Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Argument-Based Evaluation for AI Recommendations in E-Commerce Decision Support: A Toulmin–FCM Approach

Li Niu [1] and Shihui Fu [2,3,*]

[1] Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Serdang 43400, Malaysia; gs61428@student.upm.edu.my
[2] School of Economics and Management, Taiyuan University of Technology, Taiyuan 030024, China
[3] Shanxi Key Laboratory of Data Factor Innovation and Economic Decision Analysis, Taiyuan 030024, China
* Correspondence: fushihui@tyut.edu.cn

Abstract

Artificial intelligence (AI) is increasingly being used to generate recommendations for e-commerce decision support. Existing XAI evaluation typically focuses on explanation attributes or user outcomes such as clarity, usefulness, trust, and acceptance but provides less insight into how an explanation justifies the recommendation itself. This study develops an argument-based evaluation framework by integrating Toulmin's model of argumentation with fuzzy cognitive maps (FCMs). Toulmin's model structures explanations through claim, ground, warrant, backing, qualifier, and rebuttal, while FCMs quantify the dependencies among these components and their combined support for a recommendation. The framework is applied to feature-based, example-based, and model-based explanations in an AI-assisted hotel-pricing scenario. The three explanation stimuli produce different argumentative profiles and Claim-support levels. A separate experiment using acceptance and trust as external outcomes shows a consistent pattern. The study contributes to XAI research by introducing argument-based evaluation, operationalizing argumentative structure quantitatively, and providing a common basis for comparing different explanation forms.

Keywords: explainable artificial intelligence; argument-based evaluation; AI recommendations; e-commerce decision support; fuzzy cognitive maps; Toulmin model

1. Introduction

Artificial intelligence (AI) has become a key technological infrastructure for e-commerce decision support. AI systems increasingly assist managers in decisions concerning pricing, promotion, inventory allocation, assortment planning, and revenue management [1,2]. In these tasks, machine learning models combine demand patterns, competitor information, transaction records, and contextual signals to generate commercially relevant recommendations [3,4]. However, the complexity that supports predictive performance can also make these recommendations difficult to interpret [5]. Explainable artificial intelligence (XAI) seeks to address this problem by making AI outputs and model behavior intelligible to relevant users [6,7]. By communicating the reasons underlying an AI output, explanations can improve transparency, facilitate scrutiny, and increase confidence in AI-generated recommendations [8,9]. This role becomes especially important in high-stakes managerial tasks such as dynamic pricing and revenue management, where AI recommendations may have substantial commercial consequences and managers remain accountable for the

Check for updates

Academic Editors: Ting Chi and Tunmin (Catherine) Jai

Received: 21 July 2026
Revised: 19 September 2026
Accepted: 20 September 2026
Published: 23 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

final decision [10]. In such settings, explanations should not only make recommendations understandable, but also help users assess their plausibility, recognize uncertainty, and make more informed judgments before acting on them.

Existing research has developed a wide range of criteria and outcomes for evaluating AI explanations, including explanation goodness, understanding, trust, acceptance, reliance, and task performance [11,12]. These measures are valuable for assessing whether explanations are useful and how users respond to them, but they provide less direct insight into how an explanation justifies the recommendation itself or where its reasoning is deficient. Argumentation-based XAI provides a promising way to address this limitation because it treats an explanation as a structured justification in which reasons can support, qualify, or challenge a conclusion [13–16]. Existing work has shown how argumentation can be used to represent explanations, construct interactive recommendations, and formalize structured forms of justification. However, considerably less attention has been devoted to translating argument structure into a systematic quantitative evaluation framework that can represent dependencies among its components and compare alternative forms of AI explanation [16]. This gap is particularly relevant to managerial decision support, where users must assess whether an AI recommendation is sufficiently justified before acting on it.

To address this gap, this study develops a framework for evaluating explanations of AI-generated recommendations by integrating Toulmin's model of argumentation with fuzzy cognitive maps (FCMs). Toulmin's model represents practical arguments through six interrelated elements: claim, ground, warrant, backing, qualifier, and rebuttal [17]. Its emphasis on evidence, inferential support, qualification, and exceptions makes it relevant to explanations that must justify recommendations without presenting them as universally valid [13–15]. FCMs complement this structure by representing concepts and their directional dependencies in a transparent computational network [18,19]. The integration therefore provides a basis for evaluating whether an AI explanation forms a coherent and appropriately bounded argument.

This study contributes to the literature in three ways. First, it shifts XAI evaluation from a predominant focus on user perceptions and downstream outcomes toward the internal justificatory structure of AI explanations. Second, it conceptualizes AI explanations as structured arguments whose elements operate jointly. This view clarifies the relationships among claim, ground, warrant, backing, qualifier, and rebuttal. Third, it operationalizes this argument-based view through FCMs, making it possible to model interdependencies among explanation elements and compare explanation types in terms of their support for AI outcomes.

The remainder of the paper is organized as follows. Section 2 reviews research on XAI evaluation and argumentation-based explanations. Section 3 introduces FCMs and their applicability to the present framework. Section 4 develops the Toulmin–FCM evaluation framework. Section 5 presents the hotel-pricing application and empirical validation. Section 6 discusses the findings, theoretical and practical implications, limitations, and future research directions.

2. Literature Review: XAI and Argumentation-Based Explanations

2.1. Explainable AI

Explainable artificial intelligence (XAI) broadly refers to methods and mechanisms that make the outputs, behavior, or reasoning of AI systems understandable to relevant users [20–23]. Early XAI research focused mainly on opening the "black box" of complex machine learning models and improving technical interpretability [24]. This stream has produced a variety of technical methods, such as SHAP, LIME, saliency-based methods,

together with technical criteria such as fidelity and consistency [25,26]. As AI has become increasingly embedded in practical decision-making, however, explainability has expanded beyond technical transparency toward a more human-centered perspective [27]. Explanations are now expected not only to reveal how an AI system produces an output but also to help users understand, scrutinize, and appropriately use that output in a specific decision context [6,20,22]. Accordingly, a second stream of research has focused on user-centered outcomes, including understanding, trust, acceptance, reliance, cognitive load, and decision confidence [28–32].

Although these measures provide a rich basis for explaining AI decisions, they offer only a partial account of explanation quality. Most focus either on the perceived or technical properties of an explanation or on downstream user responses and decision outcomes. They are therefore well suited to determining whether an explanation is clear, useful, trustworthy, or behaviorally effective, but they provide less direct insight into why an explanation constitutes a well-supported justification for an AI output. High trust or acceptance, for example, does not necessarily indicate that the evidence supporting a recommendation is sufficient, that the inferential connection between evidence and conclusion is explicit, or that uncertainty and exceptional conditions have been appropriately communicated. A related concern in AI-assisted professional decision-making is automation bias and over-reliance [33]. Decision-makers may place excessive weight on automated recommendations, particularly when system outputs appear credible or authoritative. Such reliance can weaken independent scrutiny and become especially consequential in high-stakes settings where managers remain responsible for the final decision [10,33]. Explanation evaluation should therefore consider not only whether an explanation increases trust or acceptance, but also whether it provides an adequate basis for users to critically assess when a recommendation should be accepted, qualified, or questioned.

2.2. Argumentation-Based Explanation

Argumentation-based XAI provides such a perspective by treating an explanation as a structured justification rather than merely a collection of informative attributes. Argumentation focuses on how a conclusion is supported by reasons, evidence, assumptions, qualifications, and possible objections [13–15,34]. From this perspective, the quality of an explanation depends not only on what information it contains, but also on how different pieces of information are connected to support, qualify, or challenge a conclusion. This orientation is particularly relevant to AI-assisted decision-making, because users often need to determine not only what an AI system recommends but also why the recommendation follows from the available information and under what conditions it should be questioned.

The use of argumentation in XAI reflects a broader effort to move beyond the disclosure of model information toward making AI outputs more justifiable and contestable. Argumentation-based approaches can make supporting reasons explicit, reveal assumptions underlying an inference, represent competing considerations, and expose possible counterarguments. In decision settings where users remain responsible for the final action, these functions are especially important because decision-makers must assess whether a recommendation is sufficiently supported rather than merely understandable.

Existing argumentation-based XAI research has explored different ways of representing and communicating structured reasoning. Studies have used argumentation to organize supporting and opposing reasons, generate interactive explanations for recommendations, and model relations among explanatory features [13,15,34]. More recent work has also begun to formalize properties for systematically evaluating explanations with increasingly rich argumentative structure [16]. These studies demonstrate that argumentation

can provide an explicit account of why an AI-generated conclusion should or should not be accepted.

2.3. Toulmin's Model of Argumentation

Among different theories of argumentation, Toulmin's model is particularly suitable for analyzing practical reasoning. Developed in *The Uses of Argument*, the model was intended to explain how conclusions are justified in real-world settings in which reasoning is often context-dependent, provisional, and open to challenge [17]. Toulmin's structure has also been used to organize argumentative explanations for recommender systems, providing a direct precedent for representing a recommendation as a claim supported by different forms of evidence and reasoning [35]. Unlike formal deductive logic, Toulmin's model focuses on the functional components that enable a practical claim to be supported and qualified. This orientation makes it well suited to AI-assisted decision contexts, where recommendations are rarely universally valid and must instead be assessed in light of available evidence, uncertainty, and possible exceptions.

The model represents an argument through six functionally related elements. At its core are the ground, warrant, and claim: the ground supplies evidence, the warrant authorizes the inferential move from that evidence, and the claim states the conclusion [17]. Backing provides additional support for the warrant, the qualifier indicates the strength of the claim, and the rebuttal identifies circumstances in which the claim or warrant may not hold. Prior recommender-system studies have shown that Toulmin-style argumentation can combine supportive and critical information and can be used to compare users' perceptions of argumentative explanation designs [15,35]. The model therefore depicts a layered and defeasible structure in which evidence, inference, support, uncertainty, and exceptions jointly determine how strongly a conclusion can be justified. The six elements and their basic relationships are summarized in Figure 1.

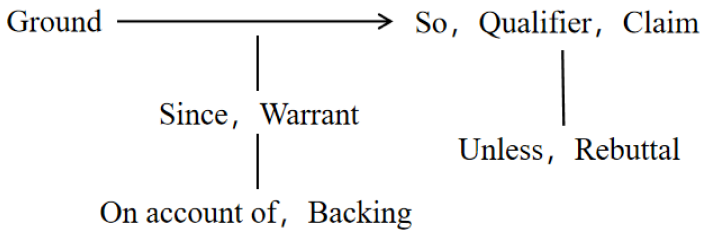


Figure 1. Toulmin's argument structure.

This structure maps naturally onto explanations of AI-generated recommendations. The claim corresponds to the recommendation produced by the AI system; the ground refers to the data, observations, or other evidence offered in support of that recommendation; the warrant explains why those grounds justify the recommended action; and the backing provides further model-, data-, or domain-based support for that reasoning. The qualifier communicates the confidence or uncertainty of the recommendation, while the rebuttal identifies exceptional conditions or countervailing considerations under which the recommendation may need to be reconsidered. This mapping enables explanation evaluation to move beyond asking whether information is present toward examining the specific justificatory function that each piece of information performs. Another important implication of Toulmin's model is that the quality of an explanation cannot be understood by treating its components as independent attributes.

Taken together, Toulmin's model provides a theoretically grounded basis for evaluating not only the presence of explanatory information, but also the structural adequacy of the reasoning that supports an AI-generated recommendation. However, an important gap

remains between conceptual argument structure and operational explanation evaluation. Existing evaluation approaches provide limited insight into how different justificatory components jointly determine the strength of an explanation, while argumentation-based XAI has devoted comparatively less attention to the systematic and quantitative evaluation of such structures. The present study addresses this gap by developing an operational framework that translates Toulmin's argumentative structure into a quantitative model for evaluating and comparing AI explanations. The overall conceptual framework is shown in Figure 2.

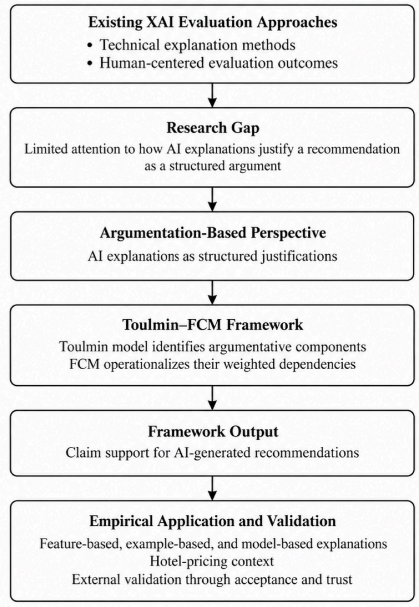


Figure 2. Conceptual framework of this study.

3. Fuzzy Cognitive Maps (FCM) for Operationalizing Argument Structure

3.1. Applicability of FCM to the Present Study

The literature review highlights the need to translate Toulmin's conceptual argument structure into an operational evaluation model. Such a model should preserve the dependencies among argumentative components and accommodate judgments whose strength may be difficult to express precisely. Fuzzy cognitive maps (FCMs) provide a suitable methodological basis for this purpose.

FCMs are well suited to theory-guided evaluation problems in which relationships among concepts are specified in advance while their strengths must be assessed through expert judgment. Toulmin's model defines the functional relationships among argumentative components, but it does not provide a quantitative mechanism for representing their relative influence. FCMs can express these relationships as directed links and assign graded weights to their strength. This allows the conceptual structure of an argument to be converted into a computable network.

Several alternative methods can also represent multiple variables or relationships, but they serve different analytical purposes. Weighted scoring and multicriteria decision methods are effective for aggregating evaluation criteria, yet they usually assign each criterion an independent contribution to an overall score. This is less suitable when the effect of one component depends on another component in the argument structure. Path analysis and structural equation modeling can represent mediated relationships, but they are mainly designed to estimate statistical associations from observed sample data. Bayesian networks represent conditional probabilistic dependencies and therefore require

a probabilistic interpretation of the relationships. In the present study, the relationships among argumentative components are theoretically specified, while their strengths are elicited from expert judgments. FCMs fit this setting because they can represent directional influence through graded and fuzzy weights. The comparison of alternative modeling approaches is shown in Table 1.

Table 1. Comparison of Alternative Modeling Approaches.

Method	Primary Purpose	Limitations for the Present Study
Weighted scoring/MCDM	Aggregate multiple evaluation criteria	Usually treats criteria as independent or directly weighted, which makes it difficult to represent structured dependencies among argumentative components
AHP/ANP	Derive criterion weights and interdependencies from expert judgments	More suitable for weighting and prioritization than for representing directional influence and propagation across an argument structure
DEMATEL	Identify and quantify influence relationships among factors	Mainly focuses on identifying cause–effect relationships rather than propagating node states through a structured evaluation network
Path analysis/SEM	Estimate structural relationships among observed or latent variables	Primarily designed for statistical estimation and hypothesis testing, rather than for representing graded influence within a predefined argument structure
Bayesian network	Represent probabilistic dependencies among variables	Requires probabilistic interpretation and conditional probability specification, which does not match the graded argumentative influence

3.2. Principles of FCM

A fuzzy cognitive map (FCM) is a computational method for representing concepts and the directional relationships among them [18,19,36,37]. Concepts are represented as nodes, and directed edges indicate the influence of one concept on another. Each edge is assigned a weight whose sign indicates the direction of influence and whose magnitude indicates its strength. FCMs have also been discussed as a useful representation for explainable AI because their weighted causal structure is directly inspectable [38].

FCMs are particularly useful when the structure of a problem is theoretically specified while the strength of relationships must be assessed through expert judgment. Linguistic assessments can be represented through fuzzy numbers and subsequently transformed into numerical weights. Once node values and edge weights are specified, influences can be propagated through the network according to its structural configuration. For FCMs with recurrent or feedback relationships, node activations are commonly updated iteratively until convergence or another stopping criterion is satisfied. For acyclic FCMs, by contrast, the effects of upstream concepts can be propagated directly to downstream nodes in a single forward pass following the direction of the network. Through this process, FCMs allow the effects of different concepts to be transmitted through the network. The resulting state reflects the combined influence of the specified relationships. When the underlying theory implies that two components are jointly necessary, their influences can also be combined non-additively—for example, through multiplicative aggregation—so that a low value in one component cannot be compensated for by a high value in another.

4. Construction of the Toulmin–FCM Evaluation Framework

4.1. Operationalization of the Argumentative Components

The framework contains six argumentative components: claim, ground, warrant, backing, qualifier, and rebuttal. The claim is the AI-generated recommendation whose justificatory support is evaluated. It serves as the terminal component of the argument structure. The ground represents the strength and adequacy of the evidence supporting the recommendation. The warrant captures the strength of the inferential connection between the evidence and the recommendation. The backing provides additional support for that inferential connection. The qualifier reflects the degree of strength or certainty attached to the recommendation. A stronger qualifier indicates that the recommendation is asserted with greater confidence, whereas a weaker qualifier indicates a more tentative conclusion. The rebuttal represents the strength of exceptional or countervailing conditions that constrain the warrant or the recommendation. A stronger rebuttal therefore indicates a greater challenge to the recommendation.

This operationalization distinguishes the argumentative force of information from the quality of its presentation. Clear disclosure of uncertainty or limitations may improve transparency, while the disclosed information itself may strengthen or weaken support for the recommendation. The operational definitions are shown in Table 2.

Table 2. Operational Definitions of the Toulmin-Based FCM Components.

Argumentative Component	Definition in the Present Framework
Claim	The AI-generated recommendation or conclusion whose justificatory support is evaluated
Ground	The strength and adequacy of the evidence supporting the recommendation
Warrant	The strength of the inferential connection linking the evidence to the recommendation
Backing	The strength of additional support that reinforces the warrant
Qualifier	The degree of strength or certainty attached to the recommendation
Rebuttal	The strength of exceptional or countervailing conditions that constrain the warrant or recommendation

4.2. Structural Relationships Among the Argumentative Components

The six elements form an argumentative dependency structure. Ground, warrant, and claim constitute the core chain: evidence supports reasoning, and reasoning supports the conclusion. Backing provides supporting information for the warrant. The qualifier adjusts the strength or certainty of the claim, while the rebuttal identifies exceptions that constrain the warrant or claim [17]. Figure 3 presents the resulting Toulmin-based FCM structure.

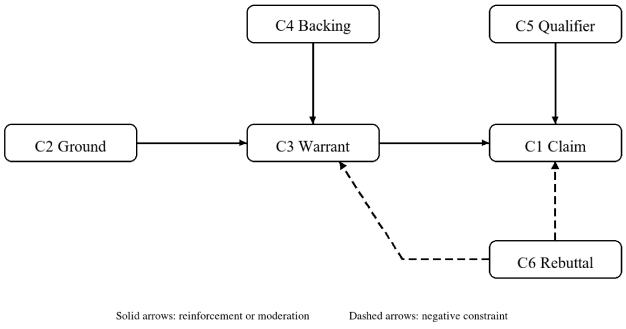


Figure 3. Dependencies among Toulmin elements.

These relationships preserve the functional logic of Toulmin's model. The framework therefore differs from an additive evaluation in which each component contributes independently to an overall score. The effect of one component may depend on its position within the argument structure. Strong grounds, for example, provide limited support when the warrant connecting the evidence to the recommendation is weak. Backing primarily contributes through its effect on the warrant. The negative paths associated with rebuttal require a specific interpretation. They represent the substantive constraining effect of exceptional conditions on the argument. They do not indicate that communicating limitations reduces explanation quality. The completeness or clarity of rebuttal disclosure is a separate property of explanation presentation.

4.3. Evaluation Procedure

The framework can be applied to different AI-supported decision contexts through three stages. First, the Toulmin-based structure specifies the relationships among the argumentative components. The presence and direction of these relationships are grounded in the theoretical logic described above. Their strengths can then be quantified for a particular decision context using appropriate empirical information or structured judgments.

Second, an explanation is represented through ratings of the argumentative components. Five components—ground, backing, qualifier, rebuttal, and warrant—are rated directly, indicating the strength of the evidence, the supporting basis, the qualification, the countervailing conditions, and the perceived inferential link contained in the explanation. The claim is treated as the terminal component whose support is generated through the network. It therefore does not require an explanation-specific initial quality rating.

Third, the FCM-based propagation mechanism combines the rated components according to the specified structure and weights. The effective warrant is first obtained by blending the rated warrant with its backing, discounted by the rebuttal; a core justification index then combines the ground and the effective warrant multiplicatively, so that evidence contributes to the conclusion only through the inferential link; the qualifier and rebuttal subsequently adjust this index; and a sigmoid mapping yields the final activation of the claim. This value represents the degree to which the recommendation is supported within the Toulmin structure.

The framework separates this structural assessment from broader user evaluations. Measures such as explanation goodness, trust, or acceptance capture users' perceptions and responses. They can be examined independently to assess whether stronger argumentative support corresponds with more favorable evaluations in a specific application. The framework itself does not depend on a specific AI application or explanation format. The structural relationships remain theory-guided, while their empirical weights and node activations can be estimated according to the characteristics of the focal decision context. The following section applies this framework to AI-generated hotel pricing recommendations.

5. AI-Generated Hotel Pricing Recommendations as an Empirical Context

5.1. Context and Development of Explanation Stimuli

This study applies the proposed framework to an AI-assisted hotel pricing scenario. Hotel pricing provides an appropriate setting because pricing decisions depend on changing demand and market conditions, while managers remain responsible for evaluating and implementing algorithmic recommendations. The empirical application uses a standardized hypothetical scenario in which an AI-based pricing system recommends a room rate based on hotel operating conditions and market information.

The recommendation and explanation materials were constructed as experimental stimuli. They were designed to represent three commonly used forms of AI explanation:

feature-based, example-based, and model-based explanations. Feature-based explanations describe how individual input factors contribute to a recommendation or prediction [39]. Example-based explanations present similar or comparable prior cases to help users interpret the current recommendation by analogy [9,40]. Model-based explanations provide a simplified representation of the decision logic or internal structure through which the system arrives at its output [41].

The selection of these three explanation forms was guided by prior XAI taxonomies and the characteristics of the focal decision context. The existing XAI literature has identified feature-based, example-based, and model-based approaches as important families of explanation methods, reflecting different sources of explanatory information: input-factor attribution, case-based reasoning, and model or decision-logic representation [20,23,39,42]. These three forms are particularly suitable for AI-assisted hotel pricing decisions because pricing recommendations typically require decision-makers to understand three types of information: which market and operational factors drive the recommended price, whether similar historical situations support the recommendation, and how the underlying decision logic connects available information to the suggested action. Therefore, feature-based, example-based, and model-based explanations provide complementary perspectives for evaluating how different forms of explanatory information contribute to the justification of AI-generated recommendations.

To reduce confounding across conditions, the three stimuli were developed around the same pricing scenario, recommendation, and core market information. The manipulation concerned the form in which the supporting information was organized and presented. The feature-based condition emphasized the contribution of individual pricing factors, the example-based condition presented comparable prior cases, and the model-based condition represented the decision logic leading to the recommendation. The stimuli were not generated from a deployed hotel-pricing model. They should therefore be interpreted as experimental representations of the three explanation forms, rather than as evidence of the technical behavior of a specific machine learning model. This design allows the decision context and recommended action to remain consistent while the form of explanatory information varies across conditions.

As shown in Figure 4, the feature-based explanation presents how different input features contribute to the room-rate recommendation. In this example, an active event increases the recommended rate by CNY 25 per night, strong OTA search popularity adds CNY 10, and the historical room-rate trend adds CNY 5. By contrast, competitor prices reduce the recommendation by CNY 10, while real-time occupancy and bookings and consumer price-comparison behavior each reduce it by CNY 5. Customer profile, regional economic indicators, and weather volatility have little influence on the current prediction. Together, these contributions produce a net upward adjustment of CNY 20 from the baseline rate. This explanation helps practitioners identify which market conditions drive the AI-generated recommendation and assess whether the direction and magnitude of the adjustment are commercially reasonable.

As shown in Figure 5, the example-based explanation presents two comparable historical cases: a product exhibition on 17 July with 800–1000 attendees, when the room rate was CNY 299 per night and occupancy reached 98%, and a technology forum on 24 July with 800–900 attendees, when the room rate was CNY 305 and occupancy reached 96%. These high-demand event cases provide an analogical basis for the current AI-generated pricing recommendation, allowing practitioners to assess whether the suggested price is consistent with the rates and occupancy levels observed under similar market conditions.

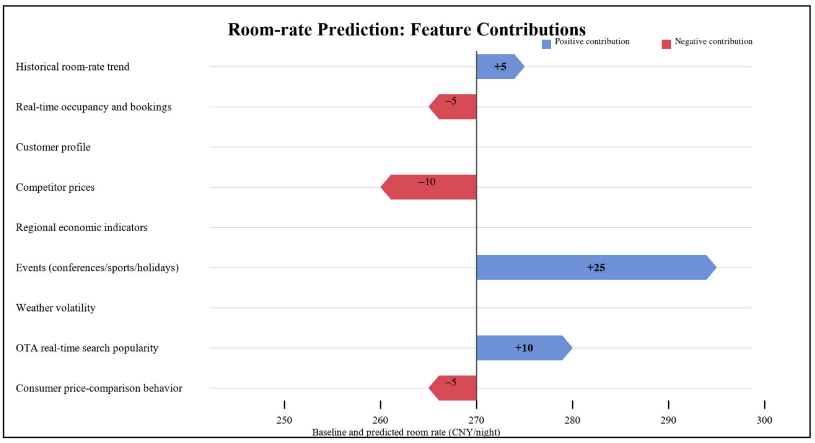


Figure 4. Feature-based explanation.

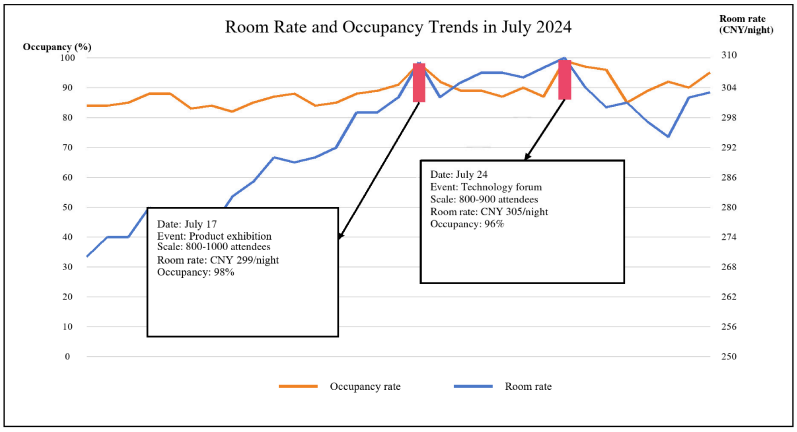


Figure 5. Example-based explanation.

As shown in Figure 6, the model-based explanation uses an interpretable decision tree to represent the decision logic underlying the pricing recommendation. The tree presents a simplified sequence of decision rules based on hotel operational, market, and consumer-related information. For example, the displayed path illustrates how an active event, recent room-rate trends, and competitor prices jointly lead to the recommended room rate. This representation helps practitioners understand the decision logic through which the recommendation is derived.

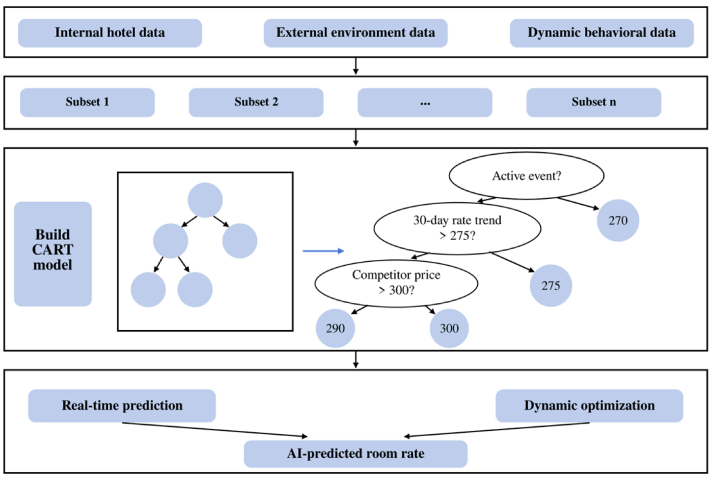


Figure 6. Model-based explanation.

5.2. Expert Sample and Data Collection

The FCM was constructed using judgments from hotel-industry practitioners familiar with pricing and revenue-management decisions. Candidate respondents were recruited through local hotel associations and direct contacts with hotel-management organizations. To ensure that respondents could evaluate the argumentative relationships in the pricing context, participants were required to have direct experience in hotel pricing, revenue management, operations, marketing, or related decision-support activities.

The screening procedure consisted of three stages. First, hotel managers or department heads recommended potential participants whose work involved pricing-related decisions. Second, candidates completed an eligibility form covering their current position, industry experience, pricing responsibilities, and familiarity with AI-enabled decision-support systems. Third, the research team confirmed their eligibility through a brief offline or telephone screening before the formal survey. A total of 160 questionnaires were distributed. After excluding responses with failed attention checks, excessive missing data, abnormal completion patterns, or highly repetitive responses, 105 valid questionnaires were retained. Among the valid respondents, 57.14% were male and 42.86% were female. In terms of hotel-industry experience, 48.57% had one to three years of experience, 29.52% had three to five years, and 21.90% had more than five years. Approximately 71.43% reported prior experience with AI-enabled decision-support tools.

The structural links included in the FCM were specified in advance according to the Toulmin-based framework developed in Section 4. Experts therefore evaluated only the theoretically defined directed relationships rather than all possible pairwise combinations among the six nodes. Six relationships were evaluated: Ground → Warrant, Warrant → Claim, Backing → Warrant, Qualifier → Claim, Rebuttal → Warrant, and Rebuttal → Claim. All remaining entries in the adjacency matrix were fixed at zero because no direct relationship was specified by the theoretical framework.

Before completing the questionnaire, respondents were provided with definitions of the six Toulmin components and a standardized description of the AI-assisted hotel-pricing context. Each evaluation item described one directed relationship and asked respondents to assess its influence strength. For example, the Ground → Warrant item was presented as follows: "In an AI-assisted hotel-pricing recommendation, to what extent does the strength and adequacy of supporting evidence influence the strength of the reasoning that connects the evidence to the pricing recommendation?" Equivalent wording was used for the remaining relationships.

Respondents evaluated each directed relationship using five linguistic categories: strong negative influence, weak negative influence, no influence, weak positive influence, and strong positive influence. These linguistic judgments were represented using triangular fuzzy numbers, as shown in Table 3.

Table 3. Linguistic Scale and Triangular Fuzzy Numbers.

Linguistic Judgment	Triangular Fuzzy Number
Strong negative influence	(−1.00, −0.90, −0.70)
Weak negative influence	(−0.70, −0.50, −0.30)
No influence	(−0.10, 0.00, 0.10)
Weak positive influence	(0.30, 0.50, 0.70)
Strong positive influence	(0.70, 0.90, 1.00)

Each linguistic judgment was represented by a triangular fuzzy number. The fuzzy assessments provided by the experts were aggregated across respondents by averaging the lower, modal, and upper values of the corresponding fuzzy numbers. The aggregated

fuzzy values were then converted into crisp edge weights using centroid defuzzification. The resulting weights range from negative to positive values, with the sign indicating the direction of influence and the absolute value reflecting its strength.

The aggregated relationship weights are reported in Table 4. Warrant exerted the strongest positive influence on Claim (w = 0.81), indicating that the inferential connection between evidence and recommendation plays a central role in determining the justificatory support for the recommendation. Ground had a relatively strong positive effect on Warrant (w = 0.72), while Backing also strengthened Warrant (w = 0.65). Qualifier showed a moderate positive relationship with Claim (w = 0.58). Rebuttal exerted negative effects on both Warrant (w = −0.55) and Claim (w = −0.49).

Table 4. Aggregated Relationship Weights and Dispersion.

Relationship	Mean Weight	SD	IQR
Ground → Warrant	0.72	0.16	0.20
Warrant → Claim	0.81	0.13	0.18
Backing → Warrant	0.65	0.18	0.22
Qualifier → Claim	0.58	0.19	0.25
Rebuttal → Warrant	−0.55	0.20	0.27
Rebuttal → Claim	−0.49	0.21	0.28

The dispersion statistics indicate moderate agreement across the six evaluated relationships. The strongest agreement was observed for Warrant → Claim, whereas judgments concerning the two rebuttal relationships showed somewhat greater variation.

To further assess the consistency of expert judgments, Kendall's coefficient of concordance was calculated across the six relationships. The resulting coefficient was 0.68 ($p < 0.01$), indicating a substantial degree of agreement among respondents. Although the judgments were not completely homogeneous, the level of concordance was sufficient to support aggregation into a common adjacency matrix.

5.3. FCM Initialization

Before rating the explanation stimuli, respondents were instructed to evaluate each explanation in terms of the strength of its argumentative components in the standardized hotel-pricing scenario. The presentation order of the three explanation stimuli was randomized across respondents to reduce potential order effects. Respondents were asked to evaluate each explanation using the following questions—Ground: "To what extent does the explanation provide strong and adequate evidence for the recommendation?". Warrant: "To what extent does the explanation provide a convincing reasoning link between the evidence and the recommendation?". Backing: "To what extent does the explanation provide additional support that strengthens the reasoning?". Qualifier: "To what extent is the recommendation expressed with a strong degree of certainty or confidence?". Rebuttal: "To what extent does the explanation contain strong exceptional or countervailing conditions that challenge or constrain the recommendation?". All items were rated on five-point scales ranging from 1 (very weak) to 5 (very strong).

Mean scores were calculated for each component and normalized to the interval [0, 1] using $(x - 1)/4$. Warrant was rated directly because the perceived strength of the inferential link is a distinct property of the explanation itself, reflecting how convincingly the evidence is connected to the recommendation. Direct rating also allows explanations that primarily clarify the reasoning process, such as model-based explanations, to be distinguished from those that mainly present evidence or analogous cases. The Claim was not rated directly; as the terminal component of the argument structure, its support is derived through the propagation procedure described below.

5.4. Propagation Procedure

The FCM specifies the directional dependencies among the Toulmin components and provides expert-derived relationship weights. In the present framework, these weights are incorporated into aggregation rules that reflect the argumentative roles of the corresponding links. Figure 3 therefore represents the dependency structure, while the equations below specify how these dependencies are operationalized in the evaluation procedure.

The propagation procedure combines the rated components according to the Toulmin structure and the expert-derived weights. First, the effective warrant is calculated as $W^{'} = (1 - w_{BW})W + w_{BW}B + w_{RW}R$. Because the effective warrant W' may theoretically become negative under strong rebutting conditions, it is bounded at zero before the subsequent calculation: $\widetilde{W} = \max\left(0, W^{'}\right)$. This ensures that a non-positive effective warrant is treated as providing no positive inferential support and that the following fractional-power transformation remains well defined.

The core justification index is then calculated as $J = G^{w_{GW}}\widetilde{W}^{w_{WC}}$, after which Qualifier and Rebuttal adjust the index, $S = J + w_{QC}Q + w_{RC}R$, and the final Claim-support value is obtained through the sigmoid transformation $C = \frac{1}{1+e^{-k(S-\theta)}}$. Here, k = 5 and θ = 0.5. The resulting Claim value represents the degree of justificatory support accumulated for the recommendation within the Toulmin–FCM structure. Table 5 reports the component value.

Table 5. Mean (SD) of Rated Components and Derived Claim Score.

Component	Feature-Based	Example-Based	Model-Based
Ground	0.80 (0.12)	0.73 (0.18)	0.69 (0.21)
Backing	0.72 (0.17)	0.78 (0.13)	0.65 (0.26)
Qualifier	0.66 (0.19)	0.61 (0.22)	0.59 (0.17)
Rebuttal	0.30 (0.14)	0.34 (0.19)	0.39 (0.23)
Warrant (rated)	0.62 (0.20)	0.58 (0.20)	0.72 (0.15)
Effective warrant (W′)	0.520	0.523	0.460
Core justification (J)	0.501	0.472	0.408
Claim (output)	0.766	0.689	0.574

5.5. Evaluation Results

The final Claim value represents the degree of justificatory support accumulated for the AI-generated recommendation within the Toulmin–FCM structure. This interpretation follows the argumentative logic of the framework: evidence supports the recommendation through the inferential link, while qualification and countervailing conditions modulate and constrain the resulting support. A higher Claim-support value therefore indicates that the recommendation receives stronger overall argumentative support within the specified structure.

As shown in Table 5, among the three explanation stimuli examined, the feature-based explanation produced the highest Claim-support value (0.766), followed by the example-based explanation (0.689) and the model-based explanation (0.574). The core justification index J shows the same ordering (0.501, 0.472, and 0.408, respectively). Notably, the model-based explanation received the highest direct warrant rating (0.72), reflecting its strength in representing the decision logic; however, its relatively weak ground limited its core justification, because under the multiplicative structure a strong warrant cannot compensate for weak evidence.

These differences reflect the combined configurations of the argumentative components. The feature-based explanation combines relatively strong Ground and Qualifier values with a lower Rebuttal value, producing the strongest overall support for the Claim. The example-based explanation receives relatively strong Backing, which raises its effective

warrant, but this advantage is partly offset by lower Ground and Qualifier values and a somewhat stronger Rebuttal. The model-based explanation exhibits the weakest supporting evidence and the strongest Rebuttal among the three conditions, resulting in the lowest core justification index and Claim-support value despite its strongest Warrant.

A brief sensitivity analysis was conducted to examine the robustness of the results to the modeling parameters. The edge weights were varied by ±10%, and alternative values of the sigmoid parameters were examined. Across these specifications, the ordering of the three explanation conditions remained unchanged, with the feature-based explanation producing the highest Claim-support value, followed by the example-based and model-based explanations. This suggests that the observed ranking is not driven by a narrow choice of parameter values.

5.6. Empirical Validation

A separate online experiment was conducted to examine whether the FCM evaluation was consistent with users' responses to the AI-generated recommendations. Acceptance and trust were selected as external outcomes because stronger justificatory support would be expected to correspond with more favorable user evaluations of the recommendation.

Participants were recruited through a professional online survey platform in China. Eligibility was restricted to adults with work experience and basic familiarity with AI-enabled decision-support applications. A total of 248 participants were initially recruited. After excluding incomplete responses and those failing the attention-check questions, 216 valid responses were retained for analysis. Participants were randomly assigned to one of the three explanation conditions using the random allocation function of the online survey platform, with 72 participants in each group. Each participant was exposed to only one explanation condition to avoid potential carryover effects.

Regarding demographic characteristics, the sample consisted of 112 males (51.9%) and 104 females (48.1%). The participants ranged in age from 22 to 55 years, with an average age of 34.6 years. In terms of educational background, 78.2% of participants held at least a bachelor's degree. Regarding occupational background, participants came from diverse professional fields, including business and management (28.2%), technology-related occupations (21.3%), service industries (18.5%), education and research (12.0%), and other sectors (20.0%).

Since the study examines how users evaluate AI-generated recommendations rather than the decision behavior of hotel professionals specifically, participants were not required to have direct experience in hotel revenue management. However, 62.5% of participants reported prior experience using AI-enabled applications (e.g., intelligent assistants, recommendation systems, or automated decision-support tools), and 31.9% indicated some familiarity with hotel pricing, online booking, or related service-selection decisions. These characteristics suggest that the participants represent general users with working experience and varying degrees of exposure to AI-enabled decision-support systems, which is appropriate for evaluating users' perceptions of AI-generated recommendations in a realistic decision-support context.

The measurement of acceptance and trust was adapted from recent research in AI-assisted recommendations and advice [29,32,43–46]. Acceptance was measured using three items: "I would accept the AI-generated pricing recommendation," "I would follow the AI-generated pricing recommendation," and "I would be willing to adopt the AI-generated pricing recommendation." Trust was measured using three items assessing the credibility, reliability, and trustworthiness of the recommendation: "I consider the AI-generated pricing recommendation credible," "I consider the recommendation reliable," and "I consider the

recommendation trustworthy." Translation and back-translation procedures were used to ensure consistency between the English scale items and the Chinese questionnaire.

The measurement properties of the two constructs were satisfactory. Standardized factor loadings ranged from 0.81 to 0.89 for acceptance and from 0.82 to 0.91 for trust. Cronbach's alpha values were 0.87 and 0.89, respectively. Composite reliability values exceeded 0.85, and average variance extracted values exceeded 0.65 for both constructs. The discriminant validity between acceptance and trust was also acceptable, with an HTMT ratio of 0.76.

As shown in Table 6, the feature-based explanation received the highest acceptance and trust scores, followed by the example-based and model-based explanations. One-way ANOVA indicated significant differences across conditions for acceptance, $F(2, 213) = 22.45$, $p < 0.001$, $\eta^2 = 0.17$, and trust, $F(2, 213) = 16.59$, $p < 0.001$, $\eta^2 = 0.13$. Tukey's HSD tests further showed significant pairwise differences across all three explanation conditions for both acceptance and trust, as reported in Table 7.

Table 6. External Validation Results.

Explanation Form	Acceptance M (SD)	95% CI	Trust M (SD)	95% CI
Feature-based	4.18 (0.62)	[4.03, 4.33]	4.10 (0.64)	[3.95, 4.25]
Example-based	3.84 (0.68)	[3.68, 4.00]	3.78 (0.69)	[3.62, 3.94]
Model-based	3.42 (0.74)	[3.25, 3.59]	3.44 (0.73)	[3.27, 3.61]

Table 7. Tukey HSD Pairwise Comparisons.

Outcome	Comparison	Mean Difference	95% CI
Acceptance	Feature-based vs. Example-based	0.34 **	[0.07, 0.61]
Acceptance	Feature-based vs. Model-based	0.76 ***	[0.49, 1.03]
Acceptance	Example-based vs. Model-based	0.42 **	[0.15, 0.69]
Trust	Feature-based vs. Example-based	0.32 *	[0.05, 0.59]
Trust	Feature-based vs. Model-based	0.66 ***	[0.39, 0.93]
Trust	Example-based vs. Model-based	0.34 **	[0.07, 0.61]

Note: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

The experimental results show a pattern consistent with the Claim-support scores generated by the Toulmin–FCM framework. The feature-based condition received the highest acceptance and trust evaluations, followed by the example-based and model-based conditions. This correspondence provides external evidence for the application of the proposed framework across the three explanation stimuli.

6. Discussion and Implications

6.1. Discussion

This study develops an argument-based approach to evaluating explanations of AI-generated recommendations. Instead of treating explanation evaluation as a collection of independent attributes or user responses, the framework asks how the information contained in an explanation functions together to support, qualify, or constrain a recommendation. The Toulmin model provides the underlying argumentative structure, while the FCM operationalizes the dependencies among these components. In this sense, the main value of the framework lies in shifting attention from whether an explanation is generally perceived as clear, useful, or trustworthy to how strongly the recommendation is justified by the explanation itself.

The hotel-pricing application illustrates how different explanation forms can generate different argumentative profiles. Within the three explanation stimuli examined, the

feature-based explanation achieved the highest Claim support, followed by the example-based and model-based explanations. These differences arise from the configurations of the underlying argumentative components. The feature-based explanation benefits from stronger Ground and Qualifier values and a lower Rebuttal value, while the example-based explanation receives comparatively stronger Backing. The model-based explanation produces weaker overall Claim support under the specific configuration examined. Another noteworthy finding is that the Rebuttal node has the lowest activation values across all three explanation types, indicating that the countervailing conditions represented in these explanation stimuli are relatively weak. This finding highlights the importance of adequately considering system limitations, exceptions, and failure conditions in AI explanations. Recent XAI research identifies knowledge limits, falsifiability, and the communication of uncertainty as central challenges for the field [6,7]. For e-commerce decision support, such countervailing conditions are important because managers need to know when a recommendation may be unreliable, such as during abnormal demand shocks, data sparsity, or sudden competitor reactions. Insufficient consideration of these conditions may encourage overreliance and weaken critical evaluation [33]. The empirical validation shows a corresponding ordering in users' acceptance and trust. This correspondence provides additional support for examining explanations through their argumentative configurations. This pattern also illustrates the non-compensatory logic of the framework: a strong warrant cannot offset weak evidence in the core justification of a recommendation.

6.2. Theoretical Implications

The first theoretical contribution of this study is to introduce argument-based evaluation as a complementary perspective on XAI evaluation. Existing approaches commonly assess explanations through attributes such as clarity, usefulness, completeness, or satisfaction, or through downstream outcomes such as trust and acceptance [11,27]. These measures are valuable for understanding how users perceive and respond to explanations, but they provide less direct insight into how an explanation justifies the recommendation itself. The present framework shifts attention to the internal justificatory structure of an explanation by treating the AI-generated recommendation as a Claim supported and constrained by functionally related argumentative components.

The second contribution lies in operationalizing this perspective through FCM. Toulmin's model provides a conceptual account of how evidence, inference, support, qualification, and counterarguments contribute to a practical conclusion, but it does not specify how their combined influence can be quantitatively represented [17]. FCM provides a tractable way to model these directional relationships and propagate their effects through the argument structure. This preserves the dependency among components and distinguishes the framework from additive evaluation approaches in which each criterion contributes independently to an overall score.

The third contribution is to provide a common analytical structure for comparing technically different forms of AI explanation. Feature-based, example-based, and model-based explanations communicate different kinds of information, making direct comparison difficult when evaluation relies only on format-specific attributes [42]. The Toulmin-based framework instead evaluates the justificatory function performed by the information contained in each explanation. This makes it possible to compare heterogeneous explanation forms in terms of how they support the recommendation, while retaining the distinct roles of evidence, inferential logic, supplementary support, qualification, and countervailing conditions.

6.3. Practical Implications

The framework offers practical value for the design and evaluation of AI-supported decision systems. Developers and managers can use the Toulmin structure as a diagnostic lens to assess whether an explanation provides sufficient evidence, establishes a clear inferential connection to the recommendation, communicates appropriate qualifications, and identifies relevant countervailing conditions. This is particularly important in commercial decision settings such as dynamic pricing, where recommendations often need to be scrutinized before implementation. At the same time, providing richer explanatory information may increase cognitive load. Interface designers should therefore balance informational completeness with usability. Progressive disclosure offers a useful design strategy: core information can be presented first, while additional qualifications, exceptions, and risk-related information are made available when users require deeper scrutiny. Such a layered approach can help managers access critical information without being overwhelmed, while still supporting informed and accountable decision-making.

6.4. Limitations and Future Research

This study has several limitations. First, the empirical application is limited to the hotel-pricing context. Although hotel pricing represents an important AI-supported decision-making scenario, its characteristics, such as demand volatility, occupancy constraints, and revenue-management practices, may differ from other e-commerce contexts. Therefore, the findings should not be generalized directly to all e-commerce decision-support settings. Future research could further examine the applicability of the proposed Toulmin–FCM framework across different domains.

Second, the experimental comparison involves three explanation forms that differ not only in argumentative emphasis but also in presentation format and information organization. In particular, the model-based condition used a relatively complex decision-tree representation, whereas the other conditions presented information in different visual forms. Therefore, the observed differences may partly reflect variations in cognitive processing or visualization characteristics. Future research could further isolate the effects of argumentative structures by manipulating individual Toulmin components while keeping presentation formats consistent. Future studies could also incorporate manipulation checks to examine whether users perceive the intended argumentative roles of different explanation forms.

In addition, the experimental stimuli were constructed representations of explanation forms rather than outputs from a validated operational pricing model. Future research can extend the framework by comparing explanation designs while holding presentation format constant, by systematically varying the presence of Toulmin components, and by applying the framework to explanations generated from deployed AI systems. Such studies could also examine technical fidelity, model accuracy, and calibrated reliance alongside argumentative support.

Finally, the present study focuses on operationalizing the Toulmin–FCM framework rather than systematically comparing alternative aggregation procedures. Future research could examine the framework alongside simpler additive models or other expert-based weighting approaches to further assess the contribution of structural dependencies and different aggregation rules.

Author Contributions: Conceptualization, L.N.; methodology, L.N.; formal analysis, L.N.; investigation, S.F.; writing—original draft preparation, L.N.; writing—review and editing, S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Program of Shanxi Province (202403021211230).

Institutional Review Board Statement: The survey and experimental procedures were reviewed and approved by the Research Ethics Committee of the School of Economics and Management, Taiyuan University of Technology (No. 0020260350024, approved on 14 June 2026).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Guha, A.; Grewal, D.; Kopalle, P.K.; Haenlein, M.; Schneider, M.J.; Jung, H.; Moustafa, R.; Hegde, D.R.; Hawkins, G. How artificial intelligence will affect the future of retailing. *J. Retail.* **2021**, *97*, 28–41. [CrossRef]
2. Huang, M.-H.; Rust, R.T. A strategic framework for artificial intelligence in marketing. *J. Acad. Mark. Sci.* **2021**, *49*, 30–50. [CrossRef]
3. Lee, S.W.; Sharma, A. AI versus human-generated hotel pricing: Unpacking consumer perceptions of trustworthiness and fairness. *Tour. Econ.* **2026**, *32*, 825–844. [CrossRef]
4. Fasihi, Z.; Rezaei-Pandari, M.; Torkaman, F.; Razzazi, M. The AI-driven evolution of dynamic pricing: A semi-systematic review and novel hierarchical classification. *Appl. Oper. Anal.* **2026**, *2*, 1–31. [CrossRef]
5. Rai, A. Explainable AI: From black box to glass box. *J. Acad. Mark. Sci.* **2020**, *48*, 137–141. [CrossRef]
6. Longo, L.; Brcic, M.; Cabitza, F.; Choi, J.; Confalonieri, R.; Del Ser, J.; Guidotti, R.; Hayashi, Y.; Herrera, F.; Holzinger, A.; et al. Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Inf. Fusion* **2024**, *106*, 102301. [CrossRef]
7. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Inf. Fusion* **2023**, *99*, 101805. [CrossRef]
8. Nunes, I.; Jannach, D. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Model. User-Adapt. Interact.* **2017**, *27*, 393–444. [CrossRef]
9. Gedikli, F.; Jannach, D.; Ge, M. How should I explain? A comparison of different explanation types for recommender systems. *Int. J. Hum.-Comput. Stud.* **2014**, *72*, 367–382. [CrossRef]
10. Sahoh, B.; Choksuriwong, A. The role of explainable artificial intelligence in high-stakes decision-making systems: A systematic review. *J. Ambient Intell. Humaniz. Comput.* **2023**, *14*, 7827–7843. [CrossRef]
11. Nauta, M.; Trienes, J.; Pathak, S.; Nguyen, E.; Peters, M.; Schmitt, Y.; Schlötterer, J.; van Keulen, M.; Seifert, C. From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable AI. *ACM Comput. Surv.* **2023**, *55*, 295. [CrossRef]
12. Saeed, W.; Omlin, C. Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowl.-Based Syst.* **2023**, *263*, 110273. [CrossRef]
13. Vassiliades, A.; Bassiliades, N.; Patkos, T. Argumentation and explainable artificial intelligence: A survey. *Knowl. Eng. Rev.* **2021**, *36*, e5. [CrossRef]
14. Hahn, U.; Tešić, M. Argument and explanation. *Philos. Trans. R. Soc. A* **2023**, *381*, 20220043. [CrossRef]
15. Rago, A.; Cocarascu, O.; Bechlivanidis, C.; Lagnado, D.; Toni, F. Argumentative explanations for interactive recommendations. *Artif. Intell.* **2021**, *296*, 103506. [CrossRef]
16. Kotonya, N.; Toni, F. Towards a framework for evaluating explanations in automated fact verification. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italy, 20–25 May 2024*; ELRA and ICCL: Paris, France, 2024; pp. 16364–16377. Available online: https://aclanthology.org/2024.lrec-main.1422/ (accessed on 2 September 2026).
17. Toulmin, S.E. *The Uses of Argument*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2003. [CrossRef]
18. Kosko, B. Fuzzy cognitive maps. *Int. J. Man-Mach. Stud.* **1986**, *24*, 65–75. [CrossRef]
19. Papageorgiou, E.I. *Fuzzy Cognitive Maps for Applied Sciences and Engineering: From Fundamentals to Extensions and Learning Algorithms*; Springer: Berlin/Heidelberg, Germany, 2014. [CrossRef]

20. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
21. Adadi, A.; Berrada, M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [CrossRef]
22. Camilleri, M.A. Opening the black box: Operational principles, tools and frameworks that advance explainable artificial intelligence (XAI) models. *Technol. Forecast. Soc. Change* **2026**, *229*, 124710. [CrossRef]
23. Carvalho, D.V.; Pereira, E.M.; Cardoso, J.S. Machine learning interpretability: A survey on methods and metrics. *Electronics* **2019**, *8*, 832. [CrossRef]
24. Lipton, Z.C. The mythos of model interpretability. *Commun. ACM* **2018**, *61*, 36–43. [CrossRef]
25. Gunning, D.; Aha, D. DARPA's explainable artificial intelligence (XAI) program. *AI Mag.* **2019**, *40*, 44–58. [CrossRef]
26. Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A survey of methods for explaining black box models. *ACM Comput. Surv.* **2018**, *51*, 93. [CrossRef]
27. Mohseni, S.; Zarei, N.; Ragan, E.D. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Trans. Interact. Intell. Syst.* **2021**, *11*, 24. [CrossRef]
28. Hamm, P.; Klesel, M.; Coberger, P.; Wittmann, H.F. Explanation matters: An experimental study on explainable AI. *Electron. Mark.* **2023**, *33*, 17. [CrossRef]
29. Kahr, P.K.; Rooks, G.; Willemsen, M.C.; Snijders, C.C.P. Understanding trust and reliance development in AI advice: Assessing model accuracy, model explanations, and experiences from previous interactions. *ACM Trans. Interact. Intell. Syst.* **2024**, *14*, 29. [CrossRef]
30. Wang, X.; Yin, M. Effects of explanations in AI-assisted decision making: Principles and comparisons. *ACM Trans. Interact. Intell. Syst.* **2022**, *12*, 17. [CrossRef]
31. Rezaeian, O.; Bayrak, A.E.; Asan, O. Explainability and AI confidence in clinical decision support systems: Effects on trust, diagnostic performance, and cognitive load in breast cancer care. *Int. J. Hum.-Comput. Interact.* **2026**, *42*, 4477–4497. [CrossRef]
32. Glikson, E.; Woolley, A.W. Human trust in artificial intelligence: Review of empirical research. *Acad. Manag. Ann.* **2020**, *14*, 627–660. [CrossRef]
33. Romeo, G.; Conti, D. Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. *AI Soc.* **2026**, *41*, 259–278. [CrossRef]
34. Lertvittayakumjorn, P.; Toni, F. Argumentative explanations for pattern-based text classifiers. *Argum. Comput.* **2023**, *14*, 163–234. [CrossRef]
35. Hernandez-Bocanegra, D.C.; Ziegler, J. Explaining review-based recommendations: Effects of profile transparency, presentation style and user characteristics. *i-com* **2020**, *19*, 181–200. [CrossRef]
36. Papageorgiou, E.I.; Salmeron, J.L. A review of fuzzy cognitive maps research during the last decade. *IEEE Trans. Fuzzy Syst.* **2013**, *21*, 66–79. [CrossRef]
37. Felix, G.; Nápoles, G.; Falcon, R.; Froelich, W.; Vanhoof, K.; Bello, R. A review on methods and software for fuzzy cognitive maps. *Artif. Intell. Rev.* **2019**, *52*, 1707–1737. [CrossRef]
38. Apostolopoulos, I.D.; Groumpos, P.P. Fuzzy cognitive maps: Their role in explainable artificial intelligence. *Appl. Sci.* **2023**, *13*, 3412. [CrossRef]
39. Burkart, N.; Huber, M.F. A survey on the explainability of supervised machine learning. *J. Artif. Intell. Res.* **2021**, *70*, 245–317. [CrossRef]
40. Perlmutter, M.; Gifford, R.; Krening, S. Impact of example-based XAI for neural networks on trust, understanding, and performance. *Int. J. Hum.-Comput. Stud.* **2024**, *188*, 103277. [CrossRef]
41. Sagi, O.; Rokach, L. Explainable decision forest: Transforming a decision forest into an interpretable tree. *Inf. Fusion* **2020**, *61*, 124–138. [CrossRef]
42. Kim, D.; Song, Y.; Kim, S.; Lee, S.; Wu, Y.; Shin, J.; Lee, D. How should the results of artificial intelligence be explained to users? Research on consumer preferences in user-centered explainable artificial intelligence. *Technol. Forecast. Soc. Change* **2023**, *188*, 122343. [CrossRef]
43. Chua, A.Y.K.; Pal, A.; Banerjee, S. AI-enabled investment advice: Will users buy it? *Comput. Hum. Behav.* **2023**, *138*, 107481. [CrossRef]
44. Bashkirova, A.; Krpan, D. Confirmation bias in AI-assisted decision-making: AI triage recommendations congruent with expert judgments increase psychologist trust and recommendation acceptance. *Comput. Hum. Behav. Artif. Hum.* **2024**, *2*, 100066. [CrossRef]

45. Fink, L.; Newman, L.; Haran, U. Let me decide: Increasing user autonomy increases recommendation acceptance. *Comput. Hum. Behav.* **2024**, *156*, 108244. [CrossRef]
46. Qin, H.; Zhu, Y.; Jiang, Y.; Luo, S.; Huang, C. Examining the impact of personalization and carefulness in AI-generated health advice: Trust, adoption, and insights in online healthcare consultations experiments. *Technol. Soc.* **2024**, *79*, 102726. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.