10.1145/3128572.3140450.

[107] K. Sokol, A. Hepburn, R. Santos-Rodriguez, and P. Flach, "BLIMEy: Surrogate prediction explanations beyond LIME," 2019, *arXiv:1910.13016*.

[108] R. Alizadeh, J. K. Allen, and F. Mistree, "Managing computational complexity using surrogate models: A critical review," *Res. Eng. Des.*, vol. 31, no. 3, pp. 275–298, Jul. 2020, doi: 10.1007/s00163-020-00336-7.

[109] M. Atzmueller, "Declarative aspects in explicative data mining for computational sensemaking," in *Proc. Int. Conf. Appl. Declarative Program. Knowl. Manage.*, vol. 1099, 2018, pp. 97–114, doi: 10.1007/978-3-030-00801-7_7.

[110] S. Vojíř and T. Kliegr, "Editable machine learning models? A rule-based framework for user studies of explainability," *Adv. Data Anal. Classification*, vol. 14, no. 4, pp. 785–799, Dec. 2020, doi: 10.1007/s11634-020-00419-2.

[111] J. Fürnkranz, T. Kliegr, and H. Paulheim, "On cognitive preferences and the plausibility of rule-based models," *Mach. Learn.*, vol. 109, no. 4, pp. 853–898, Apr. 2020, doi: 10.1007/s10994-019-05856-5.

[112] W. Guo, "Explainable artificial intelligence for 6G: Improving trust between human and machine," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 39–45, Jun. 2020, doi: 10.1109/MCOM.001.2000050.

[113] L. Longo, R. Goebel, F. Lecue, P. Kieseberg, and A. Holzinger, "Explainable artificial intelligence: Concepts, applications, research challenges and visions," in *Machine Learning and Knowledge Extraction*, vol. 12279, Aug. 2020, pp. 1–16, doi: 10.1007/978-3-030-57321-8_1.

[114] Y.-S. Lin, W.-C. Lee, and Z. B. Celik, "What do you see? Evaluation of explainable artificial intelligence (XAI) interpretability through neural backdoors," 2020, *arXiv:2009.10639*.

[115] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI—Explainable artificial intelligence," *Sci. Robot.*, vol. 4, no. 37, p. 1, Dec. 2019, doi: 10.1126/scirobotics.aay7120.

[116] P. Ponsa and D. Guasch, "A human–computer interaction approach for healthcare," *Universal Access Inf. Soc.*, vol. 17, no. 1, pp. 1–3, Mar. 2018, doi: 10.1007/s10209-016-0515-7.

[117] I. Clarke, O. Oluwaseun, and I. Rhoda, "State of the art: A study of human-robot interaction in healthcare," *Int. J. Inf. Eng. Electron. Bus.*, vol. 9, no. 3, pp. 43–55, May 2017, doi: 10.5815/ijieeb.2017.03.06.

[118] O. Korn, "Social robots—A new perspective in healthcare," *Res. Outreach*, vol. 114, pp. 78–81, 2020, doi: 10.32907/ro-114-7881.

[119] K. Stowers and M. Mouloua, "Human computer interaction trends in healthcare: An update," in *Proc. Int. Symp. Hum. Factors Ergon. Heal. Care*, vol. 7, no. 1, 2018, pp. 88–91, doi: 10.1177/2327857918071019.

[120] F. Wiser, C. Durst, and N. Wickramasinghe, "Activity theory: A comparison of HCI theories for the analysis of healthcare technology," Tech. Rep., 2018, pp. 235–249, doi: 10.1007/978-3-319-72287-0_15.

[121] A. Søgaard Neilsen and R. L. Wilson, "Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: An integrative literature review," *Int. J. Mental Health Nursing*, vol. 28, no. 1, pp. 22–39, Feb. 2019, doi: 10.1111/inm.12527.

[122] A. Ahmad and P. Mozelius, "Critical factors for human computer interaction of eHealth for older adults," in *Proc. the 5th Int. Conf. e-Soc., e-Learn. e-Technol. (ICSLT)*, 2019, pp. 58–62, doi: 10.1145/3312714.3312730.

[123] B. Kaysi and E. K. Kesler, "Human computer interaction and visualization tools in health care services," in *Proc. Int. Conf. Inf. Knowl. Eng.*, vol. 1, 2018, pp. 55–61. [Online]. Available: https://csce.ucmss.com/cr/books/2018/LFS/CSREA2018/IKE3668.pdf

[124] A. Melder, T. Robinson, I. McLoughlin, R. Iedema, and H. Teede, "An overview of healthcare improvement: Unpacking the complexity for clinicians and managers in a learning health system," *Internal Med. J.*, vol. 50, no. 10, pp. 1174–1184, Oct. 2020, doi: 10.1111/imj.14876.

[125] A. Blandford, "HCI for health and wellbeing: Challenges and opportunities," *Int. J. Hum.-Comput. Stud.*, vol. 131, pp. 41–51, Nov. 2019, doi: 10.1016/j.ijhcs.2019.06.007.

[126] M. E. Matheny, D. Whicher, and S. T. Israni, "Artificial intelligence in health care: A report from the national academy of medicine," *JAMA J. Amer. Med. Assoc.*, vol. 323, no. 6, pp. 509–510, 2020, doi: 10.1001/jama.2019.21579.

[127] A. Mahajan, T. Vaidya, A. Gupta, S. Rane, and S. Gupta, "Artificial intelligence in healthcare in developing nations: The beginning of a transformative journey," *Cancer Res., Statist., Treatment*, vol. 2, no. 2, p. 182, 2019, doi: 10.4103/crst.crst_50_19.

[128] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: Past, present and future," *Stroke Vascular Neurol.*, vol. 2, no. 4, pp. 230–243, 2017, doi: 10.1136/svn-2017-000101.

[129] N. Noorbakhsh-Sabet, R. Zand, Y. Zhang, and V. Abedi, "Artificial intelligence transforms the future of health care," *Amer. J. Med.*, vol. 132, no. 7, pp. 795–801, Jul. 2019, doi: 10.1016/j.amjmed.2019.01.017.

[130] S. Kanza and J. G. Frey, "A new wave of innovation in semantic web tools for drug discovery," *Expert Opinion Drug Discovery*, vol. 14, no. 5, pp. 433–444, May 2019, doi: 10.1080/17460441.2019.1586880.

[131] K. Fritchman, K. Saminathan, R. Dowsley, T. Hughes, M. De Cock, A. Nascimento, and A. Teredesai, "Privacy-preserving scoring of tree ensembles: A novel framework for AI in healthcare," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 2413–2422, doi: 10.1109/BigData.2018.8622627.

[132] H. D. Dataset, "Explainable AI meets healthcare: A study on heart disease dataset," 2020, *arXiv:2011.03195*.

[133] A. Adadi and M. Berrada, "Explainable AI for healthcare: From black box to interpretable models," in *Embedded Systems and Artificial Intelligence*, 2020, pp. 327–337, doi: 10.1007/978-981-15-0947-6_31.

[134] A. Lucieri, A. Dengel, and S. Ahmed, "Deep learning based decision support for medicine—A case study on skin cancer diagnosis," 2021, *arXiv:2103.05112*.

[135] M. Winckler and U. Chatterjee, *Human Computer Interaction and Emerging Technologies: Workshop Proceedings From the INTERACT 2019 Workshops*, 2020.

MOBEEN NAZAR received the bachelor's and master's degrees in software engineering. She is currently pursuing the Ph.D. degree in information technology with the Malaysian Institute, Universiti Kuala Lumpur, Malaysia. She is also associated with the Department of Software Engineering, Bahria University, Karachi Campus, as a Faculty Member. Her research interests include machine learning, artificial intelligence, human–computer interaction, usability engineering, and explainable artificial intelligence.

MUHAMMAD MANSOOR ALAM received the M.S. degree in system engineering and the M.Sc. degree in computer science from France, U.K., and Malaysia, and the Ph.D. degree in computer engineering and the Ph.D. degree in electrical and electronics engineering. He is currently a Professor of computer science. He is also working as an Associate Dean with CCSIS and the HOD of the Department of Mathematics, Statistics, and Computer Science. He is working as an Adjunct Professor with Universiti Kuala Lumpur (UniKL) and supervising 12 Ph.D. students. He is enjoying 20 years of research and teaching experience in Canada, England, France, Malaysia, Saudi Arabia, and Bahrain. He has the honor to work as an Online Laureate (Facilitator) for the MSIS Program run by Colorado State University, USA, and Saudi Electronic University, Saudi Arabia. He has also established research collaboration with UniKL and Universiti Malaysia Pahang (UMP). He has done a postdoctoral research from Malaysia in machine learning approaches for efficient prediction and decision making. He has authored more than 150 research articles which are published in well-reputed journals of high impact factor, Springer Link book chapters, and Scopus indexed journals and IEEE conferences. The Universite de LaRochelle awarded him Très Honorable (Hons.) Ph.D. due to his research impact during his Ph.D. degree.

EIAD YAFI (Senior Member, IEEE) received the B.Sc. degree in mathematics from Al-Baath University, in 2000, and the master's degree in computer applications and the Ph.D. degree in computer science from Jamia Hamdard University, India, in 2011. His research interests include machine learning, data analytics, cloud computing, human–computer interaction, and ICTD and blockchain.

MAZLIHAM MOHD SU'UD received the Diploma degree in science, the bachelor's degree in electronics electrotechnics and automation, the master's degree in electronics electrotechnics and automation, and the master's degree in electronics from the University de Montpellier II, France, the master's degree in electrical and electronics engineering from the University of Montpellier, in 1993, the Ph.D. degree in computational intelligence and decision from the University De La Rochelle, France, and the Ph.D. degree in computer engineering from the Université de La Rochelle, in 2007. Since 2013, he has been the President/the CEO of Universiti Kuala Lumpur, Malaysia. He has vast experience of publishing in high quality international scientific journals and conference proceedings. He has numerous years of experience in the industrial and academic field.