Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 3, No. 3, 2023

Impact and Prediction of AI Diagnostic Report Interpretation Type on Patient Trust

Yunqing Tang [1], Jinliang Cai [2]

[1] Management & Economics of Hebei University of Technology, Tianjin, China
[2] School of Nanchang University, Jiangxi, China

Abstract: With the rapid development of AI technology and the rise of AI in health care, AI diagnostic techniques are gaining attention. Studies have been conducted to enhance the reliability of AI in terms of algorithmic accuracy and "black box" nature, but few studies have explored the impact of AI interpretation type on patient trust. In this paper, we use subjective scales and objective eye-tracking techniques based on the elaboration likelihood model (ELM) and cognitive load theory to explore the trust and prediction of patients with different health literacy on global and partial interpretations of AI diagnostic reports. Firstly, based on the existing AI diagnostic report form, we remove the distracting information and restore the AI diagnostic report display elements by Axure RP9, and construct the patient health literacy and patient trust evaluation scales using the questionnaire method; then we conduct scenario simulation experiments using eye-tracking technology to analyze and compare the patient trust perception and objective eye-movement measurement results; finally, we use Pearson correlation test. Partial least squares method was used to construct a relationship model between patient trust and eye movement index, and the validity of the model was verified. The results showed that patients with different health literacy differed in their trust in different AI interpretation types; patients with different health literacy differed in their gaze levels for different interpretation types of diagnostic reports; and the relationship model between patient trust and eye movement indicators could effectively predict patient perceived trust. The results of the study complement the research on the calibration trust of eye-tracking technology in the medical field, while providing a reliable scientific basis for the design and developers of intelligent diagnostic technology applications.

Keywords: Artificial Intelligence Diagnosis; AI Interpretation Type; Patient Trust; Eye Movement Behavior.

1. Introduction

In recent years, artificial intelligence has emerged in medical care, with AI technologies helping doctors diagnose or examine cases, such as deep neural networks and knowledge graphs that have been used in genetic diagnosis or imaging [1,2]. There is growing evidence that patients are becoming interested in AI systems[3,4], but because of the "black box" nature of AI systems[5,6], users cannot see the inner workings of the system, have difficulty understanding AI systems, and have difficulty determining whether they can trust intelligent diagnoses and make decisions, despite many efforts by academics[7-9], there is still no widespread acceptance by users Despite many efforts by scholars[10], there is still no general acceptance by users . Because the healthcare industry is a high-risk field without rigorous evaluation criteria and extensive experience, it is difficult for designers and developers of AI diagnostic systems to obtain guidance on how to develop meaningful, interpretable, and highly acceptable AI diagnostic systems [11,12]. Research has been done to theoretically sort out AI explanations and demonstrate the importance of AI system explanations [13-16], which have focused on developing and improving system models to make AI systems more interpretable [17-19]. In addition, studies have emphasized the importance of explaining AI predictions for users to understand and accept the results provided by the system [11,20].

It has been shown that users have a cautious attitude towards the decision making of AI systems and when there is a discrepancy between the user's idea and the decision outcome of the AI system, the user will want to understand the operating mechanism of the AI system and form user perceptions to promote trust and preference for the system through the processing of information explained by the system [21]. In healthcare, it has been shown that lack of trust in AI by physicians and patients is a significant barrier to the adoption of AI systems in medical care[22,23], Asan et al.[24] showed that lack of trust in AI systems is a major issue in clinical adoption of this technology, and clinicians may have bias towards AI leading to trust deficits in the technology, which hinders the acceptance and adoption of AI, physicians' lack of trust in AI lack of trust in the system may result in misuse of devices and waste of resources, and lack of trust in the AI system by patients may cause congestion in the hospital system, both of which can be barriers to the development of AI systems within the healthcare environment. The literature suggests that a major barrier to facilitating user trust and acceptance of AI systems is the "black box" problem [25], where users cannot understand why an AI system produces information output [6]. To address this issue, researchers suggest making systems more transparent to enhance user trust in AI [20,26,27].

From the user perspective, user trust is the emotional experience of user interaction with an interface; therefore, understanding the relationship between interpretable AI and user trust can help in the design and development of AI systems. To date, several types of metrics have been used for measuring user trust, such as subjective ratings, behavioral data, and psychophysiological measures. The vast majority of studies have relied on rating scales to measure [28], but this method does not measure change over time in real time. With the development of physiological measurement techniques, it has become possible to measure dynamic emotions during user interactions, and eye-tracking techniques are beginning to be noticed and used as a continuous, real-time, non-invasive way to analyze emotions during user interactions

[29,30]. Lu et al. [31] applied eye-tracking techniques to automated trust and system reliability calibration studies and confirmed that eye-tracking techniques are reliable tools for inferring trust and supporting trust calibration; Shanet al. [32] extracted user valued information and constructed a cognitive trust model by eye tracking technique. In summary, the existing results have had good theoretical and practical value in their respective fields, but there are still shortcomings in the field of medical high-risk: (i) few studies on explainable AI systems have empirically explored different types of explanations on patient trust; (ii) few studies have considered the impact of patient health literacy on trust in explainable AI systems from the perspective of patient information processing and processing; (iii) few studies have explored patients' trust in interpretable AI based on eye-movement metrics in the context of high risk in healthcare. To address the above deficiencies, this paper uses a combination of questionnaire evaluation scale and eye-tracking technology to analyze patient eye-movement indicators and subjective trust evaluations, and constructs and validates a patient trust prediction model for interpretable AI diagnostic reports.

2. Theoretical Basis and Literature Review

2.1. Elaboration Likelihood Model

Psychologists Petty and Cacioppo proposed the Elaboration likelihood Model (ELM) in 1980, which is considered to be the most influential theoretical model of consumer information processing. "Elaboration" refers to the degree to which individuals think about information and process it in different ways depending on their "motivation" and "ability" to influence their decision attitudes Petty and Cacioppo confirm that there are two main ways in which people process information: (i) When individuals have high "motivation" and "ability", they think carefully about the content of the information, and evaluate the concepts and logic proposed by the information in depth, so as to decide whether to change their attitude toward things, and this way of information processing is called the "central path". (ii) When individuals are weak in "motivation" and "ability", they are often not willing to go through more efforts and process and analyze the information, and are more easily persuaded by the apparent characteristics, and use heuristic clues to understand the information through peripheral factors, and decide the credibility of the information, and this way of information processing is called the "edge path"[33]. In recent years, the ELM model has been further developed in the study of online users' information reception. Studies have combined the ELM model with visual perceptual physiological signals to study user behavior and its effects. Trivedi et al. [34] based on ELM model combined with eye-tracking technology to explore how user information access, message characteristics and social media users' health literacy level predict the evaluation of message credibility. Wan et al. [35] combine ELM model with eye-tracking technology from the perspective of visual perception to explore the relationship between community users' information browsing behavior and message judgment.

In this paper, based on the classification of explanation types and ELM model research, AI explanation types are classified into two categories according to two paths of user information processing: (i) central path - local explanation (AI diagnosis report gives diagnosis results based on the degree of correlation between illness and disease) requires patients to extract the degree of matching between symptoms and disease after comparative analysis to determine the disease they are suffering from; (ii) Edge path global interpretation (AI diagnostic report gives the overall analysis logic of diagnostic results) does not require much effort and information processing. In addition, patients were grouped according to their health literacy levels to explore the impact of patient health literacy and AI interpretation type on patient trust.

2.2. Cognitive Load Theory

Cognitive psychologist John Sweller first mentioned the cognitive load theory in 1988, which states that cognitive load is a multidimensional structure that represents the load on the cognitive system of the learner when processing a specific task, to reflect the causal dimension of the interaction between the learning task andlearner characteristics, and to reflect the assessment dimension of measurable concepts such as mental load, mental effort, and performance of the learner. Learning tasks mainly include task form, task complexity, and time pressure, and learner characteristics mainly include expertise level, age, and spatial ability [36]. Sweller[37] proposes that cognitive load consists of two independent sources: internal cognitive load and external cognitive load. Internal cognitive load is determined by the interaction of information in the learning material and the learners' expertise level, while external cognitive load is generated by poorly designed instructional design. Previous authors measured learners' cognitive effort during learning by means of subjective scale measures [38], but due to the limitations of subjective and discontinuous scales, with the development of physiological measurement techniques, many scholars have adopted physiological measures for the study of learners' cognitive load [39-41]. Örün et al.[42]used EEG to study the effects of different learning environments on students' cognitive load under multitasking; Holden et al.[43] and Khairat et al.[44] explored the effect of electronic health records on patient safety hazards by highlighting the cognitive performance of physicians' electronic health records; Wu et al.[45] used eye-tracking devices to refine the effect of information display methods in different scenarios on users' cognition and memory effects based on perceptual load theory, providing a realistic basis for human-computer interaction.

This paper explores the evaluation and prediction of patient health literacy level and AI interpretation type on patient trust with the help of eye-tracking technology based on Sweller's concept of independent sources of internal and external cognitive load by considering patient health literacy level as a factor in internal cognitive load and AI interpretation type as a factor in external cognitive load.

3. Research Model

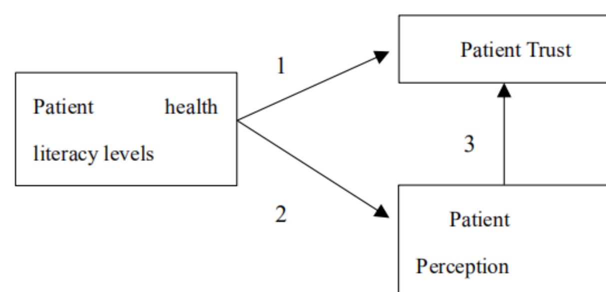

Fig 1. Conceptual model diagram

Study Objective 1: To analyze the impact of the type of AI interpretation and the level of patient health literacy on patient trust.

In this paper, we classify AI interpretation types into two categories, local interpretation and global interpretation, based on the two pathways of patient information processing based on the interpretation type classification and ELM model study. In addition, patient types are distinguished based on health literacy levels to explore the effects of patient health literacy levels and AI interpretation types on patient trust.

Study Objective 2: To explore the differences in cognitive load of patients reading diagnostic reports with different interpretation types by analyzing patient gaze levels.

Based on previous studies, this paper explores the effects of health literacy level and AI interpretation type on patients' cognitive load by considering health literacy as a factor in internal cognitive load and AI interpretation type as a factor in external cognitive load with the help of eye-tracking technology.

Study Objective 3: To predict patient trust by patient gaze duration and gaze frequency.

This paper explores patients' perceived trust in different interpretation types of diagnostic reports of AI diagnostic systems through eye-movement gaze metrics to provide reference for designers and developers of AI systems. The study selects the number of gaze and gaze duration as the eye-movement metrics in this paper based on the related user perception [36] and information processing eye-movement metrics [32,37] research.

4. Research Methodology

4.1. Experimental Subjects

Sixty-six school volunteers were recruited as subjects to participate in the experiment, and subjects whose visual acuity could not be calibrated in the oculomotor experiment due to visual astigmatism were excluded. 64 valid samples were obtained, with 48.4% of males and 51.6% of females, and all subjects had normal or corrected visual acuity in both eyes, no history of allergies or major diseases, and were informed for the follow-up experiment.

4.2. Experimental Materials

Based on the existing programs and APPs that provide AI disease self-diagnosis, the well-known Left Hand Doctor and Coretek Q Medical were selected as the AI self-diagnosis primary materials for the experiment. In order to prevent the interference of other information elements in the interface, remove the interfering information elements and ensure the original design information layout as much as possible, Axure RP9 was used to simulate the results of the patient's diagnosis report, and two web interfaces were designed to present the diagnosis report in a global interpretation and in a partial interpretation as the experimental materials, and the two diagnosis reports were identical except for the different types of interpretation.

4.3. Experimental Equipment and Eye Movement Index Selection

In this paper, the Tobii T120 (Tobii Technology) eye-tracking device was used. The eye-tracking device has a built-in eye-tracking server integrated into a 17-inch thin film transistor display with a resolution of 1280 × 1024 and a sampling rate of 120 Hz. Subjects are required to view the experimental material on the display, and the eye-tracking device automatically records and saves the subject's eye-tracking data through the Tobii Pro Glasses Controller software, and extracts the subject's eye-tracking data through ErgoLAB 3.0 The subjects' eye movement data were extracted using ErgoLAB 3.0. Based on the related user trust [34,35] and information processing eye movement index [46,47] research, the number of gaze and gaze duration were selected as the eye movement indexes in this paper.

4.4. Experimental Design and Procedure

The artificial intelligence diagnosis report was designed to meet the research content, and the user consultation process was simulated according to the existing artificial intelligence diagnosis system, thus making the subjects more realistic in the experimental process, and the task experimental scenario was as follows:

Suppose you have been suffering from runny nose, nasal congestion and headache for a long time, runny nose for more than two weeks, nasal congestion and headache for about a week, with symptoms of sore throat, dry throat and itchy nose. The color of the nasal discharge is white and sticky, and the nasal congestion is often smooth in one nostril and not in the other, and occasionally smooth at the same time, repeatedly. The headache is located in the position above the temples, and the pain level is bearable. Now you choose artificial intelligence disease self-diagnosis system to detect the condition.

The whole experiment was conducted in the oculomotor laboratory with only one subject and the instructor for each experiment to avoid interference from other people. The subjects were first briefly introduced to the purpose of the experiment, the procedure and precautions, and were asked to fill in the personal information form and the health literacy level scale. The subject was then asked to wear an eye-tracking device to calibrate the eye-tracking system. Afterwards, the subject was guided to sit in a comfortable position in a chair, and the Tobii Pro Glasses 2 software was activated to position the subject's line of sight in the display at the best range of information, and the subject's line of sight was further calibrated to ensure that the subject's eye movement data were accurately collected. After the subjects understood the simulated condition, they entered the disease self-diagnosis system, selected the corresponding disease and interacted with the AI diagnosis system, and finally received the disease diagnosis report. After reading the diagnostic report, the subjects completed the trust assessment scale, which was based on a five-star Likert scale. Other interference factors such as noise were strictly controlled during the whole experiment.

4.5. Experimental Data Processing

The oculomotor experimental data were derived by ErgoLAB 3.0 data processing software for each plotted dynamic area of interest for the duration of gaze and number of gazes in each area of interest during the experimental time, patient trust data were summarized by Excel, and the data were grouped and analyzed using SPSS software, and the oculomotor index and patient trust data were modeled using MATLAB.

5. Analysis of Data Results

5.1. Patient Trust Analyst

The differences in the effects of AI diagnosis report interpretation type and health literacy level on patient trust were analyzed by independent sample t-test. The effect of the type of AI interpretation on patient trust under the difference in the level of expertise was first analyzed, and the results are shown in Table 1: (i) patients with weak ealth literacy perceived trust in the local interpretation (M_{global}=2.50, M_{local}=3.90, p=0.000) were significantly higher than global explanations; (ii) patients with strong health literacy had significantly higher perceived trust in global explanations (M_{global}=4.24, M_{local}=3.70, P=0.010) were more significant.

Table 1. Patient health literacy t-test table

	Conditions					
	Weak health literacy			Strong health literacy		
	M_{global}	M_{local}	p-value	M_{global}	M_{local}	p-value
trust	2.50	3.90	0.000***	4.24	3.70	0.010**

5.2. Analysis of Eye Movement Data Results

The effects of AI interpretation type and health literacy level on patients' gaze levels were analyzed by independent samples t-test. From the perspective of this paper on the internal cognitive load hypothesis to analyze the differences in gaze of patients with different health literacy underwater and flat, the results are shown in Table 2: (i) under the global explanation type, the gaze duration and gaze number of patients with strong health literacy were significantly smaller than those with weak health literacy; (ii) under the local explanation type, the gaze duration and gaze number of patients with strong health literacy were significantly smaller than those with weak health literacy.

Table 2. Differences in gaze of patients with different levels of health literacy

Explanation Type	Health Literacy	Attention span			Number of gazes		
		Average value	t-value	p-value	Average value	t-value	p-value
Global	Strong	76.072	-7.496	0.000***	124.30	-2.270	0.040*
	weak	100.372			141.33		
Parti	Strong	45.117	-13.175	0.000***	92.113	-13.180	0.000***
	weak	67.00			161.00		

The differences in patients' gaze under different AI interpretation types were analyzed from the perspective of the hypothesis of external cognitive load in this paper, and the results are shown in Table 3: (i) patients with strong health literacy had significantly less gaze duration and number of gaze for the local interpretation than for the full interpretation; (ii) patients with weak health literacy had no statistically significant differences in gaze duration for the global interpretation and the local interpretation, but the number of gaze for the local interpretation was significantly higher than for the global (ii) The difference in gaze duration between global and local explanations was not statistically significant.

Table 3. Differences in user attention under different AI interpretation types

Health literacy	Interpretation type	Fixation duration			Number of gazes		
		Average value	t value	p value	Average value	t value	p value
strong	global	76.072	8.949	0.000***	124.30	9.098	0.000***
	local	45.117			67.00		
weak	global	100.372	1.431	0.180	141.33	-2.315	0.041*
	local	96.970			161.00		

5.3. Patient Trust Prediction Model Construction and Testing

Pearson correlation analysis was performed to verify the correlation between patient trust and oculomotor indicators. In this paper, PLS (partial least squares regression) method was used for modeling the relationship between patient perceived trust and oculomotor indexes. MATLAB was used to program the solution according to the PLS calculation steps. The results of the correlation analysis are shown in Table 4: (i) patients with higher health literacy showed significant positive correlations between perceived trust and gaze duration (PCC=0.751, sig=0.012) and gaze number (PCC=0.792, sig=0.006) in the global interpretation of diagnostic reports; (ii) patients with higher health literacy showed significant positive correlations between perceived trust in the local interpretation type of diagnostic reports were significantly and positively correlated with both gaze duration (PCC=0.893, sig=0.001) and gaze frequency (PCC=0.765, sig=0.016); (iii) patients with weaker health literacy showed a significant positive correlation between perceived trust and gaze duration (PCC=0.820, sig=0.046) and gaze frequency (PCC=0.829, sig=0.041) in the diagnostic report of the global explanation type; (iv) patients with weaker health literacy had a significant positive correlation between perceived trust and gaze duration (PCC=0.773, sig=0.042) and gaze frequency (PCC=0.830, sig=0.021) in the local interpretation type.

The results of the above correlation analysis show that there is a correlation between patients' perceived trust and eye movement indexes. The relationship model between patients'

perceived trust and gaze level indexes can be further constructed. The patient gaze duration and gaze number were selected as independent variables, and the mean value of patient perception trust score was taken as the dependent variable. The PLS method was used to construct the relationship model between patient perception trust and oculomotor indexes with different AI interpretation types through MATLAB programming to achieve the prediction of patient perception trust for diagnostic reports, and the relationship model equations were obtained according to the results of MATLAB runs as follows:

$$Y_1 = 0.0990 + 0.0280X_1 + 0.0160X_2 \quad (1)$$

$$Y_2 = 1.5014 + 0.0408X_1 + 0.0049X_2 \quad (2)$$

$$Y_3 = -2.4155 + 0.0260X_1 + 0.0158X_2 \quad (3)$$

$$Y_4 = -0.6302 + 0.0254X_1 + 0.0126X_2 \quad (4)$$

Where $Y1$ denotes the mean score of perceived trust rating of global explanation type diagnostic reports for patients with strong health literacy, and $Y2$ represents the mean score of perceived trust in the diagnostic report of the local interpretation type for patients with strong health literacy, and$Y3$ represents the mean score of perceived trust in the global explanatory type of diagnostic report for patients with weak health literacy, $Y4$ the mean score representing the perceived trust rating of patients with weak health literacy for the local interpretation type of diagnostic report;$X1$ is the duration of gaze, and$X2$ is the number of gazes.

Table 4. Differences in patient gaze under different types of AI interpretation

Health Literacy	Explanation Type	Attention span		Number of gazes	
		PCC	sig	PCC	sig
strong	global	0.751	0.012*	0.792	0.006**
	local	0.893	0.001**	0.765	0.016*
weak	global	0.820	0.045*	0.829	0.041*
	local	0.773	0.042*	0.830	0.021*

To test the accuracy of the constructed model, four subjects with strong expertise and eight subjects with weak expertise were reselected, and the diagnostic reports with both global and partial explanations were included.

Table 5. Paired samples t-test of model predicted and actual values

Health Literacy	Explanation Type	Actual value		Predicted value		t	df	p
		Average value	Standard deviation	Average value	Standard deviation			
strong	global	3.835	0.233	3.863	0.073	-2.248	1	0.845
	local	3.335	0.474	3.185	0.223	0.847	1	0.553
weak	global	2.500	0.240	2.691	0.201	-6.780	1	0.093
	local	4.165	0.234	4.091	0.290	1.872	1	0.312

As shown in Table 5, there was no significant difference between the predicted and actual values of the partial least squares regression model between patients' perceived trust and eye movement indexes, i.e., the significance levels between the predicted and actual values were all greater than 0.05, and this result verified the accuracy of the PLS model.

6. Discussion

6.1. Patient Trust Evaluation Discussion

There were significant differences in trust in diagnostic reports of different interpretation types among patients with different health literacy. Patients with weaker health literacy had higher levels of trust in diagnostic reports that provided a local interpretation type, while patients with stronger health literacy had higher levels of trust in diagnostic reports that provided a global interpretation type. The possible reason for this is that in healthcare contexts with high risk, patients use to be more concerned about the accuracy and reliability of the AI system and evaluate AI system explanations based on a priori knowledge when the AI system is interacting [48]. Thus, patients' trust in AI diagnostic systems is not only influenced by the type of explanation but also related to patients' health literacy.

6.2. Discussion of Eye Movement Data Results

The existing literature on patients' gaze behavior towards the interpretation interface of artificial intelligence systems in medical environments is scarce, and studies have shown that gaze duration and gaze number respond to the user's sensitivity to the processing of information content within the interest area, which is an important basis for measuring cognitive load. The analysis of gaze behavior data by eye-tracking technology can obtain high-value information about users' interest in content, ease of accessing information and behavioral models [49]. The longer the total gaze time, the more difficult it is for users to process information and the higher the cognitive load [47,50]. The more the number of gaze sessions, the greater the effort required for users to process the information and the higher the cognitive load [46]. Some studies have found that users' ability to access health information on social media has an impact on the trustworthiness of information assessed by individuals, and users with lower health literacy spend more time on information processing. In this paper, patients with strong health literacy were more sensitive to disease diagnosis information and had higher information processing ability, so the length of gaze and number of gazes were significantly smaller than those with weaker health literacy levels; patients with stronger health literacy had higher trust in global explanations, while patients with weak health literacy had more trust in partial explanations of information. From the correlation between patients' trust and gaze level, it is reasonable that patients with strong health literacy gaze at globally explained information more than locally explained information, and patients with weak health literacy gaze at locally explained information significantly more than globally explained information.

6.3. Discussion of the Relationship Model Between Patient Perceived Trust and Eye Movement Indicators

From the correlation test and PLS model correlation coefficients, it can be seen that patients' perceived trust in the information of diagnostic report interpretation type is positively correlated with gaze level, and the correlation coefficient of gaze duration is greater than that of gaze number, which can be interpreted as the influence of gaze duration on patients' perceived trust is greater than that of gaze number. In addition, comparing the model predicted values and patient perceived trust results by paired sample t-test reveals that the model predicted values do not differ significantly from the actual values, i.e., it gives support to the accuracy of the model constructed using the PLS method for the relationship between gaze level and patient perceived trust evaluation.

7. Conclusion

In this paper, an AI diagnostic experimental study is conducted to evaluate the effects of different AI explanation types and patients' health literacy levels on patients' trust, and a prediction model of perceived trust levels is constructed in combination with an eye-tracking device. The main findings are as follows: (i) Patients with different health literacy levels have different levels of trust in global interpretation and partial interpretation reports, and patients may pay more attention to the accuracy of information rather than understanding, and intelligent diagnostic services should focus on cultivating patients' health literacy levels and pushing disease-related knowledge to promote patients' trust levels. (ii) Patients with different health literacy differ in their ability to process information, which is expressed through the gaze level. Also, the type of AI diagnostic report interpretation affects patients' small gaze level, indicating that eye-movement indicators can reflect patients' intrinsic cognition and extrinsic cognitive load. (iii) By correlating perceived trust with gaze level, the relationship model of patients' perceived trust was constructed, which can accurately and effectively make objective prediction of patients' trust. The above findings can enrich the study of eye movement metrics in medical high-risk areas and the construction of patient trust models, and provide developers and designers of AI diagnostic systems with a basis for pre-design development. However, there are some shortcomings in this paper. First, this paper uses university students as subjects, which is limited at the age level and can be explored for different age groups in the future. Second, the relationship model was constructed by choosing two commonly used eye movement indicators, but whether other physiological indicators have correlation needs to be further explored in future experiments.

References

[1] He J, Sally L, Xu J, et al. The practical implementation of artifcial intelligence technologies in medicine. Nat. Med, 2019, 25(1): 30-36.

[2] Erickson B J. Machine learning for medical imaging. Radiographics, 2017, 37(2): 505-515.

[3] Sunarti S, Fadzlul R, Ferry N, et al. Artificial intelligence in healthcare: opportunities and risk for future. Gaceta Sanitaria, 2021, 35(S1): S67-S70.

[4] Rasheed J, Jamil A, Hameed A, et al. A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. Chaos, Solitons and Fractals: the interdisciplinary journal of Nonlinear Science, and Nonequilibrium and Complex Phenomena, 2020, 79(20): 307323.

[5] Xu H H. On credibility of medical artificial intelligence technologies based on physician-patient trust. Journal of Engineering Studies, 2020, 12(3): 252-259.

[6] Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access, 2018, 6: 52138-52160.

[7] Caruana R, Lou J, Gehrke P, et al. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, 1721-1730.

[8] Lu L, Cai R, Gursoy D. Developing and validating a aervice robot integration willingness scale. International Journal of Hospitality Management, 2019, 80: 36-51.

[9] Cook R D. Detection of influential observation in linear regression. Technometrics, 1977, 19: 15-18.

[10] Yokoi R, Eguchi Y, Fujita T, et al. Artificial intelligence is trusted less than a doctor in medical treatment decisions: influence of perceived care and value similarity. International Journal of Human-Computer Interaction, 2020, 11: 1-10.

[11] Bussone A, Stumpf S, Sullivan D O. The role of explanations on trust and reliance in clinical decision support systems. IEEE International Conference on Healthcare Informatics, 2015.

[12] Vorm E S, Andrew D M. Assessing the value of ttransparency in recommender systems: an end-yser perspective. ACM Conference on Recommender Systems, 2018.

[13] Jonathan L. Herlocker J, Joseph A, et al. Explaining collaborative filtering recommendations. Computer supported cooperative work, 2000.

[14] Wang J, Ye B. "Responsible" algorithmic transparency: The Trend of media ethics Construction in the era of Artificial Intelligence. Studies in Dialectics of Nature, 2020, 36(12): 66-72.

[15] Upol E, Tambwekar R, Larry C, et al. Automated rationale generation: a technique for explainable AI and its effects on human perceptions. CoRR, 2019, 1: 037290.

[16] Kong X W, Tang J Z, Wang Z M. A review of research on the interpretability of artificial intelligence decisions. Systems Engineering-Theory & Practice, 2021, 41(2): 524-536.

[17] Zhang Z, Xie Y, Xing F, et al. MDNet: a semantically and visually interpretable medical image diagnosis network. IEEE Computer Society. IEEE Computer Society, 2017: 3549-3557.

[18] Tang Z, Chuang K V, Decarli C, et al. Interpretable classification of alzheimer's disease pathologies with a convolutional neural network pipeline. Nature Communications, 2019, 10: 071010.

[19] Karim M R, Dohmen T, Cochez M, et al. Deep COVID explainer: explainable COVID-19 diagnosis from chest X-ray images. IEEE International Conference on Bioinformatics and Biomedicine, 2020.

[20] Alam L, Mueller S. Examining the effect of explanation on satisfaction and trust in AI diagnostic systems. BMC Medical Informatics and Decision, aking, 2021, 21(1): 178-184.

[21] Ehsan U, Tambwekar P, Chan L, et al. Automated rationale generation: a technique for explainable AI and its effects on human perceptions. IUI '19: Proceedings of the 24th International Conference on Intelligent User Interfaces, 2019.

[22] Larosa E, Danks D. Impacts on trust of Hhealthcare AI. The 2018 AAAI/ACM Conference, 2018.

[23] Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Medical Informatics and Decision Making, 2020, 20(1): 129-136.

[24] Asan O, Bayrak E, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. Journal of Medical Internet Research, 2020, 22(6): e15154.

[25] Aniek F, Jan A, Peter R. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. Journal of biomedical informatics, 2021, 113(0): 103655.

[26] Shin D. How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. Computers in Human Behavior, 2020, 109: 106344.

[27] Schoonderwoerd T, Jorritsma W, Neerincx M A, et al. Human-Centered XAI: developing design patterns for explanations of clinical decision support systems. International Journal of Human-Computer Studies, 2021, 154(19): 102684.

[28] Bashir C, Masooda J, Hoff D, et al. Trust in automation: integrating empirical evidence on factors that influence trust. Human Factors the Journal of the Human Factors Society, 2015, 57: 407-434.

[29] Li S, Sun M, Xu N, et al. Development of a model to measure consumer emotions associated with web browsing. Industrial Engineering and Management, 2013, 1: 106-110.

[30] Liu C, Guo F, Liu W L, et al. Evaluation and prediction of the elderly's perceived usability of shopping website homepage interface. Industrial Engineering and Management, 2018, 23(6): 101-107.

[31] Lu Y, Sarter N. Eye tracking: a process-oriented method for inferring trust in automation as a function of priming and system reliability. IEEE Transactions on Human-Machine Systems, 2019, PP (99): 1-9.

[32] Wei S, Ying L, et al. The influence of physician information on patients' choice of physician in mhealth services Uusing China's chunyu doctor App: eye-tracking and questionnaire study. JMIR mHealth and uHealth, 2019, 7(10): e15544.

[33] Petty R E, Cacioppo J T. The elaboration likelihood model of persuasion. Advances in Experimental Social Psychology, 1986, 19: 123-205.

[34] Trivedi N, Lowry M, Gaysynsky A, et al. Factors associated with cancer message believability: a mixed methods study on simulated facebook posts. Journal of Cancer Education, 2021.

[35] Wan X Y, Wan J. An empirical study on users' information trust Judgment in online trivia community: evidence from eye tracking experiment. Library and Information Service, 2021, 65(23): 11.

[36] John B, Sweller H. Cognitive Load During Problem Solving: Effects on Learning[J]. Cognitive Science, 1988.

[37] [59]-37 Sweller, John. Cognitive load theory, learning difficulty, and instructional design. Learning & Instruction, 1994, 4(4): 295-312.

[38] Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. Educational Psychologist, 2003, 38(1): 1-4.

[39] Dong S Y, Kim B K, Lee S Y. EEG-based classification of implicit intention during self-relevant sentence reading. IEEE Transactions on Cybernetics, 2015: 1-8.

[40] Ayaz H, Shewokis P A, Bunce S, et al. Optical brain monitoring for operator training and mental workload assessment. Neuroimage, 2012, 59(1): 36-47.

[41] Müller J A, Wendt D, Kollmeier B, et al. Comparing eye tracking with electrooculography for measuring individual sentence comprehension duration. PLoS ONE, 2016, 11(10): e0164627.

[42] Orun O, Akbulut Y. Effect of multitasking, physical environment and electroence phalography use on cognitive load and retention. Computers in Human Behavior, 2019, 92(MAR): 216-229.

[43] Holden R J. Cognitive performance-altering effects of electronic medical records: an application of the human factors paradigm for patient safety. Cognition Technology & Work, 2011, 13(1): 11-29.

[44] Khairat S, Coleman C, Ottmar P, et al. Association of electronic health record use with physician fatigue and efficiency. JAMA Network Open, 2020, 3(6): e207385.

[45] Wu X D, Zhang X Y, Yue D M, et al. Location and color effect on interface display under gender difference. Industrial Engineering and Management, 2022, 27(6): 141-148.

[46] Yan G L, Xiong J P, Cang C L, et al. Review of eye-movement measures in reading research. Advances in Psychological Science, 2013, 21(4): 17-21.

[47] Busjahn T, Bednarik R, Schulte C. What influences dwell time during source code reading? Analysis of element type and frequency as factors. Studies in Foundations & Combinatorics, 2014.

[48] Samek W, Binder A, Montavon G, et al. Evaluating the visualization of what a deep neural network has learned. IEEE Transactions on Neural Networks and Learning Systems, 2016, 28(11): 2660-2673.

[49] Wang L, Guo M X. Is broeing theory-oriented or biology-driven?--an empirical research based on the eye-tracking experiment. Journal of the China Society for Scientific and Technical Information, 2015, 34(12): 12-18.

[50] Rodeghero P, Mcmillan C, Mcburney P W, et al. Improving automated source code summarization via an eye-tracking study of programmers. Computing Reviews, 2015, 56(4): 390-401.