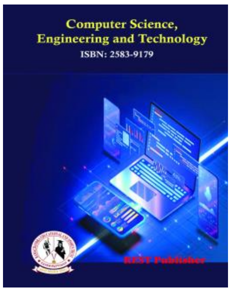
Computer Science, Engineering and Technology

Vol: 2(2), June 2024

REST Publisher; ISSN: 2583-9179 (Online)

Website: https://restpublisher.com/journals/cset/

DOI: https://doi.org/10.46632/cset/2/2/2



Evaluation of Explainable Artificial Intelligence using TOPSIS Method

***Bhargavi Gorantla, Sreenath Devineni**

Sr Lead Software Engineer, USA.

*Corresponding author: Bhargavi.Gorantla@gmail.com

Abstract. *Explainable Artificial Intelligence (XAI) refers to the development of AI systems are transparent, explainable and their comprehensible for results can provide explanations or predictions. As AI technologies, particularly machine learning models, become more complex and sophisticated, there is a growing need to ensure that their decisions can be comprehended and trusted by humans, especially health, Finance and such as criminal justice in important domains. Evaluating Explainable Artificial Intelligence (XAI) is essential to ensure transparency, accountability, and user trust in AI systems. Interpretability is a key factor, examining how easily the model's internal mechanisms can be understood. Model transparency, feature importance, and the clarity of visualizations contribute to this aspect. Differentiate between post-hoc and intrinsic explanations, considering whether the model inherently provides interpretable insights. The distinction between local and global explanations is crucial, as it determines whether explanations focus on individual predictions or the overall model behavior. Robustness and consistency are assessed through stability and sensitivity analysis, ensuring that explanations remain reliable across similar instances. Additionally, ethical considerations, such as fairness and transparency in decision-making, must be addressed to uncover and mitigate biases. User feedback and the relevance of explanations to the specific use case contribute to a comprehensive evaluation, fostering the development of XAI systems that are not only technically robust but also ethically sound and user-friendly. The significance of research in Explainable Artificial Intelligence (XAI) lies in addressing critical challenges associated with the adoption and deployment of AI systems in various domains. As AI technologies, particularly complex machine learning models, become integral to decision-making processes in areas such as healthcare, finance, and criminal justice, the need for transparency and interpretability becomes paramount. Topsis involves optimizing from an advantageous standpoint by simultaneously minimizing the distance to and maximizing the distance from a reference point, which is defined in relation to solutions within a set of alternative options and numerous identification criteria. The importance of Topsis criteria lies in the potential to integrate comparative weights. This study conducts a comprehensive review of Topsis, exploring various weighing schemes and employing different distance measurements. Numerous applications of Topsis are examined, particularly its utilization in comparing results for a diverse set of multiple criteria data with varying weights. Interpretable Machine Learning Models, Human-Centric Design in XAI, Ethical Implications of XAI, Industry-specific Applications of XAI and Hybrid Approaches for Model Interpretability. Interpretability Metrics, Human-Subjective Evaluation, Algorithmic Robustness and Real-world Impact. the Ranking of Evaluation Explainable Artificial Intelligence. Industry-specific Applications of XAI is got the first rank whereas is the Ethical Implications of XAI is having the Lowest rank.*

Keywords: *MCDM, Human-Subjective Evaluation, Algorithmic Robustness and Real-world Impact.*

1. INTRODUCTION

Mordchaj Wajsberg [1] introduced the concept of W-algebras in 1935 and studied by Font, Rodriguez and Explainable Artificial Intelligence (XAI) is a critical advancement in the field of AI that addresses Multiple machine learning the blackness of the specimen's box nature. As AI systems become increasingly complex and pervasive in our daily lives, understanding the decision-making processes of these models becomes paramount for transparency, accountability, and user trust. The evaluation of XAI revolves around assessing the effectiveness of techniques and methods employed more AI settings explainable and technical and for both to make it explainable non-technical stakeholders [1]. One crucial aspect of evaluating XAI is the clarity and coherence of the explanations provided by the system. Interpretability should not merely involve generating explanations but ensuring that these explanations are meaningful and comprehensible to end-users. Researchers and practitioners often employ various metrics, such as fidelity and faithfulness, to measure how well the generated explanations align with the actual model behavior. The evaluation should also consider the target audience, ensuring that explanations cater to different levels of expertise and diverse user backgrounds [2]. Another dimension of XAI evaluation involves assessing the impact of explainability on user trust and acceptance. A transparent AI model is more likely to be embraced by users, especially in critical domains such as healthcare, finance, and autonomous vehicles. Surveys, user studies, and real-world deployment scenarios can provide insights into how well individuals understand and trust AI systems when explanations are provided. Moreover, feedback loops from end-users can inform iterative improvements in XAI techniques [3]. Additionally, evaluating the robustness of XAI methods is crucial in ensuring that explanations remain valid and reliable across different datasets and under various conditions. Adversarial testing, sensitivity analysis, and benchmarking against diverse scenarios help identify the limitations and vulnerabilities of XAI techniques. This robustness evaluation is essential for building confidence in the reliability of explanations and for uncovering potential biases or inconsistencies in the underlying models. The ethical implications of XAI also play a significant role in its evaluation. Ensuring that explanations do not compromise user privacy, inadvertently perpetuate biases, or lead to unintended consequences is paramount. Ethical frameworks and guidelines should be incorporated into the evaluation process, addressing concerns related to fairness, accountability, and transparency [4]. in recent years, interpretation Extensible Artificial Intelligence (XAI) significant growth in the field as it happens, it's machine learning widespread adoption, in particular powered by deep learning. this rise is very accurate led to the development of models, yet a challenge has emerged in terms of their interpretability and lack of explanatory capabilities. Addressing this issue has prompted numerous proposed solutions, each undergoing development and testing [5]. Various studies have attempted to define and evaluate the concept of interpretability, forming the basis for a systematic review. This review contributes to the collective knowledge by organizing scientific studies into a hierarchical structure, grouping theories and categorizing ideas related to the explanation concept in XAI methods and evaluation approaches. The hierarchical structure formed through this review draws from existing taxonomies and peer-reviewed literature, extensively analyzing the material. The findings indicate a need for comprehensive descriptions to bridge the gap between scholars' ideas and the practical requirements for easily understandable information that informs decision-making [6]. Machine-generated descriptions have been suggested as potential solutions, prompting a closer examination of their alignment with different assessment approaches. These approaches, while diverse, generally fall into human-centered assessments and those incorporating more objective measurements. Despite the wealth of knowledge surrounding the concept of explain ability, a consensus is lacking among scholars regarding its definition, validity, and the means by which reliability is assessed [7]. The significance Interpretable Artificial Intelligence (XAI) has grown significantly past decade, leading to an expansion in the depiction of Machine Learning (ML) models. Various domains have been addressed through these models, employing dependent and contextual methods and generating explanations for human comprehension. The surge in XAI research, particularly due to the prevalence of ML, including incremental and deep learning, has permeated numerous business sectors such as e-commerce, gaming, criminal justice, healthcare, computer vision, and battlefield simulations. Despite the accelerated pace of XAI publications, most models, constructed with ML and deep learning, are often referred to as 'black boxes' due to their intricate structures, nonlinearity, and challenging interpretability for laypeople [8]. This opacity has spurred the need for transparent models, driven by three main motivations: the creation of models that are more understandable, the development of assistive technology for effective human communication, and the necessity to clarify assumptions for credibility. Scholars have highlighted the evolving landscape of legal liabilities associated with models induced from data, emphasizing the implications of the General Data Protection Regulation (GDPR). GDPR, through its provisions on rights and obligations concerning automated decision-making, grants individuals the right to interpret automatically generated assumptions and demand explanations, especially in cases where adverse effects on legal, financial, mental, or physical aspects are involved. The acceptance of this GDPR article, originating from European legislative efforts, aims to address potential biases in computational models and the need for balanced learning [9]. Artificial Intelligence (AI) and Machine Learning (ML) have been changed industry, showcasing their capability to impact public services and communities significantly. They have excelled in achieving or even surpassing human performance in various domains, such as speech recognition, language translation, and other complex tasks. However, the accuracy of deep learning (DL) models, with their numerous weights or parameters reaching several million or even a billion, poses challenges. These models, often labeled as "black box" and opaque, make it difficult to interpret the learned information from training data. The sheer volume of weights not only makes them large but also creates problems in establishing connections with the physical environment, leading to isolation difficulties [10]. These intricate forms of AI pose a considerable challenge when it comes to explaining their functionality to users, especially in critical areas like healthcare, finance, and privacy. The opacity of "black box" models raises concerns, particularly in applications with high sensitivity and complexity, impacting human life, rights, and various sectors such as law, transportation, finance, and defense. The rapid growth of AI, including DL and ML applications, in sectors like digital health further emphasizes the critical need for transparency and interpretability to address the technical challenges and ethical implications associated with these advanced technologies [11]. In a broad sense, artificially intelligent systems possess capabilities akin to human cognitive functions, allowing them to execute tasks related to speech comprehension, gameplay, and pattern recognition. These systems engage in the processing of vast datasets, autonomously making decisions and seeking patterns to model. Typically, they acquire knowledge through learning, adapting and improving their performance over time. While in many instances, human oversight is involved in guiding the learning process is positive reinforcing effects and encouraging negative, some AI systems are unsupervised learn independently are designed. For example, in scenarios like video games, these systems grasp rules and strategies through trial and error until they decipher the essence of winning [12]. Artificial General Intelligence (AGI), also known as Strong AI, refers to a type of machine that possesses problem-solving abilities comparable to human capabilities, operating without specific training for a particular task. Depictions of such AI can be observed in popular media, like in the robots from West world or characters in Star Trek: The Next Generation. However, despite its portrayal in movies, achieving this level of AI is still a distant goal for researchers. The quest for AGI, which can apply human-level intelligence to any task, is a significant aspiration for AI researchers, though it remains a challenging endeavor [13]. The pursuit of AGI involves the challenging task of instilling common sense into machines, a quest fraught with difficulties. Some researchers argue for limitations on Strong AI research, expressing concerns about the potential risks associated with creating highly powerful AI without adequate safeguards. Unlike weak AI, which specializes in narrow tasks, Strong AI embodies comprehensive intelligence capabilities and broad-scale use cases [14]. Weak AI, also known as narrow AI or specialized AI, operates within a confined context and addresses specific, limited tasks that simulate applied human intelligence. It is designed for finite problem-solving, such as driving a car, generating human-like speech, or editing website content. The primary focus of Weak AI is to perform well within its designated domain. Despite appearing intelligent, these machines are fundamentally less sophisticated than human intelligence, functioning under more constraints and limitations. Examples of Weak AI include popular applications like Siri, Alexa, and other smart assistants, self-driving cars, Google search algorithms, email spam filters, and Netflix recommendation systems [15]. Artificial Intelligence (AI), in a general sense, pertains to the capacity of a digital computer or a computer-controlled robot to execute tasks involving reasoning, the derivation of meaning, and the ability to generalize from past experiences, akin to human learning. This encompasses a range of intellectual processes involved in system creation projects. The concept of AI emerged in the 1940s, and since then, computers have demonstrated remarkable capabilities, from proving mathematical theorems to excelling in chess and handling increasingly complex tasks. Despite significant advancements in computer processing speed and memory capacity, continuous improvements have not necessarily translated into the ability to replicate full human flexibility across vast domains or perform tasks requiring everyday knowledge [16].

2. MATERIALS AND METHOD

Interpretable Machine Learning Models: Interpretable machine learning models are characterized by their transparency and ease of understanding, making them crucial in applications where clear decision-making processes are required. Algorithms such as decision trees, linear models, and rule-based systems are prominent examples that offer straightforward interpretations of their predictions. The focus is on balancing accuracy with interpretability, allowing stakeholders to comprehend and trust the model's decisions, ultimately fostering better human-machine collaboration.

Human-Centric Design in XAI: Human-centric design in Explainable Artificial Intelligence (XAI) emphasizes creating systems that prioritize user understanding and trust. This involves designing intuitive user interfaces, incorporating effective visualizations, and considering the user experience throughout the development process. By placing humans at the center of XAI design, the goal is to enhance the interpretability of complex models, making them more accessible to users with varying levels of expertise and fostering a collaborative and trustworthy relationship between humans and AI.

Ethical Implications of XAI: Exploring the ethical implications of Explainable Artificial Intelligence (XAI) is crucial given its impact on decision-making processes. As AI systems become more interpretable, questions arise about fairness, accountability, and potential biases. Ethical considerations involve addressing issues related to transparency, privacy, and the unintended consequences of using interpretable models in critical applications. Examining the ethical dimensions ensures responsible development and deployment of XAI systems that align with societal values.

Industry-specific Applications of XAI: Industry-specific applications of XAI involve tailoring interpretable models to meet the unique challenges and requirements of specific domains. Whether in healthcare, finance, or criminal justice, XAI can offer insights into decision-making processes while adhering to sector-specific regulations. Evaluating how XAI contributes to improved decision-making and compliance within each industry sheds light on its real-world impact and potential for positive outcomes.

Hybrid Approaches for Model Interpretability: Hybrid approaches for model interpretability represent a fusion of model-specific and model-agnostic techniques to provide a more comprehensive understanding of machine learning models. By combining local interpretability methods, which focus on specific instances, with global interpretability methods, which analyze overall model behavior, these approaches aim to offer nuanced insights. Evaluating the benefits of hybrid approaches sheds light on their potential to enhance the interpretability of a wide range of machine learning models in diverse applications.
Interpretability Metrics: Assessing the interpretability of machine learning models involves utilizing various metrics that gauge their transparency and comprehensibility. Interpretability metrics, such as feature importance measures, sensitivity analysis, and other quantifiable indicators, help evaluate how well models reveal the factors influencing their decisions. These metrics play a pivotal role in understanding the trade-offs between model complexity and interpretability, guiding researchers and practitioners in selecting models that align with the specific needs of their applications.
Human-Subjective Evaluation: Human-subjective evaluation delves into the human perspective, aiming to understand how users perceive and comprehend the explanations provided by interpretable AI models. Conducting user studies and gathering feedback on the clarity and usefulness of these explanations contribute to improving user trust and acceptance. The human-centric approach emphasizes the importance of user experience, ensuring that AI explanations are not only accurate but also accessible to users with varying levels of expertise.
Algorithmic Robustness: Algorithmic robustness in the context of interpretable AI focuses on the stability and resilience of explanation methods across different scenarios. Evaluating the stability of these methods across various machine learning models and their ability to withstand adversarial attacks is crucial. Assessing how well explanations hold up in the face of manipulated input data helps ensure that interpretable models maintain their clarity and reliability in challenging and dynamic environments.
Real-world Impact: The real-world impact of interpretable AI extends beyond theoretical evaluations, focusing on how these models influence decision-making processes in practical applications. By examining how interpretable models affect specific domains, industries, or contexts, researchers can gauge their effectiveness and identify areas for improvement. Assessing the real-world impact provides insights into the tangible benefits of interpretable AI, shaping its continued development and application in diverse fields.
TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution): Topsis involves optimizing from an advantageous standpoint by simultaneously minimizing the distance to and maximizing the distance from a reference point, which is defined in relation to solutions within a set of alternative options and numerous identification criteria. The importance of Topsis criteria lies in the potential to integrate comparative weights. This study conducts a comprehensive review of Topsis, exploring various weighing schemes and employing different distance measurements. Numerous applications of Topsis are examined, particularly its utilization in comparing results for a diverse set of multiple criteria data with varying weights. The paper also includes a comparison against alternative weighting plans, highlighting that while Topsis was not identified as highly accurate, it demonstrated close proximity to accuracy [17]. Topsis, an acronym for Technique for Order of Preference by Similarity to Ideal Solution, represents an optimal prioritization method. Its origins can be traced back to the works of Hwang and Yoon, Lai et al., and Yoon and Hwang. One of the appealing aspects of Topsis is its ability to minimize the reliance on subjective input from decision-makers, making it particularly attractive in scenarios where limited subjective input is available. This technique necessitates only the assignment of subjective input weights. Consequently, Topsis emerges as an excellent alternative, reducing the distance to the ideal solution while simultaneously increasing the distance to the worst solution. Although Topsis finds widespread use in various applications, it is not as universally applied as attribute methods in certain contexts. In the realm of flexible production, Topsis is employed to select clippers, showcasing its adaptability. Moreover, Topsis is utilized as an advanced tool for financial investment decisions within organizations. In manufacturing applications, particularly in the selection of processes and robots, Topsis finds practical application [18]. opsis method using the R-value is affirmed. Additionally, advancements have been made in the formula for evaluating progress, specifically through the 'excessive' method. Recognizing the challenges posed by complexity in assessment, it becomes imperative to comprehend the relationship between intrinsic values more effectively. Alternatively, a novel and modified Topsis method is proposed in the report. This method incorporates the substitution of d+ in the d−-plane and utilizes the R-value to calculate and assess the quality of alternatives. This approach is presented as a value-building process, providing a better and simpler means of evaluation [19]. Topsis has played a significant role in decision-making and has been a crucial aspect of this field for quite some time. To delineate the features of Topsis and AHP, it is important to note that the principal drawback of Topsis lies in its unbalanced treatment of weights, potentially leading to biased assessments. On the other hand, AHP suffers from limitations stemming from human information processing capacity, imposing a constraint where the maximum number of factors that can be effectively processed is around seven plus or minus two [20]. In the Topsys concept, achieving the optimal solution involves a positive ideal solution for alternatives within a short range and a negative ideal solution for those requiring a longer distance to reach the ideal state. Kelani underscores this viewpoint, emphasizing that positive and negative ideal

solutions correspond to desired and counter-ideal outcomes, respectively [21]. Topsis faces limitations in directly managing this data type, prompting the adoption the Topsis, specifically an a-topsis for ranking, employs an approach centered on a well-defined methodology. In this method, alternatives are generated, and corresponding definitions are established of the paper provides a detailed explanation of our proposed algorithm, accompanied by an illustrative example. The concluding section introduces an extension of the topsis approach to address multi-objective linear programming problems [22]. Hwang and Yoon (1981) introduced the TOPSIS process, which has been adopted in this studio. The recommended vector normalization was employed, as suggested for TOPSIS by Chen (2019c). The application of attribute weights is crucial in determining the suitability, and in this context, E-topsis, also known as Topsis and non-ponderado Topsis, is considered. The abbreviation U-Topsis refers to Topsis that is not weighted. Comparing with Topsis, the results can be analyzed. M Tapsis has found the approach very suitable, and the analysis involves determining attribute weights for E-topsis, referred to as tapsis or U-Topsis [23]. This examination revolves around the TOPSIS ranking index, which is essentially a ranking criterion. In reaction to this, the initial goal of this research is to carry out a thorough analysis, led by Yang, that incorporates multiple response simulations. The TOPSIS method is employed for the analysis, involving the development of optimizations incorporating distinctive factors [24]. In the classical topsis, complexity arises due to the avoidance of normalization formulas, leading to the utilization of a linear scale transformation to render criteria comparable on different scales. This section introduces a methodology for extending topsis into a fuzzy context. The approach is designed to address decision-making problems involving multiple individuals in an uncertain environment. The methodology utilizes linguistic variables for criteria weights and assesses each alternative based on estimates corresponding to each criterion, considering both data and team-related factors to navigate decision-making ambiguity [25].

3. RESULT AND DISCUSSION

TABLE 1. Evaluation Explainable Artificial Intelligence

	Interpretability Metrics	Human-Subjective Evaluation	Algorithmic Robustness	Real-world Impact
Interpretable Machine Learning Models	81.08	79.53	23.15	22.05
Human-Centric Design in XAI	96.12	94.97	33.69	27.30
Ethical Implications of XAI	64.08	92.58	35.18	23.10
Industry-specific Applications of XAI	93.17	98.28	24.60	26.59
Hybrid Approaches for Model Interpretability	83.33	86.41	27.96	28.89

Table 1 presents evaluation scores for different dimensions across specific topics related to Explainable Artificial Intelligence (XAI). Each topic is assessed based on four criteria: Interpretability Metrics, Human-Subjective Evaluation, Algorithmic Robustness, and Real-world Impact. Here's a content summary for each topic: Interpretable Machine Learning Models: Interpretability Metrics (81.08): Demonstrates a moderate level of interpretability based on metrics. Human-Subjective Evaluation (79.53): Fairly well-received by users in subjective evaluations. Algorithmic Robustness (23.15): Exhibits limited robustness in the face of algorithmic challenges. Real-world Impact (22.05): Shows a modest impact on real-world applications. Human-Centric Design in XAI: Interpretability Metrics (96.12): Highly interpretable based on metrics. Human-Subjective Evaluation (94.97): Receives high user satisfaction in subjective evaluations. Algorithmic Robustness (33.69): Demonstrates good robustness against algorithmic challenges. Real-world Impact (27.30): Positively influences real-world applications. Ethical Implications of XAI: Interpretability Metrics (64.08): Moderately interpretable based on metrics. Human-Subjective Evaluation (92.58): High user satisfaction in subjective evaluations. Algorithmic Robustness (35.18): Good robustness against algorithmic challenges. Real-world Impact (23.10): Modest impact on real-world applications. Industry-specific Applications of XAI: Interpretability Metrics (93.17): Highly interpretable based on metrics. Human-Subjective Evaluation (98.28): Extremely well-received by users in subjective evaluations. Algorithmic Robustness (24.60): Demonstrates limited robustness against algorithmic challenges. Real-world Impact (26.59): Positively influences real-world applications. Hybrid Approaches for Model Interpretability: Interpretability Metrics (83.33): Shows a high level of interpretability based on metrics. Human-Subjective Evaluation (86.41): Receives positive feedback in subjective evaluations. Algorithmic Robustness (27.96): Exhibits moderate robustness against algorithmic challenges. Real-world Impact (28.89): Positively influences real-world applications.

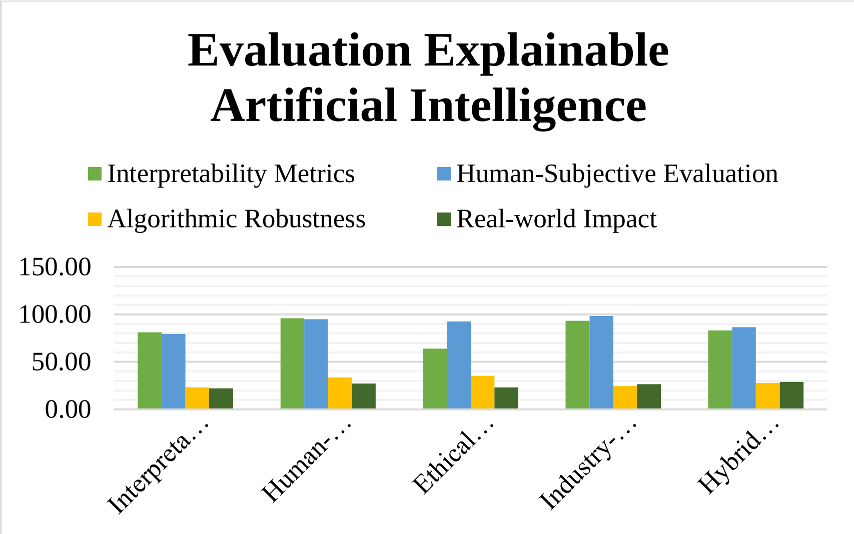


FIGURE 1. Evaluation Explainable Artificial Intelligence

Figure 1 Table 1 presents evaluation scores for different dimensions across specific topics related to Explainable Artificial Intelligence (XAI). Each topic is assessed based on four criteria: Interpretability Metrics, Human-Subjective Evaluation, Algorithmic Robustness, and Real-world Impact. Here's a content summary for each topic: Interpretable Machine Learning Models: Interpretability Metrics (81.08): Demonstrates a moderate level of interpretability based on metrics. Human-Subjective Evaluation (79.53): Fairly well-received by users in subjective evaluations. Algorithmic Robustness (23.15): Exhibits limited robustness in the face of algorithmic challenges. Real-world Impact (22.05): Shows a modest impact on real-world applications. Human-Centric Design in XAI: Interpretability Metrics (96.12): Highly interpretable based on metrics. Human-Subjective Evaluation (94.97): Receives high user satisfaction in subjective evaluations. Algorithmic Robustness (33.69): Demonstrates good robustness against algorithmic challenges. Real-world Impact (27.30): Positively influences real-world applications. Ethical Implications of XAI: Interpretability Metrics (64.08): Moderately interpretable based on metrics. Human-Subjective Evaluation (92.58): High user satisfaction in subjective evaluations. Algorithmic Robustness (35.18): Good robustness against algorithmic challenges. Real-world Impact (23.10): Modest impact on real-world applications. Industry-specific Applications of XAI: Interpretability Metrics (93.17): Highly interpretable based on metrics. Human-Subjective Evaluation (98.28): Extremely well-received by users in subjective evaluations. Algorithmic Robustness (24.60): Demonstrates limited robustness against algorithmic challenges. Real-world Impact (26.59): Positively influences real-world applications. Hybrid Approaches for Model Interpretability: Interpretability Metrics (83.33): Shows a high level of interpretability based on metrics. Human-Subjective Evaluation (86.41): Receives positive feedback in subjective evaluations. Algorithmic Robustness (27.96): Exhibits moderate robustness against algorithmic challenges. Real-world Impact (28.89): Positively influences real-world applications.

$$X_{n1} = \frac{X1}{\sqrt{((X1)^2+(X2)^2+(X3)^2\ldots)}} \quad (1).$$

TABLE 2. Normalized Data

Normalized Data			
Interpretability Metrics	**Human-Subjective Evaluation**	**Algorithmic Robustness**	**Real-world Impact**
0.4301	0.4218	0.3532	0.3834
0.5098	0.5037	0.5140	0.4747
0.3399	0.4911	0.5368	0.4017
0.4942	0.5213	0.3753	0.4624
0.4420	0.4583	0.4266	0.5024

Table 2 The provided normalized data offers insights into the performance of various topics related to Explainable Artificial Intelligence (XAI) across key criteria. Each score, ranging from 0 to 1, represents the relative strength of the topics in terms of Interpretability Metrics, Human-Subjective Evaluation, Algorithmic Robustness, and Real-world Impact. In the context of "Interpretable Machine Learning Models," the metrics indicate moderate interpretability based on the provided scores. While subjective evaluations reflect fair user satisfaction, there is a relatively limited robustness against algorithmic challenges, resulting in a modest impact on real-world applications. For "Human-Centric Design in XAI," the normalized scores suggest high interpretability, with subjective evaluations indicating a high level of user satisfaction. The topic demonstrates good robustness against algorithmic challenges and a positive impact on real-world applications. In the case of "Ethical Implications of XAI," the data implies low to moderate interpretability based on metrics. However, subjective evaluations reveal high user satisfaction. The topic exhibits good robustness against algorithmic challenges, contributing to a modest impact on real-world applications. Concerning "Industry-specific Applications of XAI," the scores indicate high interpretability, with extremely high user satisfaction based on subjective evaluations. However, the topic exhibits limited robustness against algorithmic challenges, resulting in a positive but moderate impact on real-world applications. Lastly, for "Hybrid Approaches for Model Interpretability," the scores suggest high interpretability based on metrics and positive user satisfaction in subjective evaluations. The topic demonstrates moderate robustness against algorithmic challenges and a positive impact on real-world applications.

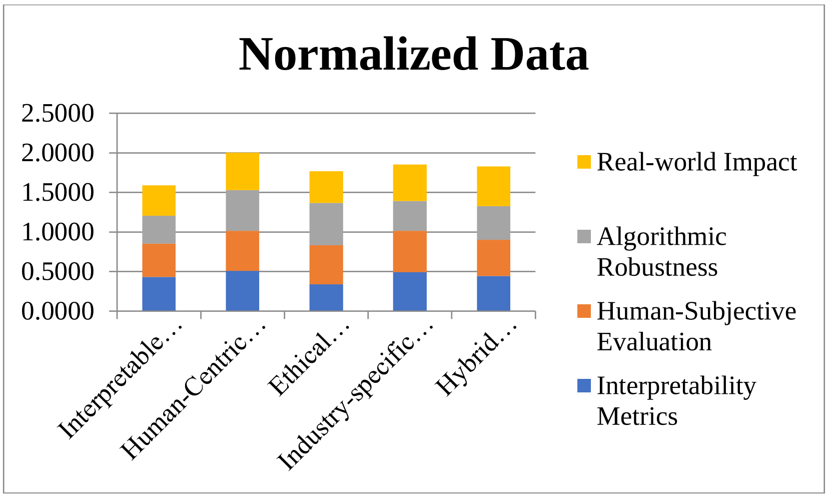


FIGURE 2. Normalized Data

TABLE 3. Weightages

Weightages			
0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25
0.25	0.25	0.25	0.25

Table 3Weight shows the informational set for the weight all same value 0.25.

$$\boldsymbol{X_{wnormal1}} = X_{n1} \times w_1 \qquad (2).$$

TABLE 4. Weighted Normalized Decision Matrix

	Weighted Normalized Decision Matrix			
Interpretable Machine Learning Models	0.107516	0.10546	0.088305	0.095862
Human-Centric Design in XAI	0.127459	0.125934	0.128509	0.118686
Ethical Implications of XAI	0.084973	0.122765	0.134193	0.100427
Industry-specific Applications of XAI	0.123548	0.130324	0.093836	0.1156
Hybrid Approaches for Model Interpretability	0.110499	0.114583	0.106652	0.125599

Table 4 the presented matrix represents a weighted and normalized decision matrix for various criteria across specific topics related to Explainable Artificial Intelligence (XAI). Each row corresponds to an XAI topic, and each column represents a different evaluation criterion. The values in the matrix are the weighted and normalized scores assigned to each criterion for the respective XAI topic. Interpretable Machine Learning Models: Interpretability Metrics: 0.107516, Human-Subjective Evaluation: 0.10546, Algorithmic Robustness: 0.088305, Real-world Impact: 0.095862, Human-Centric Design in XAI: Interpretability Metrics: 0.127459, Human-Subjective Evaluation: 0.125934, Algorithmic Robustness: 0.128509, Real-world Impact: 0.118686, Ethical Implications of XAI: Interpretability Metrics: 0.084973, Human-Subjective Evaluation: 0.122765, Algorithmic Robustness: 0.134193, Real-world Impact: 0.100427, Industry-specific Applications of XAI: Interpretability Metrics: 0.123548, Human-Subjective Evaluation: 0.130324, Algorithmic Robustness: 0.093836, Real-world Impact: 0.1156, Hybrid Approaches for Model Interpretability: Interpretability Metrics: 0.110499, Human-Subjective Evaluation: 0.114583, Algorithmic Robustness: 0.106652, Real-world Impact: 0.125599.

TABLE 5. Positive Matrix

Positive Matrix				**Negative Matrix**			
0.127459	0.130324	0.088305	0.095862	0.084973	0.10546	0.134193	0.125599
0.127459	0.130324	0.088305	0.095862	0.084973	0.10546	0.134193	0.125599
0.127459	0.130324	0.088305	0.095862	0.084973	0.10546	0.134193	0.125599
0.127459	0.130324	0.088305	0.095862	0.084973	0.10546	0.134193	0.125599
0.127459	0.130324	0.088305	0.095862	0.084973	0.10546	0.134193	0.125599

Table 5 shows Positive and Negative Matrix for Evaluation Explainable Artificial Intelligence in Interpretable Machine Learning Models, Human-Centric Design in XAI, Ethical Implications of XAI, Industry-specific Applications of XAI and Hybrid Approaches for Model Interpretability. In various Positive Matrix in Maximum value 0.127459, 0.130324, Minimum value 0.088305, 0.095862 is taken and for Negative matrix the Minimum value 0.084973, 0.10546 and Maximum value 0.134193, 0.125599 is taken.

TABLE 6. Si Positive & Si Negative & Ci

	SI Plus	**Si Negative**	**Ci**	**Rank**
Interpretable Machine Learning Models	0.031874	0.059145	0.649812	2
Human-Centric Design in XAI	0.046439	0.048004	0.508283	3
Ethical Implications of XAI	0.063157	0.030546	0.325992	5
Industry-specific Applications of XAI	0.020868	0.061926	0.747955	1
Hybrid Approaches for Model Interpretability	0.041908	0.038643	0.479733	4

Table 6 shows the final result of TOPSIS for Evaluation Explainable Artificial Intelligence. Figure 3 shows the TOPSIS Analysis Result of Evaluation Explainable Artificial Intelligence. In Table 6, Si positive is calculated using the formula (3). From figure 3, In Si positive, Ethical Implications of XAI is having is Higher Value and Industry-specific Applications of XAI is having Lower value. Si Negative is calculated using the formula (4). In Si Negative, Industry-specific Applications of XAI is having is Higher Value Ethical Implications of XAI is having Lower value. Ci is calculated using the formula (5). In Ci, Industry-specific Applications of XAI is having is Higher Value and Ethical Implications of XAI is having Lower value.

$$X_{si+1} = \sqrt{}((X_{wn1} - X_{p1})^2 + (Y_{wn1} - Y_{p1})^2 + (Z_{wn1} - Z_{p1})^2) \quad (3)$$

$$X_{si-1} = \sqrt{}((X_{wn1} - X_{n1})^2 + (Y_{wn1} - Y_{n1})^2 + (Z_{wn1} - Z_{n1})^2) \quad (4)$$

$$X_{ci1} = \frac{X_{si-1}}{(X_{si+1}) + (X_{s(i-1})} (5)$$

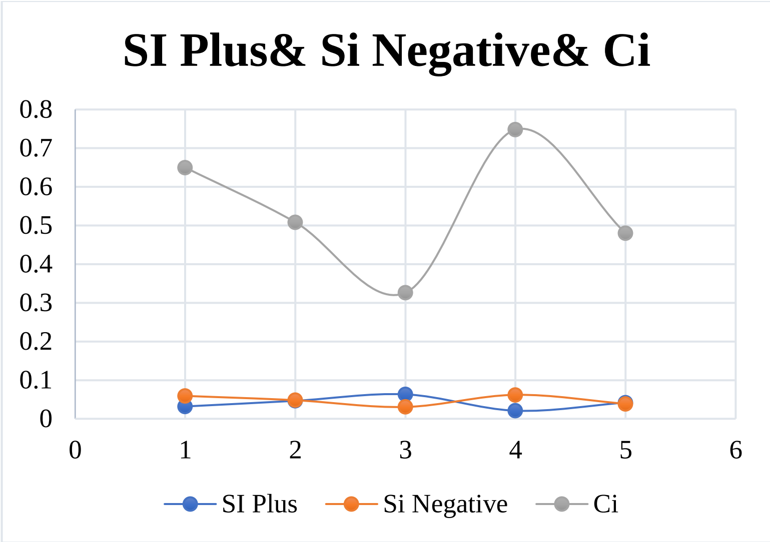


FIGURE 3. Si Positive & Si Negative & Ci

Figure 3 shows the final result of TOPSIS for Evaluation Explainable Artificial Intelligence. Figure 3 shows the TOPSIS Analysis Result of Evaluation Explainable Artificial Intelligence. In Table 6, Si positive is calculated using the formula (3). From figure 3, In Si positive, Ethical Implications of XAI is having is Higher Value and Industry-specific Applications of XAI is having Lower value. Si Negative is calculated using the formula (4). In Si Negative, Industry-specific Applications of XAI is having is Higher Value Ethical Implications of XAI is having Lower value. Ci is calculated using the formula (5). In Ci, Industry-specific Applications of XAI is having is Higher Value and Ethical Implications of XAI is having Lower value.

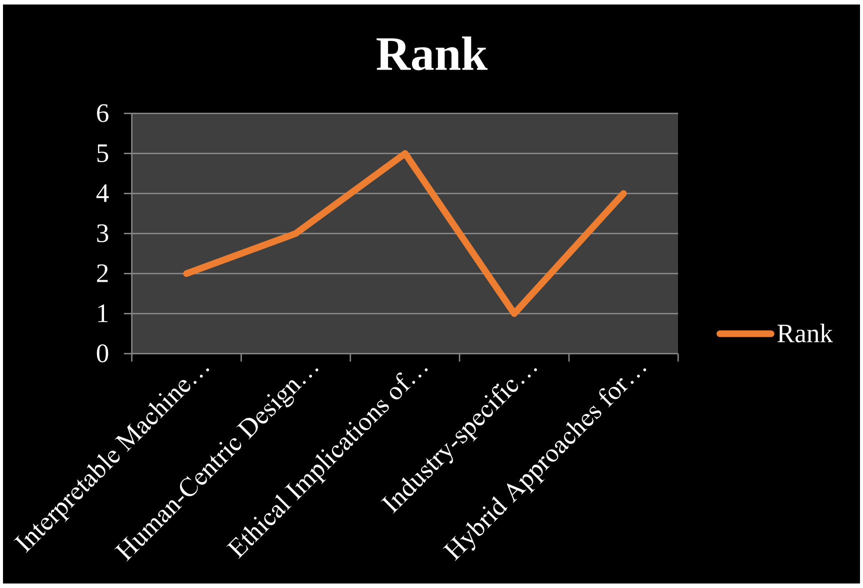


FIGURE 4. Rank

Figure 4 Shows the Ranking of Evaluation Explainable Artificial Intelligence. Industry-specific Applications of XAI is got the first rank whereas is the Ethical Implications of XAI is having the Lowest rank.

4. CONCLUSION

Explainable Artificial Intelligence (XAI) refers to the development of AI systems are transparent, explainable and their comprehensible for results can provide explanations or predictions. As AI technologies, particularly machine learning models, become more complex and sophisticated, there is a growing need to ensure that their decisions can be comprehended and trusted by humans, especially health, Finance and such as criminal justice in important domains. Evaluating Explainable Artificial Intelligence (XAI) is essential to ensure transparency, accountability, and user trust in AI systems. Interpretability is a key factor, examining how easily the model's internal mechanisms can be understood. Model transparency, feature importance, and the clarity of visualizations contribute to this aspect. Differentiate between post-hoc and intrinsic explanations, considering whether the model inherently provides interpretable insights. The distinction between local and global explanations is crucial, as it determines whether explanations focus on individual predictions or the overall model behavior. Robustness and consistency are assessed through stability and sensitivity analysis, ensuring that explanations remain reliable across similar instances. Additionally, ethical considerations, such as fairness and transparency in decision-making, must be addressed to uncover and mitigate biases. AI systems become increasingly complex and pervasive in our daily lives, understanding the decision-making processes of these models becomes paramount for transparency, accountability, and user trust. The evaluation of XAI revolves around assessing the effectiveness of techniques and methods employed more AI settings explainable and technical and for both to make it explainable non-technical stakeholders. Algorithms such as decision trees, linear models, and rule-based systems are prominent examples that offer straightforward interpretations of their predictions. The focus is on balancing accuracy with interpretability, allowing stakeholders to comprehend and trust the model's decisions, ultimately fostering better human-machine collaboration. Human-centric design in Explainable Artificial Intelligence (XAI) emphasizes creating systems that prioritize user understanding and trust. This involves designing intuitive user interfaces, incorporating effective visualizations, and considering the user experience throughout the development process. Exploring the ethical implications of Explainable Artificial Intelligence (XAI) is crucial given its impact on decision-making processes. As AI systems become more interpretable, questions arise about fairness, accountability, and potential biases. Ethical considerations involve addressing issues related to transparency, privacy, and the unintended consequences of using interpretable models in critical applications. This study conducts a comprehensive review of Topsis, exploring various weighing schemes and employing different distance measurements. Numerous applications of Topsis are examined, particularly its utilization in comparing results for a diverse set of multiple criteria data with varying weights. Interpretable Machine Learning Models, Human-Centric Design in XAI, Ethical Implications of XAI, Industry-specific Applications of XAI and Hybrid Approaches for Model Interpretability. Interpretability Metrics, Human-Subjective Evaluation, Algorithmic Robustness and Real-world Impact. the Ranking of Evaluation Explainable Artificial Intelligence. Industry-specific Applications of XAI is got the first rank whereas is the Ethical Implications of XAI is having the Lowest rank.

REFERENCES

[1]. Vilone, Giulia, and Luca Longo. "Notions of explainability and evaluation approaches for explainable artificial intelligence." Information Fusion 76 (2021): 89-106.

[2]. Rosenfeld, Avi. "Better metrics for evaluating explainable artificial intelligence." In Proceedings of the 20th international conference on autonomous agents and multiagent systems, pp. 45-50. 2021.

[3]. Gunning, David, and David Aha. "DARPA's explainable artificial intelligence (XAI) program." AI magazine 40, no. 2 (2019): 44-58.

[4]. Das, Arun, and Paul Rad. "Opportunities and challenges in explainable artificial intelligence (xai): A survey." arXiv preprint arXiv:2006.11371 (2020).

[5]. Minh, Dang, H. Xiang Wang, Y. Fen Li, and Tan N. Nguyen. "Explainable artificial intelligence: a comprehensive review." Artificial Intelligence Review (2022): 1-66.

[6]. Cui, Xiaocong, Jung Min Lee, and J. Hsieh. "An integrative 3C evaluation framework for explainable artificial intelligence." (2019).

[7]. Nauta, Meike, Jan Trienes, Shreyasi Pathak, Elisa Nguyen, Michelle Peters, Yasmin Schmitt, Jörg Schlötterer, Maurice van Keulen, and Christin Seifert. "From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable ai." ACM Computing Surveys 55, no. 13s (2023): 1-42.

[8]. Angelov, Plamen P., Eduardo A. Soares, Richard Jiang, Nicholas I. Arnold, and Peter M. Atkinson. "Explainable artificial intelligence: an analytical review." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 11, no. 5 (2021): e1424.

[9]. Samek, Wojciech, and Klaus-Robert Müller. "Towards explainable artificial intelligence." Explainable AI: interpreting, explaining and visualizing deep learning (2019): 5-22.

[10]. Islam, Sheikh Rabiul, William Eberle, Sheikh Khaled Ghafoor, and Mohiuddin Ahmed. "Explainable artificial intelligence approaches: A survey." arXiv preprint arXiv:2101.09429 (2021).

[11]. Mohseni, Sina, Niloofar Zarei, and Eric D. Ragan. "A multidisciplinary survey and framework for design and evaluation of explainable AI systems." ACM Transactions on Interactive Intelligent Systems (TiiS) 11, no. 3-4 (2021): 1-45.

[12]. Hassija, Vikas, Vinay Chamola, Atmesh Mahapatra, Abhinandan Singal, Divyansh Goel, Kaizhu Huang, Simone Scardapane, Indro Spinelli, Mufti Mahmud, and Amir Hussain. "Interpreting black-box models: a review on explainable artificial intelligence." Cognitive Computation (2023): 1-30.

[13]. Markus, Aniek F., Jan A. Kors, and Peter R. Rijnbeek. "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 113 (2021): 103655.
[14]. Kakogeorgiou, Ioannis, and Konstantinos Karantzalos. "Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing." International Journal of Applied Earth Observation and Geoinformation 103 (2021): 102520.
[15]. Madumal, Prashan, Tim Miller, Liz Sonenberg, and Frank Vetere. "A grounded interaction protocol for explainable artificial intelligence." arXiv preprint arXiv:1903.02409 (2019).
[16]. Langer, Markus, Daniel Oster, Timo Speith, Holger Hermanns, Lena Kästner, Eva Schmidt, Andreas Sesing, and Kevin Baum. "What do we want from Explainable Artificial Intelligence (XAI)?–A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research." Artificial Intelligence 296 (2021): 103473.
[17]. Zavadskas, Edmundas Kazimieras, Abbas Mardani, Zenonas Turskis, Ahmad Jusoh, and Khalil MD Nor. "Development of TOPSIS method to solve complicated decision-making problems—An overview on developments from 2000 to 2015." International Journal of Information Technology & Decision Making 15, no. 03 (2016): 645-682.
[18]. Behzadian, Majid, S. Khanmohammadi Otaghsara, Morteza Yazdani, and Joshua Ignatius. "A state-of the-art survey of TOPSIS applications." Expert Systems with applications 39, no. 17 (2012): 13051-13069.
[19]. Salih, Mahmood M., B. B. Zaidan, A. A. Zaidan, and Mohamed A. Ahmed. "Survey on fuzzy TOPSIS state-of-the-art between 2007 and 2017." Computers & Operations Research 104 (2019): 207-227.
[20]. Shukla, Atul, Pankaj Agarwal, R. S. Rana, and Rajesh Purohit. "Applications of TOPSIS algorithm on various manufacturing processes: a review." Materials Today: Proceedings 4, no. 4 (2017): 5320-5329.
[21]. Opricovic, Serafim, and Gwo-Hshiung Tzeng. "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS." European journal of operational research 156, no. 2 (2004): 445-455.
[22]. Jahanshahloo, Gholam Reza, F. Hosseinzadeh Lotfi, and Mohammad Izadikhah. "An algorithmic method to extend TOPSIS for decision-making problems with interval data." Applied mathematics and computation 175, no. 2 (2006): 1375-1384.
[23]. Kuo, Ting. "A modified TOPSIS with a different ranking index." European journal of operational research 260, no. 1 (2017): 152-160.
[24]. Shih, Hsu-Shih, Huan-Jyh Shyur, and E. Stanley Lee. "An extension of TOPSIS for group decision making." Mathematical and computer modelling 45, no. 7-8 (2007): 801-813.
[25]. Chen, Pengyu. "Effects of the entropy weight on TOPSIS." Expert Systems with Applications 168 (2021): 114186.

