Explainable Artificial Intelligence in Decision Support Systems: A Structured Literature Review

Alya Kartikasari[1*], Reyhan Firmansyah[2]
[12]Department of System of Information, Indo Global Mandiri University, Palembang, Indonesia
*Correspondence: alya.kartikasari14@gmail.com

ARTICLE HISTORY
Received: *[25/01/2026]* **Revised:** *[16/03/2026]* **Accepted:** *[20/05/2026]* **Published:** *[30/06/2026]*
Vol/No/Pages: *[Vol 1 (No.2), 53-62]*

Abstract

The use of artificial intelligence (AI) in decision support systems offers an opportunity to improve the speed, consistency, and accuracy of analysis across organizational contexts. Yet the use of complex AI models also raises concerns about transparency, interpretability, accountability, and user trust in the recommendations produced. Explainable Artificial Intelligence (XAI) has emerged as an approach to explaining the process and output of AI models so that they can be understood and evaluated by people. This article maps the concepts, methods, application contexts, evaluation approaches, and research challenges of XAI in decision support systems. The study was developed through a structured review of the literature with a scoping orientation, using a provided set of references on the conceptual foundations of XAI, interpretability methods, the evaluation of explanations, human–AI interaction, and clinical decision support. The synthesis identifies six main themes: the conceptual foundations of XAI, model-explanation methods, application to high-stakes decision-making, the evaluation of explanation quality, the integration of explanation with user needs, and computational and governance challenges. The literature indicates that the success of XAI cannot be measured by the ability to produce technical explanations alone; it must also consider the fidelity of explanations to the model, user comprehension, contextual relevance, and the effect on decisions. The article offers a conceptual framework for understanding the integration of XAI into decision support systems and identifies a research agenda oriented toward empirical evaluation, user-centered design, and responsible AI governance.

Keywords: Explainable Artificial Intelligence; decision support systems; interpretability; algorithmic transparency; decision-making; scoping review

1. Introduction

The use of artificial intelligence (AI) has expanded the capacity of decision support systems to process data, recognize patterns, predict likely events, and generate recommendations that help individuals and organizations. In environments marked by informational complexity and uncertainty, AI-based systems can support clinical decisions, risk assessment, business analysis, feasibility evaluation, and other activities that require the systematic processing of information. Even so, an increase in predictive capability is not always matched by an increase in users' ability to understand the reasoning behind an output. This becomes important when the resulting decisions affect the safety, rights, welfare, or economic interests of individuals and organizations.

A number of machine-learning models, particularly those with complex structures, are often treated as black boxes because the relationship among inputs, computational process, and outputs is difficult to grasp directly. Adadi and Berrada describe XAI as a field that has grown to open up the characteristics of models that were previously hard to interpret [1]. The broader push toward explainability is often associated with programs such as the DARPA XAI initiative [2]. Guidotti et al. likewise map a range of approaches to explaining black-box models, including techniques that seek to identify the factors contributing to a prediction [3]. This development shows that interpretability concerns not only the technical capability of a model but also the way people understand and use the information AI produces.

Arrieta et al. place XAI within the context of responsible AI through a discussion of concepts, taxonomies, opportunities, and challenges [4]. In that perspective, explanations can help users judge whether a system's output is reasonable, recognize possible errors, and understand the limits of a model. However, the availability of an explanation does not automatically guarantee that a system can be trusted or that the resulting decision is fair. Lipton shows that interpretability carries several meanings that are not always distinguished consistently in research [5]. Terms such as transparency, interpretability, explainability, and the justification of a decision therefore need to be positioned conceptually so that the evaluation of XAI does not rest on vague definitions.

In decision support systems, this issue becomes more complex because AI output forms part of a human decision-making process. A system not only produces predictions but can also influence how users identify problems, weigh alternatives, and determine action. Preece emphasizes the importance of the question "why" in explaining intelligent systems [6]. Drawing on the social sciences, Miller argues that explanation for people should be understood as a communicative practice bound up with context, needs, and the way people form understanding [7]. A technically correct explanation is therefore not necessarily the most useful explanation for a decision-maker. These challenges are especially clear in healthcare. AI-based clinical decision support systems can help interpret patient data and provide information to support professional judgment. Yet clinical decisions require an assessment of the patient context, scientific evidence, the risk of error, and professional responsibility. Antoniadi et al. identify the challenges and opportunities of XAI in machine-learning-based clinical decision support systems [8]. Xu et al. review the interpretability of clinical decision support systems from technological and medical perspectives [9]. Freyer et al., meanwhile, examine the ethical reasons underlying the need for explainability in AI decision support systems in healthcare [10]. These studies show that the need for explanation cannot be separated from the characteristics of a decision and its consequences.

Although XAI research has grown, a mapping is still needed that connects conceptual foundations, explanation techniques, user needs, evaluation, and application context within decision support systems. A review focused only on algorithms risks overlooking how explanations are used in practice. Conversely, a discussion that emphasizes only user trust may neglect the fidelity of explanations to a model's behavior. Kostopoulos et al. specifically review developments in XAI-based decision support systems [11], while Saranya and Subhashini map developments in XAI models and applications more broadly [12]. Both studies provide an important basis for situating the present work within the existing literature.

Against this background, this article aims to map the scope of research on applying XAI to decision support systems. Specifically, it addresses four research questions: (1) how are the concepts and approaches of XAI understood in the context of decision support systems; (2) which explanation methods are relevant to supporting an understanding of AI model outputs; (3) how are explanations evaluated in technical, cognitive, and contextual terms; and (4) what are the main challenges and future research opportunities in integrating XAI into decision support systems? The contribution of this article lies in a conceptual mapping that connects XAI methods with the needs of decision-makers and the consequences of their application. The discussion is therefore oriented not only toward the algorithmic ability to produce explanations but also toward how explanations can be used appropriately, critically, and responsibly in the decision-making process.

2. Method

Research Design

This article is designed using a structured literature review with a scoping-review orientation. A scoping review is used, in principle, to map the breadth and characteristics of research in a field, to identify core concepts, to group types of evidence, and to locate research gaps. Unlike a systematic review that focuses on a narrower question or the evaluation of specific evidence, a scoping approach allows the exploration of the diverse perspectives and methods that shape a field of study.

In this study, the mapping focuses on the application of Explainable Artificial Intelligence to decision support systems. The literature underlying the analysis includes conceptual work on XAI, surveys of model-explanation methods, discussions of interpretability evaluation, research on human–AI interaction, and reviews of clinical decision support systems. Using references across these categories is intended to illuminate the relationships among a model's technical capability, the form of explanation, user needs, and the characteristics of a decision.

It should be emphasized that the sources analyzed in this draft are drawn from a provided reference list. Documentation of the databases, search dates, keywords, number of records, deduplication, and screening outcome is not available. This article therefore cannot be treated as a scoping-review report that has completed a reproducible search and selection procedure.

Scope of the Review

The scope of the study is limited to research concerned with the concept and application of XAI to support the understanding, evaluation, or use of AI output in decision-making. The literature is grouped into five analytical domains.

The first is conceptual foundations, covering the definition, taxonomy, purpose, and limits of interpretability. This category draws on Adadi and Berrada [1], Lipton [5], Arrieta et al. [4], and Vilone and Longo [13].

The second is explanation methods, covering techniques for explaining models locally or globally, feature-attribution approaches, and techniques for interpreting complex models. This discussion refers, among others, to Ribeiro et al. [14], Lundberg and Lee [15], Guidotti et al. [3], and Samek et al. [16].

The third is the evaluation of explanations, encompassing fidelity to the model, technical validity, comprehensibility, and the fit of explanations to user needs. The main literature in this category includes Doshi-Velez and Kim [17], [18], Tomsett et al. [19], and Vilone and Longo [13].
The fourth is human–AI interaction and user needs. This category highlights how explanations are understood, questioned, and used by decision-makers, drawing on Miller [7], Preece [6], Kaur et al. [20], and Liao et al. [21].
The fifth is application to decision support systems and high-stakes contexts, particularly healthcare. This category includes Kostopoulos et al. [11], Antoniadi et al. [8], Xu et al. [9], Freyer et al. [10], Gerdes [22], and Ezeji et al. [23].

Analytical Procedure

The analysis was carried out through a conceptual thematic synthesis. First, the references were grouped by their main focus—such as interpretability theory, explanation techniques, evaluation, interaction design, clinical application, and computational complexity. Second, each group was analyzed to identify the central issues, conceptual contributions, and limitations relevant to applying XAI to decision support systems. Third, the relationships among groups were formulated to build a mapping that connects explanation methods with user goals and decision characteristics. Fourth, these themes were used to identify challenges and an agenda for future research.
This synthesis does not produce publication frequencies, annual trends, or a quantitative distribution of articles, because the bibliographic data and selection process required for such calculations have not been assembled systematically. The grouping presented is an analytical framework for reading the provided literature, not the result of counting a corpus of publications screened with a complete scoping-review protocol.
To develop this into a scoping-review article ready for submission to a journal, the procedure would need to be supplemented with a transparent search protocol, inclusion and exclusion criteria, documentation of screening, a data-extraction form, and reporting of the selection flow in line with an appropriate guideline, such as PRISMA-ScR where used in the conduct of the research.

3. Results

Conceptual Foundations of Explainable Artificial Intelligence

The conceptual synthesis shows that XAI involves more than presenting information about how an algorithm works. XAI concerns the ability to produce explanations that help people understand the relationship among input data, the inference process, and a system's output. Adadi and Berrada map the development of XAI as a response to the limits of understanding complex models [1]. Arrieta et al. broaden this discussion by placing XAI within a framework of responsible AI [4], while Guidotti et al. give particular attention to methods for explaining black-box models [3].
One important conceptual finding is that interpretability and explainability are not always identical. Interpretability generally refers to the extent to which a model's behavior can be understood, whereas explainability can refer to the process of producing representations or reasons that help users understand a system's output. This distinction matters because a model that is relatively easy to interpret directly does not always require an additional explanation mechanism, while a complex model may need post-hoc methods to help users understand its predictions.
Lipton shows that the term interpretability is often used with varied meanings [5]. Vilone and Longo likewise discuss the various notions of explainability and approaches to its evaluation [13]. In decision support systems, this distinction requires developers to determine in advance what they wish to explain: the structure of a model, the contribution of features to a prediction, the general behavior of a model, or the reasons relevant to a particular decision.
Rudin raises a critique of using black-box models that are then explained through post-hoc techniques in high-stakes decisions [24]. She advocates the use of intrinsically interpretable models where such models can deliver adequate performance. This argument underscores that XAI should not always be understood as an effort to add explanations to complex models. Choosing a more transparent model from the outset can also be a strategy for supporting decisions that demand high accountability.
The conceptual foundations of XAI in decision support systems can thus be understood through three interrelated orientations: understanding a model's behavior, explaining a particular output, and supporting human judgment of whether a recommendation is appropriate.

Methods for Explaining AI-Based Decision Support

The literature shows a range of methods used to explain AI models. These methods have different aims, assumptions, and limitations. The choice of method therefore needs to consider the characteristics of the model, the type of output, the context of use, and the needs of the decision-maker.
The first approach is intrinsic interpretability—the use of models whose structure or decision rules are relatively easy to understand. Rule-based models, decision trees of limited complexity, and certain simple models can help

users trace the relationship between input variables and results. This perspective is consistent with Rudin's critique of reliance on black-box models in high-stakes decisions [24]. Even so, a readable structure does not automatically guarantee that users can understand all of a model's consequences, especially when the model contains many rules or is used in a complex context.

The second approach is post-hoc explanation—explanations generated after a model has made a prediction. One widely discussed technique is LIME, introduced by Ribeiro et al. [14]. This approach builds a local surrogate model to approximate a model's behavior around a given prediction. In this way, users can gain a sense of the factors contributing to a particular result without needing to understand the entire structure of the original model. However, the quality of the explanation depends on how the local approximation is constructed and how well the surrogate represents the model's actual behavior.

The third approach is SHAP, developed within a unified framework by Lundberg and Lee [15]. This method uses the concept of Shapley values to attribute feature contributions to a prediction. In decision support systems, feature attribution can help users see which factors raise or lower a predicted value. Although useful for analysis, feature contributions to a prediction should not be read directly as causal relationships. Attribution values can also be affected by dependence among features, computational assumptions, and the characteristics of the data.

The fourth approach includes visualization and attribution analysis for deep-learning models. Samek et al. discuss methods for understanding, visualizing, and interpreting deep-learning models [16]. Visualization techniques can help reveal the parts of an input related to a model's output, such as the regions of an image that influence a prediction. However, Tomsett et al. show the importance of checking the validity of saliency metrics [19]. A convincing-looking visualization does not necessarily provide an explanation that genuinely represents the model's mechanism.

Overall, XAI methods can be mapped by the scope of explanation: global explanations of model behavior, local explanations of particular predictions, and explanations designed to support user interaction. These can be complementary but cannot be interchanged without considering the purpose of the analysis.

XAI in High-Stakes Decision Support Systems

Applying XAI to high-stakes decisions requires particular care because the consequences of error can extend beyond the technical. In clinical decision support systems, for example, a model can help estimate risk or identify patterns related to a patient's condition. Yet a model's recommendation must still be placed within the context of medical evidence, individual circumstances, and the judgment of professionals.

Antoniadi et al. review the challenges and opportunities of XAI in machine-learning-based clinical decision support systems [8]. Xu et al. discuss interpretability from technological and medical perspectives, underscoring the need to bring computational capability together with the needs of clinical practice [9]. Freyer et al. place explainability within a framework of ethical reasons [10], while Gerdes discusses the role of explainability in AI-supported medical decision-making [22]. A broader review of XAI in healthcare, including current trends and challenges, further frames these issues [25].

From this literature, the need for explanation in clinical decisions can be understood through several considerations. First, explanation helps professionals check whether a system's recommendation is consistent with patient information and relevant medical knowledge. Second, explanation can support communication about the basis of a recommendation, although it does not replace professional responsibility. Third, explanation can help identify potential model mismatches, including when a model produces output that is implausible in a particular context.

Freyer et al. show the relevance of ethical considerations in the discussion of explainability for AI decision support systems in healthcare [10]. However, explainability should not be treated as the sole ethical requirement. A system that provides an explanation can still have problems with data quality, bias, clinical validity, or the distribution of risk. XAI therefore needs to be combined with evaluation and oversight mechanisms appropriate to the risk of the decision.

Kostopoulos et al. specifically examine XAI-based decision support systems [11]. Their work provides a basis for treating explanation as part of the design of a decision support system rather than as a stand-alone, add-on feature. Within this framing, the central question is not only whether a system can explain a prediction but whether that explanation helps users evaluate it and make a more appropriate decision.

Human-Centered Explanations and User Requirements

The comprehensibility of an explanation is a matter distinct from the availability of technical information. A detailed explanation may be useful to a data scientist but not necessarily effective for a physician, manager, analyst, or non-technical user. A decision support system therefore needs to consider users' background knowledge, responsibilities, goals, and situation.

Miller shows that explanation can be understood through the lens of the social sciences, including how people request, give, and use reasons [7]. This approach broadens XAI from a matter of feature visualization to a matter of communication. A good explanation is not always a list of all the variables contributing to a prediction. In

certain situations, users may need the main reasons, a comparison with alternatives, or the conditions that would change a recommendation.

Liao et al. examine how an understanding of users' questions about AI can inform the design of XAI user experiences [21]. This perspective is relevant to decision support systems because user needs are not always uniform. One user may wish to know why a particular recommendation arose, while another wishes to understand why a certain alternative was rejected or how a change in input affects the result.

Kaur et al. study how data scientists use interpretability tools in their work practice [20]. Their study is a reminder that interpretability tools are used within particular workflows and can be understood differently according to a user's task. The success of XAI therefore needs to be evaluated in the context of real use, not only on the basis of visual output or researchers' assessment of a method.

From the perspective of decision support systems, a user-centered explanation needs to meet at least three requirements. First, relevance—the explanation relates directly to the question at hand. Second, comprehensibility—the information can be understood by the intended user. Third, usefulness—the explanation helps the user evaluate a recommendation or determine the next step. These three requirements must be balanced against technical validity so that simplification does not remove important information or create mistaken understanding.

Evaluation of Explainability

Evaluation is one of the central challenges in developing XAI. No single measure can universally determine the quality of every kind of explanation. An explanation that is good for one purpose can be ill-suited to another. A simple explanation, for example, may help users understand a prediction but be too simple to support a technical audit.

Doshi-Velez and Kim call for the development of a more rigorous science of interpretable machine learning, including attention to how interpretability is defined and evaluated [17]. Doshi-Velez and Kim subsequently discuss considerations for evaluation and generalization in interpretable machine learning [18]. Both works underscore that claims about interpretability need to be supported by evaluation methods suited to the purpose of use.

Vilone and Longo map the various notions of explainability and approaches to XAI evaluation [13]. Based on the conceptual synthesis, evaluation can be distinguished along several dimensions. First, fidelity—the extent to which an explanation represents the behavior of the model being explained. Second, stability—the consistency of an explanation when relevant inputs undergo small changes. Third, comprehensibility—the extent to which users can understand the content of an explanation. Fourth, usefulness—the extent to which an explanation helps accomplish a decision-making task. Fifth, contextual validity—the fit of an explanation with the domain and the risk of application.

Tomsett et al. show the importance of sanity checks on saliency metrics [19]. The key implication is that a metric or visualization should not be accepted simply because it produces a plausible-looking pattern. Researchers need to check whether a method is genuinely sensitive to a model's behavior and whether the results hold under appropriate testing.

Evaluation also needs to distinguish technical assessment from human assessment. Technical evaluation tests the characteristics of an explanation on the basis of the properties of the model and the data, whereas user-based evaluation tests whether an explanation can be understood and used appropriately. The two approaches are complementary. An explanation that is mathematically consistent does not necessarily help users, while an explanation that users like is not necessarily faithful to the model.

Computational Complexity and Practical Implementation

Applying XAI also faces issues of computational complexity and technical integration. Some explanation methods require additional computation after a model produces a prediction. In systems that handle large volumes of data or must provide recommendations quickly, computational cost can affect response time, scalability, and operating cost.

Ezeji et al. specifically discuss computational complexity in explainable decision support systems [23]. This issue shows that the choice of an XAI method cannot rest on the conceptual quality of the explanation alone. A more complex method may produce more detailed information but can become less practical if it requires substantial resources or slows the decision-making process.

There is also a trade-off between level of detail and ease of understanding. An explanation that is too brief can omit important information, whereas one that is too long can overburden the user. In addition, post-hoc explanation techniques can be limited in representing a model's behavior comprehensively. Systems therefore need to be designed with attention to the purpose of the explanation, time constraints, audit needs, and users' capabilities.

Technical complexity is also bound up with system maintenance. When a model is updated, the data distribution shifts, or new features are added, the explanations produced can change. A responsible decision support system needs to provide procedures for re-testing so that explanations remain relevant to the version of the model in use.

Thematic Mapping of the Literature

Based on the conceptual synthesis, the literature can be mapped into six main themes, as shown in Table 1. This mapping shows the relationship among research focus, main contribution, and implications for decision support systems.

Table 1. Thematic Mapping of XAI Research in Decision Support Systems

Theme	Main focus	Representative references	Implications for decision support
Conceptual foundations of XAI	Definition, taxonomy, interpretability, and the limits of explanation	Adadi and Berrada [1]; Lipton [5]; Arrieta et al. [4]; Vilone and Longo [13]	Establish the meaning and purpose of explanation before choosing a method
Explanation methods	Intrinsic models, local explanations, feature attribution, and visualization	Ribeiro et al. [14]; Lundberg and Lee [15]; Guidotti et al. [3]; Samek et al. [16]	Select techniques suited to the model and the decision need
Evaluation of explanations	Fidelity, validity, stability, and user-based evaluation	Doshi-Velez and Kim [17], [18]; Tomsett et al. [19]; Vilone and Longo [13]	Avoid unsupported claims about explanation quality
Human–AI interaction	User needs, questions, communication, and use of interpretability tools	Miller [7]; Preece [6]; Kaur et al. [20]; Liao et al. [21]	Design explanations to fit users' tasks, competence, and context
High-stakes decisions	Clinical interpretability, ethical rationale, and professional responsibility	Antoniadi et al. [8]; Xu et al. [9]; Freyer et al. [10]; Gerdes [22]; Kostopoulos et al. [11]	Integrate explanation with domain validation and human oversight
Implementation and efficiency	Computational cost, scalability, and operational feasibility	Ezeji et al. [23]; Saranya and Subhashini [12]	Balance explanation quality, response time, and implementation cost

Note. Source: Authors' conceptual synthesis based on the provided references. This table is not the result of frequency counts or coding of a corpus of articles screened systematically.

The mapping shows that XAI research does not follow a single, homogeneous line of development. Some research focuses on the ability to explain models, some examines the validity and evaluation of explanations, while another body of work highlights the use of explanations in clinical decisions or in human–AI interaction. Integrating these perspectives is needed so that the development of XAI does not stop at algorithmic innovation but also considers the effectiveness and responsible use of systems.

4. Discussion

XAI as an Integrated Component of Decision Support

The literature synthesis indicates that XAI needs to be understood as part of the architecture of a decision support system rather than as an add-on component placed after a model has been developed. A decision support system essentially connects data, analytical models, the presentation of information, and user action. If explanation is not designed together with these components, the explanations produced risk being ill-suited to the needs of the decision process.

Kostopoulos et al. provide specific context on developments in XAI-based decision support systems [11]. In line with these developments, Arrieta et al. place XAI within a framework of responsible AI [4]. The two perspectives can be combined through a design approach that considers three questions from the outset: what needs to be explained, to whom the explanation is given, and which decision will be supported.

The first question concerns the target of the explanation. Developers need to determine whether users require a general understanding of model behavior, the reasons behind a particular prediction, or information about how results change when inputs change. The second question concerns the characteristics of users. An explanation for model developers may require technical measures, whereas an explanation for operational decision-makers may need to highlight the relevant factors and their practical implications. The third question concerns the consequences of the decision. The higher the risk of a decision, the more important it is to test the quality of the explanation, document its limitations, and provide oversight mechanisms.

The success of XAI should therefore not be judged merely by the presence of an explanation module. It needs to be judged by the contribution of the explanation to the process of evaluating, understanding, and appropriately using a recommendation.

The Trade-Off Between Predictive Performance and Interpretability

One central issue in XAI is the relationship between predictive performance and interpretability. Complex models can offer advantages on particular tasks, but their behavior may be harder to understand. Conversely, simpler

models can offer greater transparency, although they do not necessarily deliver the same performance on all problems.

Rudin criticizes the practice of explaining black-box models in high-stakes decisions when more interpretable models could be used [24]. This critique challenges the assumption that complex models must always be retained and then supplemented with post-hoc explanations. In some contexts, the more appropriate approach may be to choose an intrinsic model that can be understood from the outset, provided the quality of prediction and its fit to the task can be justified.

That argument does not mean, however, that all complex models should be abandoned. The suitability of a model depends on the characteristics of the problem, data quality, prediction needs, and the risk of use. Design decisions therefore need to consider an empirical comparison between intrinsically interpretable models and complex models supplemented with explanation methods.

In practice, developers can use a staged selection process. First, test whether a relatively transparent model can meet the decision need. Second, if a complex model provides substantial benefit, select an appropriate explanation technique and evaluate its limitations. Third, determine whether the results of that evaluation are sufficient to support use in the intended context. This approach avoids two extremes: treating transparency as always more important than every aspect of performance, or treating predictive accuracy as sufficient to justify the use of a model.

The Explanation Must Be Faithful and Useful

The literature on explanation methods shows that explanations can take different forms. LIME, from Ribeiro et al., seeks to approximate a model's behavior locally [14], while SHAP, from Lundberg and Lee, provides a framework for attributing feature contributions to a prediction [15]. Other visualization and attribution methods also offer ways to examine the relationship between a model's inputs and outputs [16].

The ability to produce an explanation, however, is not the same as the ability to produce a correct explanation. An explanation can appear intuitive yet not fully reflect a model's behavior. The critique of the ambiguity of interpretability discussed by Lipton [5] and the need for rigorous evaluation in Doshi-Velez and Kim [17], [18] show the importance of distinguishing a convincing explanation from a validated one.

Tomsett et al. stress the need for checks on saliency metrics [19]. In decision support systems, this principle can be applied by testing whether an explanation changes reasonably when the model or input changes, whether a method still gives plausible results under control tests, and whether the explanation reflects the factors the model genuinely uses.

Beyond technical fidelity, an explanation needs to be useful to the user. Miller shows the importance of understanding explanation through a social and cognitive lens [7]. An explanation that is too complex can reduce practical benefit even when it contains correct technical information. Conversely, an explanation that is oversimplified can remove important uncertainty or limitations.

The implication is that developing XAI requires multidimensional evaluation. An explanation must be sufficiently faithful to the model, understandable to the intended user, and relevant to the decision under consideration. These three dimensions need to be evaluated separately before concluding that an XAI technique is effective.

Human-Centered Design and Decision-Making Context

AI explanation ultimately resides in the relationship between a system and the people who use it. The design of XAI therefore needs to consider how users ask questions, understand reasons, compare alternatives, and decide whether to follow a recommendation.

Preece raises the importance of the question "why" in intelligent systems [6], while Liao et al. discuss how an understanding of users' questions can guide the design of XAI experiences [21]. Kaur et al. likewise show the relevance of using interpretability tools in data scientists' work practice [20]. This literature draws attention to the need for explanations not only to be available but also to fit users' actual activities.

In decision support systems, explanations can be structured at several levels. The first level provides a summary of the main reasons for users who need quick information. The second presents the detailed factors contributing to a prediction. The third provides technical information for audit or deeper examination. A tiered structure of this kind is a testable design implication, not a guarantee that one format will suit all users.

Users also need to understand the limits of an explanation. A feature's contribution to a prediction does not, in itself, prove that the feature causes the result. Likewise, an explanation available for one case does not automatically explain a model's behavior in all cases. The development of XAI interfaces should explicitly distinguish predictive association, uncertainty, and causal relationships where such information is relevant to a decision.

A user-centered approach thus needs to combine technical testing with empirical study. Developers need to test not only whether users like an explanation but also whether it helps them detect errors, avoid mistaken conclusions, and make more appropriate decisions.

Ethical and Professional Accountability in High-Stakes Decisions

The need for XAI becomes more pressing when systems are applied to decisions with significant consequences. In healthcare, a system's recommendation can influence the process of diagnosis, risk assessment, or the choice of treatment. Xu et al. [9], Antoniadi et al. [8], Gerdes [22], and Freyer et al. [10] provide a basis for understanding that interpretability and explainability have technical, medical, and ethical dimensions.

Explanation can help professionals judge whether a system's output fits the available information. However, explanation cannot replace clinical validation, data quality, professional judgment, or institutional responsibility. If an explanation is inaccurate or fosters excessive confidence, users may become less critical of an AI recommendation.

Explainability therefore needs to be positioned as one component of system governance. Other components include model validation, documentation of limitations, performance monitoring, error-handling mechanisms, and the assignment of responsibility to users and system managers. In high-stakes decisions, human oversight must be designed as part of the decision process rather than as a formal approval of a system's output.

This perspective also makes clear that trust is not a goal to be maximized without limit. The more appropriate goal is calibrated trust—a level of trust aligned with a system's capabilities and limitations. A good explanation should help users know when a recommendation is worth considering, when it needs to be verified, and when it should not be followed.

Research Gaps and Future Research Agenda

Based on the conceptual mapping, several research agendas can be developed to strengthen the application of XAI in decision support systems.

First, standardizing the definitions and purposes of explanation. The diversity of meanings of interpretability discussed by Lipton [5] and Vilone and Longo [13] shows the need for clear operational definitions. Future research can compare the effectiveness of different kinds of explanation for different tasks and user groups, distinguishing global, local, technical, and decision-oriented explanations.

Second, developing multidimensional evaluation. Research needs to integrate measures of technical fidelity, stability, comprehensibility, and usefulness in decision-making. The rigorous evaluation approach called for by Doshi-Velez and Kim [17], [18] needs to be applied alongside validity testing of explanation methods as discussed by Tomsett et al. [19]. Research should also avoid using a single metric as the sole indicator of XAI quality.

Third, empirical study of user behavior. Kaur et al. [20], Liao et al. [21], and Miller [7] show the importance of the human dimension in interpretability. Future research can test how explanations affect users' ability to detect errors, calibrate trust, consider alternatives, and make decisions. Evaluation should use not only subjective perceptions but also indicators of task performance.

Fourth, evaluation in high-stakes contexts. XAI research needs to test how explanations work on decisions of differing risk and complexity. In healthcare, for example, studies need to consider professional needs, patient conditions, and the limits of using a model's recommendation. Findings from one context should not be generalized directly to another without validation.

Fifth, efficiency and scalability. The computational complexity discussed by Ezeji et al. [23] opens an opportunity to evaluate the cost, response time, and resource needs of various XAI methods. Future research can compare explanation techniques by quality, efficiency, and implementation feasibility, particularly in systems that require rapid responses.

Sixth, the relationship between explanation and AI governance. Research needs to examine how explanations can be used in auditing, documentation, oversight, and accountability. Explanations should be provided not only to end users but also to support evaluation and oversight processes suited to the characteristics of a system.

These agendas show that developing XAI needs to move from an orientation focused solely on algorithms toward research that integrates technical, human, organizational, and governance aspects.

5. Conclusion

This article develops a structured literature synthesis on the application of Explainable Artificial Intelligence to decision support systems. Based on the provided references, the discussion maps six main themes: the conceptual foundations of XAI, explanation methods, application to high-stakes decisions, the evaluation of explanations, human–AI interaction, and computational and implementation complexity.

The synthesis shows that XAI cannot be understood merely as the ability to produce visualizations or the reasons behind a model's predictions. XAI needs to be positioned as a mechanism that helps users understand, evaluate, and appropriately use AI output. To achieve this, explanation methods must be chosen on the basis of the characteristics of the model, user needs, and the consequences of the decision. Explanations also need to be tested for fidelity to a model's behavior, stability, comprehensibility, and usefulness in the decision-making process.

In high-stakes contexts, particularly healthcare, explainability has technical, ethical, and professional relevance. However, explainability does not replace model validation, data quality, human oversight, or accountability mechanisms. A responsible decision support system needs to combine explanation with testing, documentation of limitations, and oversight procedures appropriate to the risk.

The contribution of this article is a conceptual mapping framework that connects XAI methods with evaluation and the needs of decision-makers. The future research agenda needs to emphasize multidimensional evaluation, user-centered design, empirical study in real contexts, computational efficiency, and the integration of XAI with AI governance. With such an approach, XAI can support decision support systems that are not only predictive but also examinable and used responsibly.

Even so, this article remains a literature synthesis based on the provided references rather than a complete scoping-review report that has undergone systematic searching and screening. Further research needs to document the search strategy, selection criteria, data-extraction process, and article-selection flow so that the coverage and reproducibility of the review can be justified academically.

Declarations

Acknowledgements. The authors have no additional individuals or institutions to acknowledge for their contributions to this research.

Author Contributions. Alya Kartikasari contributed to the conceptualization, methodology, literature synthesis, and preparation of the original manuscript. Reyhan Firmansyah contributed to the literature synthesis, interpretation of the reviewed literature, and review and editing of the manuscript. Both authors contributed to the development of the article and approved the final manuscript.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest. The authors declare that they have no conflict of interest regarding the publication of this article.

Data Availability. This study is a literature-based review. All sources used in the conceptual synthesis are listed in the reference section, together with their available DOI links or repository links.

6. References

[1] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," IEEE Access, vol. 6, pp. 52138–52160, 2018. https://doi.org/10.1109/ACCESS.2018.2870052

[2] D. Gunning, "Explainable artificial intelligence (XAI)," Defense Advanced Research Projects Agency (DARPA), 2017. [Online]. Available: https://www.darpa.mil/program/explainable-artificial-intelligence

[3] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," ACM Computing Surveys, vol. 51, no. 5, art. 93, 2018. https://doi.org/10.1145/3236009

[4] A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82–115, 2020. https://doi.org/10.1016/j.inffus.2019.12.012

[5] Z. C. Lipton, "The mythos of model interpretability," Queue, vol. 16, no. 3, pp. 31–57, 2018. https://doi.org/10.1145/3236386.3241340

[6] A. Preece, "Asking 'why' in AI: Explainability of intelligent systems—Perspectives and challenges," Intelligent Systems in Accounting, Finance and Management, vol. 25, no. 2, pp. 63–72, 2018. https://doi.org/10.1002/isaf.1422

[7] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," Artificial Intelligence, vol. 267, pp. 1–38, 2019. https://doi.org/10.1016/j.artint.2018.07.007

[8] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney, "Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review," Applied Sciences, vol. 11, no. 11, art. 5088, 2021. https://doi.org/10.3390/app11115088

[9] Q. Xu et al., "Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: A systematic review," Journal of Healthcare Engineering, vol. 2023, art. 9919269, 2023. https://doi.org/10.1155/2023/9919269

[10] N. Freyer, D. Groß, and M. Lipprandt, "The ethical requirement of explainability for AI-DSS in healthcare: A systematic review of reasons," BMC Medical Ethics, vol. 25, art. 104, 2024. https://doi.org/10.1186/s12910-024-01103-2

[11] G. Kostopoulos, G. Davrazos, and S. Kotsiantis, "Explainable artificial intelligence-based decision support systems: A recent review," Electronics, vol. 13, no. 14, art. 2842, 2024. https://doi.org/10.3390/electronics13142842

[12] A. Saranya and R. Subhashini, "A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends," Decision Analytics Journal, vol. 7, art. 100230, 2023. https://doi.org/10.1016/j.dajour.2023.100230

[13] G. Vilone and L. Longo, "Notions of explainability and evaluation approaches for explainable artificial intelligence," Information Fusion, vol. 76, pp. 89–106, 2021. https://doi.org/10.1016/j.inffus.2021.05.009

[14] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144. https://doi.org/10.1145/2939672.2939778

[15] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, vol. 30, 2017. [Online]. Available: https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

[16] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," arXiv:1708.08296, 2017. https://arxiv.org/abs/1708.08296

[17] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv:1702.08608, 2017. https://arxiv.org/abs/1702.08608
[18] F. Doshi-Velez and B. Kim, "Considerations for evaluation and generalization in interpretable machine learning," in Explainable and Interpretable Models in Computer Vision and Machine Learning. Cham, Switzerland: Springer, 2018. https://doi.org/10.1007/978-3-319-98131-4_11
[19] R. Tomsett, D. Harborne, S. Chakraborty, P. Gurram, and A. Preece, "Sanity checks for saliency metrics," in Proc. AAAI Conf. Artificial Intelligence, vol. 34, no. 4, 2020, pp. 6021–6029. https://doi.org/10.1609/aaai.v34i04.5819
[20] H. Kaur, H. Nori, S. Jenkins, R. Caruana, H. Wallach, and J. Wortman Vaughan, "Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning," in Proc. 2020 CHI Conf. Human Factors in Computing Systems, 2020. https://doi.org/10.1145/3313831.3376219
[21] Q. V. Liao, D. Gruen, and S. Miller, "Questioning the AI: Informing design practices for explainable AI user experiences," in Proc. 2020 CHI Conf. Human Factors in Computing Systems, 2020. https://doi.org/10.1145/3313831.3376590
[22] A. Gerdes, "The role of explainability in AI-supported medical decision-making," Discover Artificial Intelligence, vol. 4, art. 29, 2024. https://doi.org/10.1007/s44163-024-00119-2
[23] I. N. Ezeji, M. Adigun, and O. Oki, "Computational complexity in explainable decision support system: A review," Journal of Intelligent & Fuzzy Systems, 2024. https://doi.org/10.3233/JIFS-219407
[24] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nature Machine Intelligence, vol. 1, pp. 206–215, 2019. https://doi.org/10.1038/s42256-019-0048-x
[25] N. A. Aziz, A. Manzoor, M. D. M. Qureshi, M. A. Qureshi, and W. Rashwan, "Unveiling explainable AI in healthcare: Current trends, challenges, and future directions," medRxiv, 2024. https://doi.org/10.1101/2024.08.10.24311735

© Journal of Sinar Data and Informatics (JSDI) — Sinar Asia Institute. This article is distributed under the Creative Commons Attribution (CC BY 4.0) license.