

Type of
Contribution:

Research Paper
Review Paper
Case Study

ENERGY: JURNAL ILMIAH
ILMU-ILMU TEKNIK
Special Issue 2025 pp 400-414
DOI: 10.51747/energy.si2025.255



E-ISSN: 2962-2565

This article
contributes to:



An Epistemological Approach to Explainable Automated Assessment of Open Concept Map Propositions Using SHAP

Mega Satya Ciptaningrum^{1*}, Syaad Patmanthara¹, Didik Dwi Prasetya¹

¹ Department of Electrical Engineering and Informatics, Universitas Negeri Malang, 65145, Indonesia

*mega.satya.250534923512@students.um.ac.id

Abstract

Concept mapping is widely recognized as an effective method for supporting meaningful learning and critical thinking because it allows teachers to assess students' underlying knowledge structures. However, evaluating concept maps and providing feedback remain challenging, as these processes are time-consuming, increase teachers' workload, and can reduce instructional efficiency. To address this issue, this study applies Transformer-based architectures, which rely on large-scale pre-training and task-specific fine-tuning, to develop an automated assessment system for concept maps. In addition, Explainable Artificial Intelligence (XAI) is integrated through the SHAP (SHapley Additive exPlanations) framework to generate interpretable explanations of the model's scoring decisions. Using Transformer models such as BERT and DeBERTa, SHAP values are computed at the token level to show how individual words within each proposition contribute to the final score. The results indicate that tokens with positive SHAP values increase scores in line with correct rubric indicators, whereas negative values reduce them. Tokens that consistently show positive contributions in high-scoring outputs reflect stable and faithful model reasoning. Overall, the findings demonstrate that combining Transformer-based assessment with SHAP explanations improves epistemic transparency by aligning the model's internal reasoning with expert evaluation criteria, thereby supporting more reliable, interpretable, and trustworthy automated feedback in concept mapping-based learning.

Keywords: DeBERTa; SHAP; Epistemology; Evidence Chain; Faithfulness; Fidelity

Article Info

Submitted:

2025-10-26

Revised:

2025-12-22

Accepted:

2025-12-25

Published:

2025-12-30



This work is
licensed under a
Creative
Commons
Attribution-
NonCommercial
4.0 International
License

Publisher

Universitas
Panca Marga

1. Introduction

Transformer-based models are claimed to enhance the performance of automated assessment systems due to their ability to capture deep contextual dependencies, enabling a more comprehensive understanding of semantic meaning [1]. In particular, the BERT model and its derivatives possess an architecture that supports evaluation from two main aspects: word meaning and inter-word relationships [2]. The Transformer architecture employs a pre-trained mechanism, in which a large-scale model is initially trained on extensive text corpora and subsequently fine-tuned for specific downstream tasks, such as text classification, that can later be applied to automated grading systems [3].

From the perspective of philosophy of electrical and computer engineering, an automated assessment system is not merely a computational artifact, but an epistemic-technical system that produces, justifies, and communicates knowledge claims. In this study, the Transformer–DeBERTa model is positioned as an epistemic agent whose outputs (scores) constitute epistemic claims grounded in learned representations and expert-defined rubrics. Therefore, the central philosophical concern is not only predictive accuracy, but epistemic justification, transparency, and alignment between internal model reasoning and external human evaluation standards.

One example of a learning task that can benefit from automated assessment is concept mapping, a pedagogical method known to promote meaningful learning [3] and critical thinking, as it enables teachers to evaluate the knowledge structures constructed by students [5]. However, a persistent challenge lies in the evaluation and feedback process, which adds a significant burden on teachers due to its time-consuming nature and potential to hinder instructional efficiency [3]. [6] Conducted an evaluation of propositional quality, where propositions—representing students' connections between two concepts through linking words—were manually assessed using rubrics to produce proposition scores. Within the domain of Natural Language Processing (NLP), Large Language Models (LLMs) that adopt instance-based methods can automatically score student responses by learning the lexical characteristics of correct answers, storing them in memory, and then classifying new responses based on previously learned patterns to determine their scores [8]. Recent innovations have further leveraged the Transformer-based architecture with pre-trained models and fine-tuning techniques, making them increasingly suitable for automated text evaluation tasks [9].

Nevertheless, a major challenge of Transformer models in semantic evaluation of student responses is their black-box nature, which limits transparency, fairness, and adaptability. Consequently, educators have limited understanding of the

underlying factors influencing the model's predictions [10]. To address this issue, a trustworthy explanatory model is required, particularly through advancements in Explainable Artificial Intelligence (XAI). Among the prominent XAI methods, SHAP (SHapley Additive exPlanations) provides explanations based on the contribution values of textual features (tokens) derived from BERT-based processes. Each feature thus contributes to the reliability and interpretability of the model's explanations [4].

From an epistemological perspective, transparency in AI-based assessment and explanation—such as that achieved through SHAP—is crucial for understanding how AI systems acquire and justify knowledge. This understanding fosters human trust (from both students and teachers) by revealing the reasoning process of the algorithm, rather than merely accepting its final outputs [7]. Given the limitations of teachers in evaluating student responses in open-ended concept map propositions, along with the rapid advancements of LLM-based automatic assessment models, this study proposes an automated scoring model for open concept map propositions based on Transformer–BERT architecture, enhanced through Explainable AI (SHAP) to improve interpretability and epistemic reliability. The objectives of this study are to: (1) Epistemically validate the SHAP explanation model for student responses in open concept map propositions based on a predefined scoring rubric; and (2) Evaluate the levels of faithfulness and fidelity to rubric in SHAP-based explanations, ensuring that the AI model aligns both internally with its reasoning process and externally with expert assessment criteria.

2. Methods

The initial stage of implementation involves the preparation of data and model configuration. The data used in this experiment consist of student responses in the form of open-ended concept map propositions within the topic of Database Relations. The process of tokenization and prediction generation is carried out using the DeBERTa model, while the test data are evaluated using SHAP to provide feedback in the form of explanatory insights regarding each student's proposition. An overview of the implementation workflow is illustrated in [Figure 1](#).



Figure 1. Implementation Workflow

This study was conducted using a dataset tested within the Google Colab environment, utilizing the Python programming language and executed on a 12th Gen Intel(R) Core(TM) i5-1250P processor with 8 GB of RAM.

2.1 Dataset

The dataset used in this study consists of student responses in the form of concept map propositions from assignments on Database Relations. Each proposition was evaluated by a single expert (teacher) and assigned a score ranging from 0 to 3, which serves as the ground truth derived from the expert's assessment rubric.

The dataset comprises 691 propositions, which have been preprocessed to remove unnecessary characters; hence, text cleaning was not required. The prepared data were divided into three subsets: 80% for training, 10% for validation, and 10% for testing, resulting in 552 training samples, 69 validation samples, and 70 testing samples. Subsequently, the data were used for fine-tuning the model, followed by the application of SHAP-based evaluation to construct the explanation model.

It should be noted that the ground truth in this study is derived from a single expert. From an epistemological standpoint, this represents a form of localized epistemic authority rather than intersubjective consensus. While sufficient for validating the internal epistemic coherence of the proposed model, future work should incorporate multiple experts to strengthen epistemic robustness and reduce potential normative bias.

The research procedure consists of four main stages: (1) Data and Model Preparation, (2) Student Answer Score Prediction, (3) SHAP-based Explanation, and (4) Visualization and Interpretation.

For the automatic scoring of student responses, the DeBERTa model was employed and trained to predict scores corresponding to expert (teacher) evaluations. In addition to score prediction, the system also provides explainability for its predictions using the SHAP method, allowing the model's decision-making process to be interpretable by humans. This feature further supports epistemic analysis of the AI-generated explanations.

2.2 Fine-Tuned Transformer–DeBERTa

One of the most influential Transformer-based models, BERT (Bidirectional Encoder Representations from Transformers) [9], has revolutionized Natural Language Processing (NLP) through its ability to capture contextual word representations, thereby improving the accuracy of text comprehension [7]. Several variants of BERT, such as SciBERT, BioBERT, and other domain-specific language models, have been developed for processing technical or structured content, including experimental analysis and methodological explanations. These specialized models are trained on domain-relevant datasets to enhance performance in specific contexts [9].

Another advancement, DeBERTa (Decoding-enhanced BERT with Disentangled Attention), introduces a dual-stream architecture in which one stream

focuses on self-attention, while the other employs a decoupled attention mechanism. This design enables the model to capture complex relationships and semantic nuances within text more effectively—first, by disentangling content embeddings (representations of word meaning), and second, by separating position embeddings (representations of positional and relational information among words) [10]. The DeBERTa model, available in the Hugging Face Transformers library, can be directly implemented for fine-tuning and prediction tasks. **Figure 2.** presents the pseudo-code of the fine-tuning process, which consists of two main stages: tokenization and score prediction using the DeBERTa architecture.

```

Procedure:
  Load pretrained DeBERTa model and tokenizer:
    model_name ← "microsoft/deberta-base"
    model ← AutoModelForSequenceClassification(num_labels=1,
problem_type="regression")
    tokenizer ← AutoTokenizer.from_pretrained(model_name)

  Define max sequence length (max_len = 128)

  Tokenize text and prepare tensor dataset:
    Encode text using tokenizer
    Convert scores to tensor
    Build custom Dataset class (ProposisiDataset)

Output:
  train_dataset, val_dataset, test_dataset ready for training

```

Figure 2. Pseudo-code of the Fine-Tuned Transformer–DeBERTa Model

2.3 SHAP Explainability

One of the main advantages of Transformer-based models such as BERT is their ability to produce rich, contextualized representations, which naturally support token-level attribution methods required in Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME [9]. SHAP and LIME have emerged as leading XAI approaches, replacing earlier model-specific methods such as Decision Trees, Linear Regression, and Bayesian Networks [14]. SHAP is grounded in Shapley value theory and enables an additive decomposition of model predictions at the token level, allowing the contribution of each text component to the output to be quantified explicitly [9]. These values reflect the importance of each feature by considering all possible feature subsets and their marginal effects on the model's prediction. The SHAP value of a given feature represents its contribution or significance in producing the model output and is computed using Eq. (1). In this study, SHAP is implemented via the `shap.Explainer()` class, which provides native support for Transformer pipelines from Hugging Face.

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]$$

The SHAP value (ϕ_i) for a given feature i is calculated based on the contribution of that feature to the model's prediction, considering all possible combinations of other features.

Formally, the SHAP value represents the marginal contribution of feature i across all subsets S of the feature set F that do not include i :

ϕ_i (SHAP value) denotes the contribution of feature i to the model prediction.

S is a subset of the full set of features F , excluding feature i .

x_s represents the input feature values within subset S .

$f_{S \cup \{i\}}(x_{S \cup \{i\}})$ is the model's prediction trained with feature i , using the subset S plus feature i . $f_S(x_S)$ is the model's prediction trained without feature i , using only the subset

The difference in predictions between these two cases,

$$f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) = f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$$

2.4 Epistemological Analysis

In classical epistemology, a belief is considered knowledge when it is justified through adequate evidence. Within the context of automated assessment, SHAP values function as epistemic justifiers that bridge raw model predictions and human-interpretable reasoning. The proposed evidence chain—consisting of features (tokens), SHAP values (evidence), inferred claims, and final scores—operates analogously to a justificatory structure in epistemology. This structure ensures that each evaluative decision can be traced back to explicit linguistic evidence, thereby satisfying the requirement of epistemic justification in technical systems.

The need for transparency in AI-based automated assessment rests on two fundamental grounds: practical and normative. The practical reason concerns the necessity for clear communication about how the system produces its evaluations from an algorithmic perspective, whereas the normative reason pertains to the responsibility of providing both the scores and their accompanying explanations [7]. A rubric-based assessment model serves as a guideline for the system to evaluate the essential aspects of a student's response, thereby enhancing the interpretability and transparency of the overall assessment process [8].

Epistemological analysis can be achieved through the application of SHAP-based evaluation, as the overall significance of each feature is computed by averaging the absolute SHAP values for that feature across all predictions. The impact of each feature is determined by analyzing its contribution to the model's output based on its SHAP value. The greater the magnitude of a feature's SHAP value, the greater its influence on the prediction result. Accordingly, the interpretation of SHAP values based on feature (token) attribution is as follows []:

- a) Positive SHAP value ($\phi_i > 0$): The feature contributes positively to the prediction, increasing the predicted score.
- b) Negative SHAP value ($\phi_i < 0$): The feature contributes negatively to the prediction, decreasing the predicted score.
- c) SHAP value ($\phi_i = 0$): The feature has negligible or no significant contribution to the prediction.

To clarify how linguistic features are transformed into epistemically justified assessment outcomes, this study introduces an epistemic workflow that explicates the transition from token-level representations to evidence, epistemic claims, and final scores.

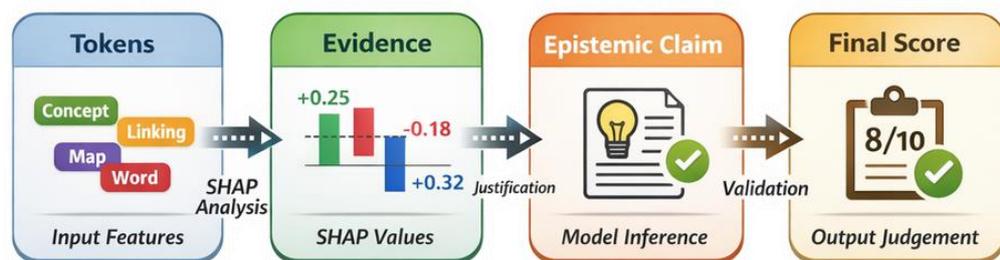


Figure 3. Epistemic Workflow of SHAP-Based Automated Assessment

Figure 3. illustrates the epistemic workflow of the proposed automated assessment system. Tokens extracted by the Transformer–DeBERTa model serve as epistemic features, which are evaluated through SHAP values as evidence. These evidence elements support epistemic claims regarding conceptual correctness, which are subsequently validated against the expert-defined rubric to produce the final assessment score.

3. Results and Discussion

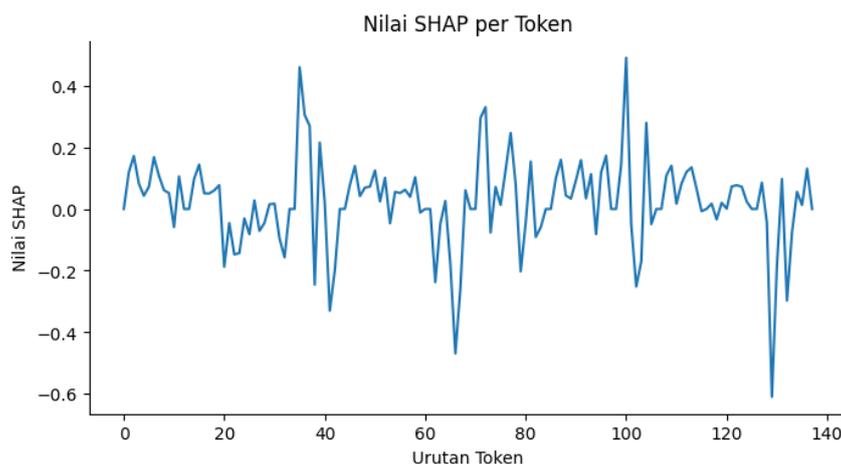
The implementation of the fine-tuned Transformer–DeBERTa model produced a set of tokens along with their predicted scores, which were compared against the ground truth derived from the expert rubric. Each token was subsequently evaluated using the SHAP method to determine its contribution value within the prediction process. A sample of the processed propositional dataset and the corresponding SHAP values for each token is presented in **Table 1**.

Although SHAP values are computed quantitatively, their interpretation is fundamentally qualitative and epistemic in nature. A positive SHAP value does not merely increase a numerical score, but represents linguistic evidence supporting a correct conceptual understanding as defined by the rubric. Conversely, negative SHAP values indicate epistemic tension or conceptual error within the student’s proposition.

Tabel 1. Sample data of SHAP values for each token in a proposition

Text	Token	SHAP_VALUE	Interpretation
Relation must be with integer column		0	Netral
Relation must be with integer column	rel	0,127181984	Increase score (correct indicator)
Relation must be with integer column	asi	-0,008977808	Netral
Relation must be with integer column	har	-0,305975117	Lower score (indicator is wrong)
Relation must be with integer column	us	0,062921144	Netral
Relation must be with integer column	den	-0,011648469	Netral
Relation must be with integer column	gan	-0,262668259	Lower score (indicator is wrong)
Relation must be with integer column	k	0,108532347	Increase score (correct indicator)
Relation must be with integer column	ol	0,057242624	Netral
Relation must be with integer column	om	-0,021399319	Netral
Relation must be with integer column	integer	-0,155591875	Lower score (indicator is wrong)
Relation must be with integer column	Token	0	Netral

Tabel 1. presents the SHAP (SHapley Additive exPlanations) values for each token within a proposition. A positive SHAP value indicates a positive contribution that increases the predicted score toward the correct indicator, whereas a negative SHAP value reflects a negative contribution, lowering the predicted score.

**Figure 4.** Token-Level Distribution of SHAP Values

The range of SHAP values extends from -1 (decreasing the score), 0 (neutral), to $+1$ (increasing the score). The graph in **Figure 4.** illustrates the distribution of SHAP values across all tokens generated by the DeBERTa model.

3.1 Epistemic Structure Analysis of the Model

The computational layer of the implemented model employs DeBERTa to represent semantic meaning and subsequently generate predicted scores. Each token

produced by the DeBERTa model is associated with a SHAP value that contributes to the explanation or justification of knowledge. DeBERTa functions as an epistemic agent that generates epistemic claims (predicted scores) based on conceptual references within the text—specifically, the expert scoring rubric—which serves as the ground truth.

Tabel 2. produced by the SHAP-based model

Token	SHAP_VALUE	Interpetation	SHAP Score
ist	0,459563583	Improve score (correct indicator)	2,64
il	0,305133373	Improve score (correct indicator)	2,64
ah	0,269202948	Improve score (correct indicator)	2,64
ad	-0,245501749	Lower score (indicator is wrong)	2,64
al	0,215181924	Increase score (correct indicator)	2,64
ah	0,020390257	Netral	2,64
foreign	-0,329763457	Lowering the score (incorrect indicator)	2,64
key	-0,199610874	Lowering the score (incorrect indicator)	2,64

Each proposition (student response), as a linguistic component, is decomposed into tokens that embody the model’s empirical experience derived from training. The content embeddings represent the semantic meaning of words, while the position embeddings encode positional and relational knowledge among words in a sentence. The content embeddings correspond to conceptual knowledge units—“what does it mean?”—whereas the position embeddings correspond to structural epistemic units—“how is the meaning related?”. These two representations together support the explainability analysis using SHAP.

Tabel 3. Summary of the evidence chain in the epistemological aspect of student answer assessment

Stage	Description	Example in the Context of DeBERTa-Proposition
Feature	The semantic unit recognized by the model (token resulting from tokenization).	Tokens such as “primary,” “key,” “foreign,” “relation,” “unique.”
Evidence	Positive or negative SHAP values indicating the contribution of each feature to the model’s evaluation.	primary (+0.24) → evidence supporting a higher score; foreign (-0.18) → evidence weakening the student’s claim.
Claim	The epistemic concept inferred by the model based on feature patterns.	The student understands the principle that the Primary Key is unique per table → epistemic claim considered correct.
Score	The evaluative representation (expert ground truth) reconstructed by the model.	The model predicts a score between 0–3 based on positive semantic evidence.

The SHAP value assigned to each token indicates the weight or contribution of that token to the model’s decision. A positive SHAP value represents linguistic

evidence supporting a correct epistemic claim, whereas a negative SHAP value indicates linguistic evidence associated with misconceptions. A higher number of positive SHAP values increases the indicator of epistemic correctness. **Tabel 2** presents the relationship between features (tokens), SHAP values (as evidence), corresponding epistemic claims or interpretations, and the resulting scores produced by the SHAP-based model. The assessed proposition states is “istilah adalah foreign key”, which received an expert score of 3. The integration of DeBERTa and SHAP demonstrates an evidence chain in epistemological analysis for evaluating student responses (propositions).

The explanation of the evidence chain in the epistemological aspect of student responses can be observed in **Tabel 3**.

3.2 Analysis of Faithfulness and Fidelity to Rubric

The faithfulness aspect is evaluated based on the consistency of SHAP token values that support the model’s decision explanations. Tokens with consistently positive SHAP values, or those that frequently appear in high-scoring predictions, indicate a higher level of trustworthiness (faithfulness) in the explanation.

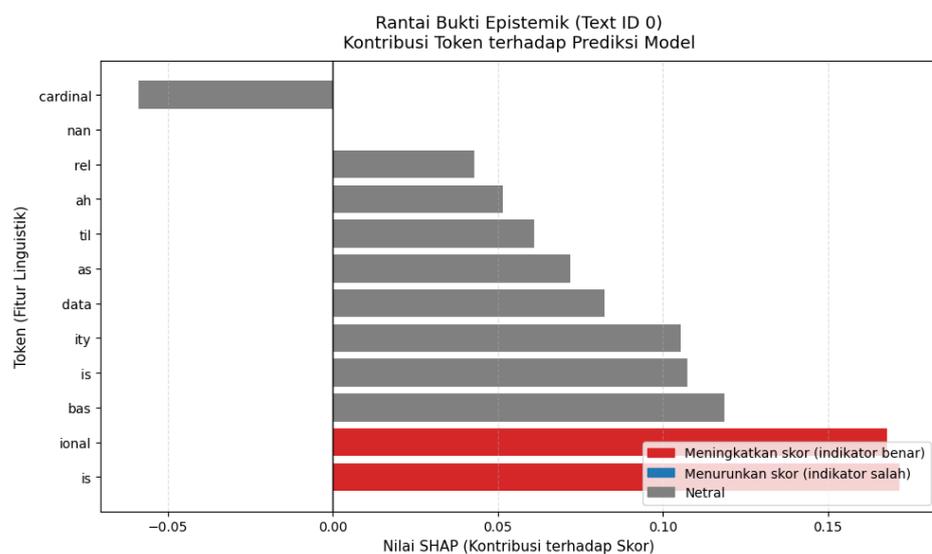


Figure 5. Relationship Between SHAP Values and Score Contribution

In terms of fidelity to the expert scoring rubric (in this case, a score range of 0–3), the model must evaluate based on the same epistemic principles as human experts, rather than relying solely on surface-level statistical patterns. This is evidenced by positive SHAP values reinforcing the correctness of proposition scores.

The SHAP analysis results indicate that the DeBERTa model develops epistemic representations that align with the expert rubric. Positive SHAP values for tokens such as “primary” and “unique” demonstrate the model’s faithfulness to transparent internal processes, while the positive correlation between rubric-related

tokens and expert-assigned scores reflects the model's fidelity to the epistemic structure of human evaluation.

In the epistemology of automated assessment, two key principles ensure that the model remains epistemically reliable:

Faithfulness (Alignment of Explanation with Model Mechanism)

In the faithfulness aspect, the explanation generated by SHAP should accurately reflect the internal reasoning process of the model. This means that the explanation truly represents how the model arrives at its decisions, as evidenced by the previously described evidence chain connecting the feature, evidence, claim, and score. **Tabel 4.** presents the indicators of faithfulness in the SHAP-based model explanation.

Tabel 4. Indicators of Faithfulness in SHAP Model Explanation

Aspect	Description	Indicator
Semantic Coherence	The relationship between key tokens is consistent with the rubric concept.	The word "primary" increases the score.
Causal Consistency	Removing a key token decreases the predicted score.	The token "key" is removed → the score decreases.
Predictive Traceability	The SHAP explanation aligns with the model's output.	The sequence of positive SHAP values corresponds to higher predicted scores.

The three aspects presented in the table—semantic coherence, causal consistency, and predictive traceability—illustrate how the model systematically understands and explains knowledge. The semantic coherence aspect indicates that the model can recognize the semantic relationships among tokens that align with the concepts defined in the evaluation rubric. For instance, the word "primary" increases the score because it corresponds to the correct concept of a primary key. The causal consistency aspect describes the cause-and-effect relationship between tokens and prediction outcomes. When an essential token such as "key" is removed, the score decreases, suggesting that the model genuinely relies on that information in its decision-making process. The predictive traceability aspect shows that the SHAP explanations are consistent with the model's predictions, allowing the decision-making process to be transparently traced and clearly interpreted. Overall, these three aspects demonstrate that the model does not merely learn statistical patterns but also constructs logical, transparent, and justifiable knowledge relationships.

Fidelity to Rubric (Alignment with Expert Scoring Rubric)

In the fidelity to rubric aspect, the model must evaluate student responses based on the same epistemic principles applied by human experts, rather than relying solely on surface-level statistical patterns. In the SHAP-based method, this is

implemented by mapping tokens to scores, where positive SHAP values align with key rubric terms defined by experts (e.g., Primary Key, unique, relationship between tables). **Tabel 5.** presents the indicators of fidelity to rubric in the model explanation.

Tabel 5. Indicators of fidelity to rubric

Rubric Principle	Relevant Tokens (SHAP +)	Indication
Unique PK per table	primary, key, unique	The model evaluates according to the conceptual rule of unique primary keys per table.
FK refers to PK of another table	foreign, reference, relation	The model identifies inter-entity relationships consistent with referential integrity.
Conceptual error	many, same, duplicate (SHAP -)	The model lowers the score when conceptual misunderstandings are detected.

The “Rubric Principle” column presents three representative examples from the fidelity to rubric analysis, demonstrating the alignment between SHAP values and the epistemic principles of the expert rubric. Principles such as “a unique Primary Key per table” and “a Foreign Key referencing the Primary Key of another table” are directly derived from the expert evaluation rubric, which serves as the ground truth in the dataset. The implementation results show a clear relationship between relevant tokens and the rubric concepts. Tokens with positive SHAP values contribute to higher predicted scores, while tokens associated with misconceptions (negative SHAP values) decrease the model’s score. Therefore, this table represents fidelity to rubric as evidence of epistemic alignment between the model and the human expert.

The evidence of fidelity to rubric is illustrated in **Figure 6**, which presents a comparison of SHAP value attributions between scores 3 and 0.

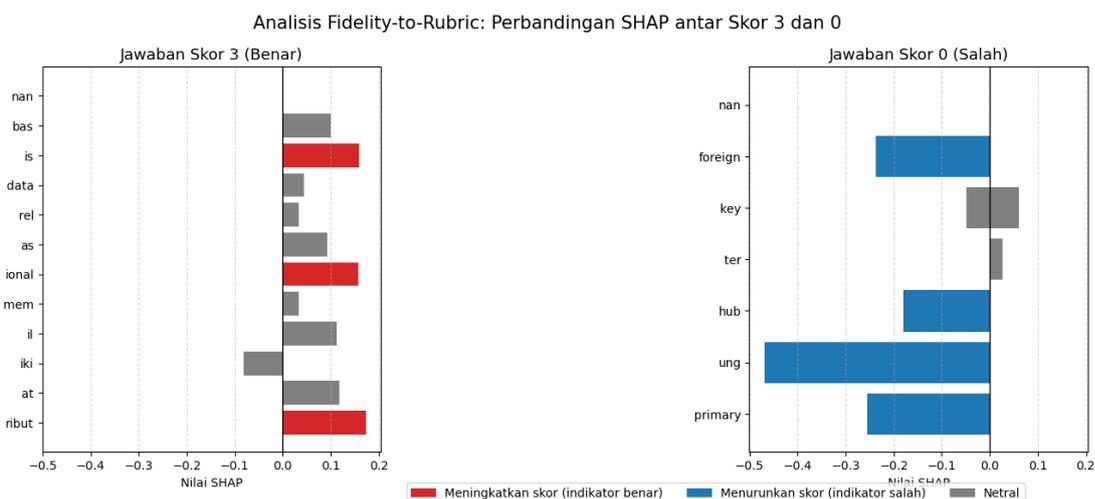


Figure 6. Fidelity-to-Rubric Analysis Based on SHAP Values for Each Token

Philosophically, faithfulness and fidelity represent two distinct but complementary epistemic constraints. Faithfulness concerns the internal validity of explanations—whether the explanation accurately reflects the actual causal and computational mechanisms of the model. Fidelity, in contrast, concerns external normative alignment—whether the model’s evaluation adheres to human-defined epistemic standards embedded in the scoring rubric. Together, these constraints ensure that the system is not only explainable, but epistemically responsible.

3.3 The Model as an Evaluative Agent in Epistemic Reflection

Based on the two aforementioned principles, faithfulness ensures that the model’s internal epistemic logic aligns with its inference process—representing what the model “knows.” Meanwhile, fidelity to rubric guarantees external epistemic alignment between the model and human experts—representing what “should be considered correct.”

In the context of this study, both principles jointly reinforce epistemic transparency, ensuring that the model operates not merely as a computational tool but as an evaluative epistemic agent capable of justifying its reasoning process. This epistemic reflection highlights that the model’s explanations (via SHAP) do not only quantify the contribution of linguistic features but also demonstrate how the model’s reasoning aligns with human epistemic standards in assessment.

4. Conclusion

The SHAP (SHapley Additive exPlanations) values attributed to each token within a proposition provide epistemic evidence of students’ answers. A positive SHAP value (+) indicates support for correctness indicators, a value of zero (0) denotes neutrality, and a negative value (–) indicates lack of support for correctness. The greater the number of tokens with positive SHAP values within a proposition, the higher the overall SHAP score, thereby strengthening the epistemic validity of the answer. This first conclusion establishes the evidence chain within the epistemological aspect of student answer assessment (propositions).

Furthermore, positive SHAP values reinforce the faithfulness of the model in providing explanations that accurately reflect its internal reasoning process, while the positive correlation between rubric-related tokens and expert-assigned scores demonstrates the model’s fidelity to the epistemic structure of human evaluation.

Authors’ Declaration

Authors’ contributions and responsibilities - The authors made substantial contributions to the conception and design of the study. The authors took responsibility for data analysis, interpretation, and discussion of results. The authors read and approved the final manuscript.

Funding - No funding information from the authors.

Availability of data and materials - All data is available from the authors.

Competing interests - The authors declare no competing interest.

Additional information - No additional information from the authors.

References

- [1] M. Al-alshaqi, D.B Rawat, C. Liu, "A BERT-Based Multimodal Framework for Enhanced Fake News Detection Using Text and Image Data Fusion", *Computers* 2025, 14, 237. <https://doi.org/10.3390/computers14060237>
- [2] Jose Campino, "Unleashing the transformers: NLP models detect AI writing in education", *Journal of Computers in Education*, 2025
- [3] Xiaojing [3]et al, "Towards transparent and trustworthy prediction of student learning achievement by including instructors as co-designers: a case study", *Education and Information Technologies*, 2024
- [3] Tanya Evans & Inae Jeong, "Concept maps as assessment for learning in university mathematics", *Educational Studies in Mathematics* (2023) 113:475–498 <https://doi.org/10.1007/s10649-023-10209-0>, 2023
- [4] S. Hariharasitaraman,et al., "Interpretable and ethical learning assessmenttransformer (IELAT): an explainable transformer model for personalized student assessments", *Cogent Education*, 2025
- [5] Danial Hoosyhar & Yeonwook Yang, "Problems With SHAP and LIME in Interpretable AI for Education: A Comparative Study of Post-Hoc Explanations and Neural-Symbolic Rule Extraction", *IEEE Access Vol 2*, 2024
- [6] Didik Dwi Prasetya,et al, "Analysis of quality of knowledge structure and students' perceptions in extension concept mapping", *Springer Open: Research and Practice in Technology Enhanced Learning*, 2022
- [7] Federica Russo et al, "Connecting ethics and epistemology of AI", *AI & SOCIETY*, 2024
- [8] Vidasha Ramnarain Seetohul; Yasmine Rosunally, Vandana Bassoo, "A Unified Conceptual Hybrid Framework for the Automated Assessment of Short Answers", *International Artificial Intelligence in Education Society*, 2025
- [9] Villegas-Ch, et al, "Explainable educational assistant integrated in Moodle: automated semantic assessment and adaptive tutoring based on NLP and XAI", *Discover Artificial Intelligence*, 2025
- [10] Shixiao Wang, "DeBERTa with hats makes Automated Essay Scoring system better", *Proceedings of the 4th International Conference on Signal Processing and Machine Learning*, 2024
- [11] Xuansheng Wu, "Unveiling Scoring Processes: Dissecting the Differences Between LLMs and Human Graders in Automatic Scoring", 2025
- [12] Wenbo Xu, , et al, "Explainable AI for education: Enhancing essay scoring via rubric-aligned chain-of-thought prompting", 2025
- [13] Gülsüm Asıksoy, "Computer-Based Concept Mapping as a Method for Enhancing the Effectiveness of Concept Learning in Technology-Enhanced Learning", *MDPI*, 2019

- [14] Xiaojing Duan, et al, "Towards transparent and trustworthy prediction of student learning achievement by including instructors as co-designers: a case study", *Education and Information Technologies*, 2024 P. Verma et al., "Meta-Heuristic Optimization Techniques Used for Maximum Power Point Tracking in Solar PV System," *Electronics*, vol. 10, no. 19, p. 2419, Oct. 2021, doi: 10.3390/electronics10192419.