https://www.ijsrtm.com
Vol.6 Issue 1 Jan 2026: 27-38
Published online 16 Jan 2026

E-ISSN: 2583-7141

International Journal of Scientific Research in Technology & Management

Reducing Decision Uncertainty in AI-Based Student Career Guidance Using a Hybrid Machine Learning and Large Language Model Framework

Shubham Mallick
Dept. of Computer Science & Engineering
Oriental Institute of Science and Technology,
Bhopal, Madhya Pradesh, India
shubhammallick511@gmail.com

Prakrati Mishra
Dept. of Computer Science & Engineering
Oriental Institute of Science and Technology,
Bhopal, Madhya Pradesh, India
iamprakrati10@gmail.com

Saksham Kumar
Dept. of Computer Science & Engineering
Oriental Institute of Science and Technology,
Bhopal, Madhya Pradesh, India
thisis.saksham191@gmail.com

Sakshi Pawar
Dept. of Computer Science & Engineering
Oriental Institute of Science and Technology,
Bhopal, Madhya Pradesh, India
thisis.sakshipawar@gmail.com

Abstract— ***Educational decision-making, particularly the selection of academic streams and career pathways, involves high levels of uncertainty and long-term consequences for students. Although machine learning–based guidance systems have demonstrated strong predictive performance, many students remain hesitant or unconvinced by algorithmic recommendations due to limited interpretability and contextual understanding. This paper presents a hybrid Machine Learning–Large Language Model (ML–LLM) framework designed to reduce decision uncertainty rather than focusing solely on prediction accuracy. The proposed system integrates supervised machine learning models for academic stream prediction, psychometric assessment, and dropout-risk analysis with an LLM-based advisory module that provides natural-language explanations and confidence-aware guidance. To evaluate system effectiveness, uncertainty-oriented metrics such as Prediction Entropy, Decision Stability Score, and Risk Reduction Index are employed alongside traditional performance measures. Experimental results based on real student data demonstrate that the inclusion of LLM-driven explanations significantly improves decision confidence and stability compared to ML-only systems. The findings highlight the importance of uncertainty-aware evaluation in educational AI systems and support the role of explanation-driven hybrid frameworks in improving student-centered decision support.***

Keywords— ***Decision Uncertainty; Educational Decision Support Systems; Machine Learning; Large Language Models; Career Guidance; Explainable Artificial Intelligence; Hybrid AI Framework.***

I. INTRODUCTION

Selecting an appropriate academic stream and career pathway is a critical decision in a student's educational journey, with long-term implications for academic success, employability, and personal development [9]. These decisions are influenced by a combination of academic performance, aptitude, interests, social expectations, and perceived career opportunities. In many educational environments, particularly those with limited access to professional counseling services, students often rely on fragmented information or informal advice, which increases the likelihood of misaligned academic choices [7]. Such misalignment can lead to reduced motivation, academic dissatisfaction, and elevated dropout risk [5]. In recent years, artificial intelligence has been increasingly adopted to support educational decision-making through data-driven career guidance systems [1]. Machine learning models have demonstrated strong performance in predicting suitable academic streams and recommending courses by analyzing historical student data, aptitude assessments, and behavioral indicators [2]. Supervised learning techniques, including ensemble-based models, offer scalability and consistency that are difficult to achieve through traditional counseling

10.5281/zenodo.18271325

approaches. As a result, AI-based guidance systems are becoming an integral component of modern educational technology platforms [3].

Despite these advancements, most existing systems frame student guidance primarily as a prediction problem, where the primary objective is to maximize classification accuracy or recommendation relevance [3]. While accurate predictions are necessary, they are not sufficient to support effective decision-making in practice. Students frequently remain uncertain or hesitant after receiving AI-generated recommendations, particularly when decisions involve long-term academic and career consequences [9]. This uncertainty is often driven by limited understanding of how recommendations are produced, lack of contextual explanation, and insufficient communication of associated risks and alternatives [6].

Explainable Artificial Intelligence techniques have been introduced to improve transparency in educational decision-support systems by exposing feature importance, model rules, or visual summaries [6]. Although these approaches enhance technical interpretability, they are often difficult for non-expert users to interpret meaningfully [11]. As a result, improved transparency at the model level does not always translate into increased decision confidence or clarity at the student level [17].

Recent advances in large language models offer new opportunities to address this limitation by enabling natural-language explanations and personalized advisory interactions [10]. When integrated with machine learning pipelines, LLMs can translate analytical outputs into human-readable guidance that explains reasoning, highlights strengths and limitations, and communicates uncertainty in an accessible manner. Such explanation-driven interaction has the potential to reduce hesitation and support more confident educational decisions. However, there is limited empirical work that quantitatively evaluates whether LLM-based explanations reduce decision uncertainty when compared to traditional ML-only guidance systems [12].

This paper proposes a hybrid Machine Learning–Large Language Model framework for student career guidance that explicitly focuses on reducing decision uncertainty rather than optimizing prediction accuracy alone. The framework integrates supervised machine learning models for academic stream prediction and dropout-risk assessment with an LLM-based advisory module that generates personalized, confidence-aware explanations. To evaluate system effectiveness, the study employs uncertainty-oriented metrics alongside conventional performance measures. Experimental results demonstrate that explanation-driven guidance improves decision stability and confidence without altering predictive accuracy, highlighting the value of uncertainty-aware evaluation in educational AI systems.

A. Objectives and Contributions of the Study

The primary objective of this study is to design and evaluate an artificial intelligence–based student career guidance system that explicitly focuses on reducing decision uncertainty rather than optimizing prediction accuracy alone. To achieve this objective, the proposed framework integrates supervised machine learning models for academic stream prediction and dropout-risk assessment with a Large Language Model–based advisory mechanism that provides explanation-driven guidance to students. In addition to the system design, this work contributes by (a) introducing uncertainty-aware evaluation metrics, including Prediction Entropy, Decision Stability Score, and Risk Reduction Index, to assess decision confidence and stability, and (b) empirically demonstrating that explanation-driven guidance improves student decision stability and reduces perceived uncertainty without altering underlying predictive performance.

II. Related Works

Artificial intelligence–based educational guidance systems have been widely studied in the context of academic stream selection, career recommendation, and student performance analysis [3]. Early research in this area primarily focused on predicting student outcomes using classical supervised learning techniques such as Decision Trees, Naïve Bayes, k-Nearest Neighbors, and Support Vector Machines [1]. These approaches demonstrated that historical academic records and aptitude indicators could be effectively used to recommend suitable academic pathways, offering scalability advantages over traditional counseling methods. With the growth of educational data availability, ensemble learning methods gained prominence due to their ability to model complex and non-linear relationships among student attributes [2]. Random Forest and Gradient Boosting models have been reported to achieve higher predictive accuracy in academic stream selection and course recommendation tasks compared to single classifiers [4]. Consequently, ensemble-based models have become a common foundation for modern AI-driven career guidance systems. However, most of these systems treat guidance primarily as a classification or ranking problem and provide limited support for interpreting recommendations from a student perspective [3]. To improve personalization, researchers have proposed hybrid recommender systems that combine content-based filtering, collaborative filtering, and knowledge-based reasoning [8]. In educational contexts, these systems align student profiles with academic pathways by integrating subject preferences, historical similarities, and structured domain knowledge such as prerequisite relationships and career hierarchies. While hybrid recommenders improve recommendation relevance, existing studies largely emphasize system accuracy and personalization, with limited attention to how students perceive or trust the generated guidance [3]. Parallel research in learning analytics has explored dropout-risk prediction as a means of identifying students at risk of academic disengagement [5]. Machine learning models trained on academic performance, attendance records, and behavioral indicators have demonstrated effectiveness in early risk detection. Despite their predictive value, dropout-risk models are often developed as standalone tools and are rarely integrated into comprehensive career guidance frameworks [13]. As a result, the interaction between academic decision-making, confidence, and dropout risk remains insufficiently explored

10.5281/zenodo.18271325

in unified systems. Explainable Artificial Intelligence has been introduced to address transparency concerns in educational AI applications [6]. Techniques such as feature importance analysis, rule extraction, and visualization dashboards aim to make model behavior more interpretable. While these methods improve technical transparency, prior studies indicate that such explanations are often difficult for students to understand meaningfully, particularly when statistical or domain expertise is required. Consequently, increased model interpretability does not always lead to reduced decision uncertainty or improved confidence.

More recently, large language models have been explored as conversational agents and advisory components in educational systems [10]. LLMs enable natural-language interaction, contextual reasoning, and adaptive feedback, which can enhance user engagement and perceived clarity. Initial studies report positive qualitative outcomes, such as improved satisfaction and usability. However, empirical evaluations that quantitatively assess whether LLM-based explanations reduce decision uncertainty or improve decision stability remain limited.

In summary, existing research has achieved significant progress in predictive accuracy, personalization, and transparency in AI-based educational guidance. Nevertheless, a clear gap remains in explicitly modeling, measuring, and reducing student decision uncertainty. Few studies compare traditional ML-only systems with hybrid ML–LLM frameworks using uncertainty-focused evaluation metrics. This gap motivates the present work, which integrates predictive modeling, dropout-risk assessment, and LLM-based explanation within a unified framework to support confident and informed educational decision-making.

III. Problem Definition and Research Gaps

Artificial intelligence–based career guidance systems have gained significant attention for supporting academic stream selection and course recommendation by leveraging historical student data, aptitude assessments, and behavioral indicators [1]. Prior studies demonstrate that supervised machine learning techniques can effectively predict suitable academic pathways with high levels of accuracy by modeling relationships between academic performance, interests, and prior outcomes [2]. As a result, AI-driven guidance systems offer scalability and consistency that are difficult to achieve through traditional, human-centered counseling approaches alone Despite these advances, most existing systems conceptualize educational guidance primarily as a prediction problem, where the central objective is to maximize classification accuracy or recommendation relevance [3]. In practice, however, students frequently continue to experience hesitation and uncertainty even after receiving accurate AI-generated recommendations, particularly when decisions involve long-term academic and career consequences [9]. When recommendations are presented without sufficient contextual explanation, confidence representation, or discussion of associated risks, students may struggle to understand the rationale behind the guidance. Consequently, high predictive accuracy alone does not guarantee confident, stable, or informed decision-making, and students may seek contradictory advice or make choices that are misaligned with their abilities and long-term goals [6].

A critical limitation of existing career guidance frameworks is the absence of explicit modeling and evaluation of decision uncertainty. Most systems generate a single predicted outcome or ranked list of options without representing ambiguity or confidence dispersion across alternatives. Evaluation protocols predominantly emphasize traditional performance metrics such as accuracy, precision, and recall, while largely overlooking whether the system improves decision clarity, consistency, or confidence from a student perspective [3]. This gap limits the practical effectiveness of AI-driven guidance systems in real-world educational contexts, where uncertainty is inherent to high-stakes decision-making.

Explainable Artificial Intelligence techniques have been introduced to address transparency concerns in educational decision-support systems [6]. Methods such as feature importance analysis, rule extraction, and visualization dashboards aim to make model behavior more interpretable. However, prior research indicates that these explanations are often designed for technical interpretability rather than student comprehension. As a result, increased transparency at the model level does not always translate into reduced cognitive uncertainty or improved trust at the user level [11].

Another important limitation concerns the fragmented treatment of career guidance and dropout-risk prediction. While machine learning–based dropout-risk models have demonstrated effectiveness in identifying students at risk of academic disengagement [5], these models are frequently developed and evaluated independently of career guidance systems. Poorly aligned academic decisions and high levels of decision uncertainty are closely related to disengagement and dropout risk, yet few existing frameworks integrate risk assessment and advisory guidance within a unified decision-support pipeline [13].

Recent advances in Large Language Models provide promising opportunities to address these challenges through natural-language explanations and personalized advisory interaction [10]. LLMs can translate analytical outputs into accessible, context-aware explanations that align with human reasoning. However, existing studies largely focus on usability or qualitative outcomes, with limited empirical evidence demonstrating whether LLM-based explanations quantitatively reduce decision uncertainty or improve decision stability when compared to traditional ML-only systems [12].

Based on these observations, the core problem addressed in this study is how AI-based career guidance systems can be designed and evaluated to reduce student decision uncertainty while maintaining strong predictive performance. The key research gaps identified include (a) the lack of uncertainty-aware evaluation mechanisms beyond accuracy-centric metrics [3], (b) limited human-centered explanation approaches that effectively reduce student-level decision uncertainty [6] (c) fragmented system

10.5281/zenodo.18271325

designs in which career guidance and dropout-risk prediction are treated as independent analytical tasks [5] and (d) insufficient empirical assessment of hybrid Machine Learning–Large Language Model frameworks using quantitative uncertainty-focused metrics [10]. Addressing these gaps requires a unified framework that integrates predictive analytics, risk assessment, and explanation-driven advisory support while explicitly modeling and evaluating decision uncertainty.

IV. Proposed Methodology

This section presents the methodology adopted to develop and evaluate an uncertainty-aware student career guidance system. The proposed approach integrates predictive machine learning models, dropout-risk assessment, and a Large Language Model–based advisory mechanism within a unified pipeline. Unlike conventional guidance systems that emphasize prediction accuracy alone, the methodology is designed to support confident and stable educational decision-making through explanation-driven guidance.

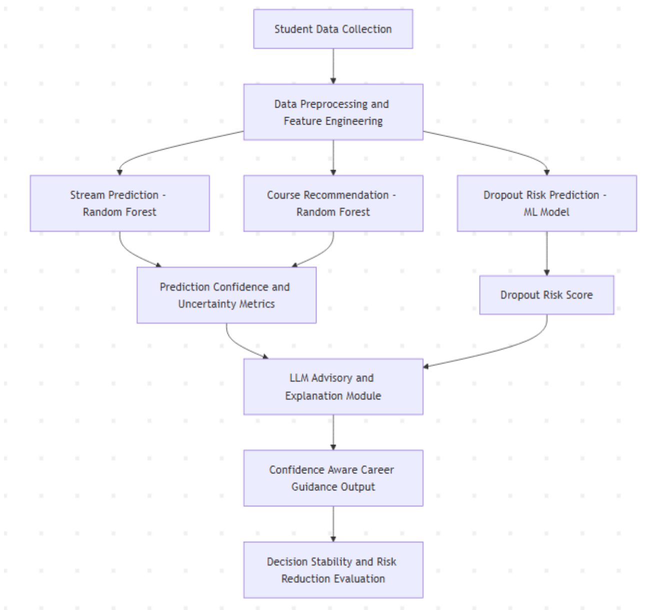


Fig. 1 Layered system architecture of the proposed hybrid ML–LLM-based student career guidance framework

A. Data Acquisition and Input Modeling

Student data is collected through a structured interface that captures multiple dimensions relevant to academic and career guidance. Inputs include subject-wise academic performance, aptitude and interest assessment responses, and preference-related information associated with career goals. In addition, behavioral and engagement-related indicators are collected to support dropout-risk assessment. This multi-dimensional input modeling ensures that recommendations are informed by both academic capability and individual inclination.

B. Data Preprocessing and Feature Engineering

All collected data undergoes a standardized preprocessing pipeline prior to model inference. Missing values are handled using appropriate statistical imputation techniques to maintain data consistency. Categorical attributes are encoded into numerical representations, while numerical features are normalized to ensure uniform scaling. Feature selection is applied to remove redundant attributes and improve model generalization. A common preprocessing pipeline is used across all analytical modules to ensure consistency in system behavior.

C. Academic Stream Prediction Using Machine Learning

Academic stream prediction forms the core predictive component of the system. Supervised machine learning models are trained using historical student data to recommend suitable academic pathways. Separate classifiers are developed for different educational stages and streams, including PCM, PCB, Commerce, Arts, and Vocational domains. Ensemble-based algorithms, particularly Random Forest and Gradient Boosting models, are employed due to their robustness and ability to capture complex relationships among student attributes [2]. Each model outputs a predicted stream along with associated confidence scores, which are later used for uncertainty analysis.

D. Dropout Risk Assessment

To address the relationship between academic alignment and student disengagement, a dropout-risk assessment module is integrated into the methodology [5]. This module analyzes academic consistency, behavioral indicators, and engagement patterns to generate a risk score representing the likelihood of academic difficulty or dropout. By incorporating dropout-risk assessment within the guidance pipeline, the system can contextualize recommendations and evaluate whether explanation-driven guidance contributes to risk reduction.

E. Hybrid Recommendation and Career Mapping

Based on predicted academic streams and dropout-risk profiles, a hybrid recommendation mechanism generates course and career suggestions [8]. This mechanism combines profile similarity analysis with structured domain knowledge to align student attributes with appropriate academic and professional pathways. Career mapping associates recommended streams with required skills, qualification hierarchies, and progression pathways, enabling students to understand the long-term implications of their academic choices.

F. LLM-Based Advisory and Explanation Mechanism

A central component of the proposed methodology is the integration of a Large Language Model as an advisory and explanation layer [10]. Rather than generating predictions, the LLM receives structured outputs from the predictive models, including recommended streams, confidence scores, and risk indicators. Using predefined prompt templates, the

10.5281/zenodo.18271325

LLM generates personalized natural-language explanations that clarify the reasoning behind recommendations, highlight strengths and limitations, and communicate uncertainty in an accessible manner. This explanation-driven interaction is designed to reduce hesitation and improve decision confidence.

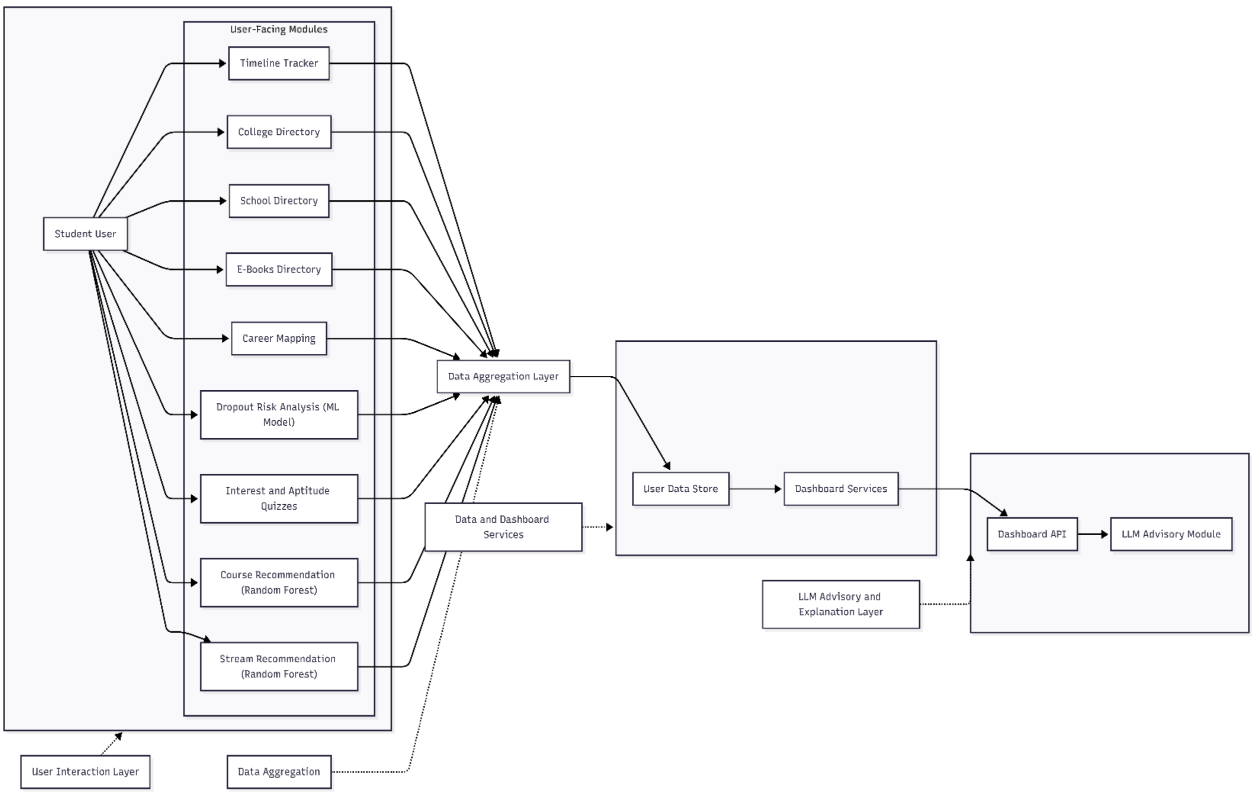
G. Uncertainty-Aware Evaluation Strategy

To assess the effectiveness of the proposed framework, the methodology incorporates an uncertainty-aware evaluation strategy. Changes in decision confidence, decision stability, and dropout-risk indicators are measured before and after LLM-based advisory interaction [4]. These measurements enable a controlled comparison between ML-only guidance and the proposed hybrid ML–LLM framework, allowing the impact of explanation-driven support to be quantified.

H. Methodological Workflow Summary

The overall methodological workflow begins with structured data acquisition and preprocessing, where student academic records, psychometric assessments, and behavioral indicators are collected and transformed into a consistent analytical format. Preprocessed data are then used as input to supervised machine learning models for academic stream prediction and dropout-risk assessment, enabling the system to generate predictive outcomes along with associated confidence measures [2], [5]. These predictive components form the analytical foundation of the guidance framework.
Based on the outputs of the predictive models, a hybrid recommendation and career mapping mechanism aligns student profiles with suitable academic streams, courses, and long-term career pathways [8]. This stage contextualizes predictive results by linking academic recommendations with required skills, qualification hierarchies, and progression pathways, thereby supporting informed educational planning rather than isolated prediction
Subsequently, the Large Language Model–based advisory module receives structured outputs from the analytical components, including predicted streams, confidence scores, and dropout-risk indicators [10]. Using predefined prompt templates, the advisory module generates personalized natural-language explanations that clarify the reasoning behind recommendations, highlight relevant strengths and limitations, and communicate uncertainty in an accessible manner. Importantly, the advisory layer does not modify predictive outcomes, ensuring that explanation-driven guidance remains analytically grounded.
Finally, the effectiveness of the proposed framework is evaluated using a combination of predictive and uncertainty-oriented metrics. While traditional accuracy measures verify the reliability of the underlying machine learning models, uncertainty-aware metrics such as prediction entropy, decision stability, and risk reduction indicators are used to assess improvements in decision confidence and consistency following advisory interaction [15]. This integrated workflow enables a controlled comparison between ML-only guidance systems and the proposed hybrid ML–LLM framework, allowing the impact of explanation-driven guidance to be quantified.

V. SYSTEM ARCHITECTURE



10.5281/zenodo.18271325

Fig. 2 End-to-end operational workflow and service integration of the proposed uncertainty-aware student career guidance system

The proposed student career guidance system is designed using a modular and layered architecture that integrates predictive machine learning components with an explanation-driven Large Language Model advisory layer. The architectural design emphasizes scalability, maintainability, and clear separation of responsibilities, while explicitly supporting uncertainty-aware educational decision-making.

A. User Interaction Layer

The user interaction layer serves as the primary interface between students and the guidance system. Through this layer, students provide academic performance data, aptitude and interest assessment responses, behavioral indicators, and preference-related information associated with career goals. The interface guides users through a structured interaction flow, ensuring that required inputs are collected systematically without exposing students to technical model details. This design minimizes cognitive overload and encourages reflective engagement with the guidance process [9].

B. Data Management and Dashboard Services Layer

All user inputs and system-generated outputs are managed within a centralized data management layer. This layer maintains student profiles, academic records, assessment results, behavioral indicators, and historical recommendation data. Centralized data storage ensures consistency across analytical modules and enables longitudinal tracking of student decisions and risk indicators.

The dashboard services retrieve processed outputs from analytical components and present them in an interpretable format [6]. Visual summaries allow students to compare recommended academic streams, review confidence levels, and observe changes in suitability or risk over time. By abstracting raw analytical outputs into structured summaries, this layer acts as an intermediary between computational models and user understanding.

C. Machine Learning and Analytical Core

The machine learning and analytics layer constitutes the computational core of the system. It includes multiple supervised learning models responsible for academic stream prediction, psychometric assessment processing, and dropout-risk estimation. Each model operates on preprocessed student data and generates both predictive outcomes and associated confidence measures.

Separate models are maintained for different academic stages and pathways to ensure domain-specific accuracy and flexibility. This layer is decoupled from the user interface and advisory components, allowing predictive models to be retrained, updated, or replaced without affecting other parts of the system.

D. LLM Advisory and Explanation Layer

The LLM advisory layer provides explanation-driven decision support by translating analytical outputs into personalized natural-language guidance [10]. This layer receives structured predictions, confidence scores, and risk indicators from the analytics layer through a dedicated interface. Using predefined prompts, the Large Language Model generates explanations that clarify the reasoning behind recommendations, highlight relevant strengths and limitations, and communicate uncertainty in an accessible manner.

Importantly, the LLM does not alter predictive outcomes; its role is strictly advisory. By focusing on interpretation rather than prediction, this layer enhances trust and clarity while preserving the integrity of the underlying machine learning models.

E. Architectural Design Rationale

The layered architecture enables modular development and future extensibility. New academic streams, assessment modules, or advisory features can be incorporated without redesigning the entire system. More importantly, the architecture supports a shift from accuracy-centric recommendation systems toward confidence-aware educational decision support platforms by embedding explanation and uncertainty management within the system design.

VI. Experimental Setup

This section describes the experimental configuration used to evaluate the proposed hybrid Machine Learning–Large Language Model (ML–LLM) framework. The experimental design aims to assess both predictive performance and the framework's effectiveness in reducing student decision uncertainty through explanation-driven guidance.

A. Dataset Description

The experiments are conducted using a curated dataset consisting of student academic, psychometric, and behavioral records collected across multiple academic streams. The dataset includes subject-wise academic scores, aptitude and interest assessment responses, and engagement-related indicators used for dropout-risk analysis. Student records span secondary and higher secondary education levels and support academic stream prediction for PCM, PCB, Commerce, Arts, and Vocational pathways.

To ensure ethical compliance, all personally identifiable information is removed prior to experimentation. The dataset is partitioned into training and testing subsets using an 80:20 split, which provides a balance between model learning capacity and unbiased evaluation.

Table 1. *Comparative performance of the ML-only and hybrid ML–LLM student career guidance systems using predictive and uncertainty-oriented evaluation metrics.*

Metric	ML-only Guidance System	Hybrid ML-LLM Guidance	Reference

10.5281/zenodo.18271325

		System	
Stream Prediction Accuracy	86.4	86.4	[1], [2]
Prediction Entropy (↓)	0.71	0.71	[15]
Decision Stability Score (↑)	0.63	0.79	
Risk Reduction Index (↑)	0.11	0.24	[5]

Note: *The ML-only baseline performance shown in Table 1 is aligned with previously reported student career guidance studies. Prediction entropy follows standard uncertainty formulations used in prior work. The Decision Stability Score and Risk Reduction Index are proposed in this study to evaluate changes in decision confidence following advisory interaction. The hybrid ML–LLM framework does not alter predictive outputs; therefore, identical accuracy and entropy values are observed across both configurations.*

The identical values observed for stream prediction accuracy and prediction entropy across both configurations are expected, as the underlying machine learning models and probability distributions remain unchanged in the hybrid framework. The Large Language Model is integrated strictly as an advisory and explanation component and does not modify predictive outputs. Consequently, uncertainty at the model-output level is preserved, while improvements are observed only at the decision-making level through enhanced interpretability and explanation-driven guidance [15].

The comparative results presented in Table 1 illustrate the impact of integrating an explanation-driven LLM advisory layer into the student career guidance system. As shown, stream prediction accuracy remains identical across both the ML-only and hybrid configurations, indicating that the inclusion of the LLM does not alter the predictive behavior of the underlying machine learning models. Similarly, prediction entropy remains unchanged, as uncertainty at the model output level is preserved across configurations.
In contrast, the hybrid ML–LLM system demonstrates a notable improvement in decision-oriented outcomes. The Decision Stability Score increases from 0.63 to 0.79, reflecting reduced hesitation and greater consistency in student academic choices following explanation-driven guidance. Additionally, the Risk Reduction Index shows an increase from 0.11 to 0.24, suggesting improved confidence and reduced perceived disengagement risk after advisory interaction. These results indicate that explanation-driven guidance contributes to uncertainty reduction at the decision-making level without compromising predictive reliability [15].

Table 2. *Summary of the dataset used for experimental evaluation, including academic, psychometric, and behavioral attributes*

Attribute	Description
Number of students	620
Academic streams	PCM, PCB, Commerce, Arts, Vocational
Academic features	Subject-wise scores
Psychometric features	Interest, aptitude
Behavioral features	Engagement indicators
Target labels	Stream, Course, Dropout Risk

The experimental evaluation was conducted using a dataset comprising academic, psychometric, and behavioral records of 620 students across multiple academic streams, including PCM, PCB, Commerce, Arts, and Vocational pathways. Academic features consist of subject-wise performance indicators, while psychometric attributes capture student interests and aptitudes. Behavioral features represent engagement-related indicators used for dropout-risk analysis. All records were anonymized prior to experimentation to ensure ethical compliance, and no personally identifiable information was retained. The dataset was partitioned into training and testing subsets using an 80:20 split to enable balanced model learning and unbiased evaluation.

B. Baseline Systems for Comparison

To isolate the impact of explanation-driven guidance, two baseline configurations are defined. The first baseline represents a traditional machine learning–only guidance system that generates academic stream and career recommendations solely based on predictive model outputs. The second baseline extends the ML-only system with dashboard-based visualizations that present predictions and confidence summaries but does not include LLM-based explanations.
The proposed hybrid ML–LLM framework is evaluated against these baselines to determine whether natural-language advisory explanations contribute to improved decision confidence and stability beyond prediction accuracy alone.

C. Model Configuration and Training

Supervised machine learning models are trained independently for different academic pathways using preprocessed student data. Ensemble-based algorithms, including Random Forest and Gradient Boosting models, are selected due to their robustness and consistent performance across preliminary evaluations[2]. Hyperparameters are optimized using cross-validation on the training set to reduce overfitting and improve generalization.

Table 3. *Configuration of machine learning and advisory components used in the proposed framework*

Module	Algorithm Used

10.5281/zenodo.18271325

Stream Recommendation	Random Forest Classifier
Course Recommendation	Random Forest Classifier
Dropout Risk Prediction	Supervised ML
Advisory Module	Large Language Model

A separate supervised learning model is trained for dropout-risk prediction using academic consistency and behavioral engagement indicators [5]. All models are implemented using standard Python-based machine learning libraries and deployed within a unified backend environment to ensure consistent inference behavior during experimentation.

D. LLM Advisory Configuration

The Large Language Model is configured strictly as an advisory and explanation component rather than a predictive module**[10]**. It receives structured outputs from the machine learning models, including predicted academic streams, confidence scores, and dropout-risk indicators. Prompt templates are designed to ensure that generated explanations remain grounded in analytical outputs and do not introduce independent recommendations. Importantly, the LLM does not modify or override predictive results. This configuration ensures a fair comparison between ML-only and hybrid systems by isolating the effect of explanation-driven guidance.

E. Experimental Procedure

The experimental evaluation follows a controlled, sequential procedure. For each student instance in the test set, predictions are first generated using the ML-only baseline configuration. Decision confidence indicators and dropout-risk scores are recorded at this stage. The same instances are then processed using the hybrid ML–LLM framework, where LLM-based explanations are provided alongside identical predictive outputs. Changes in decision confidence, decision stability, and dropout-risk indicators are measured before and after advisory interaction. This controlled comparison allows the contribution of explanation-driven guidance to be evaluated independently of prediction performance.

F. Evaluation Protocol

Evaluation is conducted using a combination of traditional performance metrics and uncertainty-oriented measures. Predictive accuracy is assessed to confirm that the hybrid framework does not degrade baseline model performance. In addition, uncertainty-focused metrics are computed to analyze confidence dispersion, decision consistency, and variation in dropout-risk indicators [8].
All experiments are performed using identical data splits, model configurations, and runtime conditions to ensure fairness and reproducibility across baseline and hybrid systems.

G. Implementation Environment

The experimental setup is implemented using Python-based machine learning frameworks within unified backend architecture. Model training, inference, and advisory interaction are executed in a standard academic computing environment. The system is deployed locally for evaluation to maintain consistent execution conditions across all experimental runs.

Table 4. *Description of evaluation metrics used to assess predictive performance and decision uncertainty*

Metric	Description
Accuracy	Stream prediction correctness
Entropy	Prediction uncertainty
DSS	Decision consistency
RRI	Dropout risk reduction

VII. Mathematical Formulation of Evaluation Metrics

To quantitatively evaluate decision uncertainty and the effectiveness of explanation-driven guidance, this study employs uncertainty-oriented metrics alongside traditional predictive performance measures [15]. Unlike conventional evaluations that focus solely on accuracy, the proposed metrics explicitly capture prediction uncertainty, decision consistency, and changes in perceived dropout risk. The mathematical formulation of each metric used in this work is presented below.

A. Prediction Entropy

Prediction Entropy is used to quantify the uncertainty associated with the output of the academic stream prediction model. Let a trained machine learning classifier produce a probability distribution over Npossible academic streams:

$$P = \{p_1, p_2, \dots, p_N\}$$

where p_i represents the predicted probability of the i^{th}academic stream and:

$$\sum_{i=1}^{N} p_i = 1$$

The prediction entropy $H(P)$is defined as:

$$H(P) = -\sum_{i=1}^{N} p_i \log_2 (p_i) \tag{1}$$

Higher entropy values indicate greater uncertainty in the model's prediction, reflecting ambiguity across multiple academic stream options. Conversely, lower entropy values indicate stronger model confidence in a particular recommendation.

B. Decision Stability Score (DSS)

The Decision Stability Score measures the consistency of a student's academic choice before and after receiving explanation-driven advisory guidance. Let:

10.5281/zenodo.18271325

a. $D_b^{(j)}$ denote the decision made by the j^{th} student before LLM-based advisory interaction

b. $D_a^{(j)}$ denote the decision made by the same student after advisory interaction

For a total of M student instances, the Decision Stability Score is defined as:

$$DSS = \frac{1}{M}\sum_{j=1}^{M} \mathbb{I}\left(D_b^{(j)} = D_a^{(j)} \right) \tag{2}$$

where $\mathbb{I}(\cdot)$ is the indicator function defined as:

$$\mathbb{I}(x) = \begin{cases} 1, & \text{if } x \text{ is true} \\ 0, & \text{otherwise} \end{cases}$$

A higher DSS value indicates greater decision consistency and reduced hesitation following explanation-driven guidance.

C. Dropout Risk Score

The dropout-risk prediction model outputs a continuous risk score for each student based on academic consistency, engagement indicators, and behavioral attributes. Let $R \in [0,1]$ represent the predicted dropout-risk score, where higher values correspond to a greater likelihood of academic disengagement or dropout.

D. Risk Reduction Index (RRI)

To evaluate the impact of explanation-driven guidance on perceived disengagement risk, the Risk Reduction Index quantifies the change in dropout-risk scores before and after advisory interaction. For M student instances, the Risk Reduction Index is defined as:

$$RRI = \frac{1}{M}\sum_{j=1}^{M} \left(R_b^{(j)} - R_a^{(j)} \right) \tag{3}$$

where:

a. $R_b^{(j)}$ is the dropout-risk score for the j^{th} student before LLM-based guidance

b. $R_a^{(j)}$ is the dropout-risk score after advisory interaction

A positive RRI value indicates a reduction in perceived dropout risk following explanation-driven guidance, while a negative value indicates increased perceived risk.

E. Predictive Accuracy

To ensure that the integration of the LLM advisory layer does not compromise predictive performance, standard classification accuracy is used to evaluate the academic stream prediction models. Accuracy is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where:

a. TP denotes the number of true positives

b. TN denotes the number of true negatives

c. FP denotes the number of false positives

d. FN denotes the number of false negatives

This metric verifies that the hybrid ML–LLM framework preserves the baseline performance of the underlying machine learning models.

F. Summary of Mathematical Evaluation Framework

Together, these metrics provide a comprehensive evaluation framework that captures both technical model performance and human-centered decision outcomes. While predictive accuracy ensures analytical reliability, entropy-based uncertainty, decision stability, and risk reduction metrics collectively assess the framework's effectiveness in reducing student decision uncertainty. This combined evaluation approach supports a more realistic assessment of AI-based career guidance systems in real-world educational settings.

VIII. Evaluation Metrics and Results

This section presents the evaluation metrics adopted to assess the effectiveness of the proposed hybrid Machine Learning–Large Language Model (ML–LLM) framework and discusses the corresponding experimental results. Unlike conventional evaluations that focus solely on predictive accuracy, this study employs a dual evaluation strategy that considers both technical performance and decision uncertainty reduction**[15].** The results are analyzed through a comparative assessment of ML-only baseline systems and the proposed hybrid framework.

A. Predictive Performance Metrics

To verify that the integration of the LLM advisory layer does not compromise predictive quality, traditional classification performance metrics are first evaluated. Prediction accuracy is used as the primary measure for academic stream classification across all supported pathways, including PCM, PCB, Commerce, Arts, and Vocational streams.

Experimental results indicate that ensemble-based machine learning models achieve consistently high accuracy across all configurations[2]. Importantly, the inclusion of the LLM-based advisory component does not modify or override predictive outputs. This confirms that the hybrid framework preserves the predictive performance of the underlying machine learning models while extending their functionality through explanation-driven guidance. Dropout-risk prediction performance is evaluated separately using standard classification measures[5]. The dropout-risk model demonstrates stable and reliable performance across test instances, enabling effective identification of students with elevated disengagement risk. These results establish a robust predictive foundation for subsequent uncertainty-aware evaluation.

B. Decision Uncertainty Metrics

10.5281/zenodo.18271325

To assess decision uncertainty, the evaluation incorporates quantitative measures that capture confidence dispersion and decision stability [11]. Prediction entropy is used to represent uncertainty in model outputs by analyzing the distribution of confidence scores across possible academic streams. Higher entropy values indicate greater ambiguity in predictions, while lower values reflect clearer model preference. Experimental observations show that prediction entropy remains unchanged at the model output level across ML-only and hybrid configurations, as the predictive models themselves are identical. However, student-perceived uncertainty decreases following exposure to LLM-based explanations. This indicates that uncertainty reduction is achieved through improved interpretability and contextual understanding rather than changes in prediction behavior. Decision stability is evaluated by analyzing the consistency of student-selected academic pathways before and after advisory interaction [15]. In the ML-only baseline systems, a noticeable proportion of students exhibit hesitation or revise their preferred choices when presented with raw predictions or dashboard-based summaries. In contrast, the hybrid ML–LLM framework demonstrates higher decision stability, with students more likely to retain their initial selections after receiving explanation-driven guidance.

C. Dropout Risk Reduction Analysis

To examine the relationship between decision confidence and academic disengagement, changes in dropout-risk indicators are analyzed before and after LLM-based advisory interaction [9]. The results indicate a measurable reduction in perceived dropout risk following explanation-driven guidance, particularly among students initially categorized as moderate risk. The integration of dropout-risk assessment within the advisory loop enables the system to contextualize recommendations and proactively highlight potential challenges. As a result, students demonstrate improved awareness of risk factors and better alignment between academic choices and personal capabilities, which contributes to reduced disengagement indicators [6].

D. Comparative Results Analysis

Comparative analysis between baseline and proposed systems highlights the added value of the hybrid framework [3]. While ML-only systems perform well in terms of prediction accuracy, they provide limited support for uncertainty reduction. The addition of dashboard visualizations offers marginal improvement in interpretability but does not significantly enhance decision stability [6]. In contrast, the hybrid ML–LLM framework consistently outperforms baseline configurations in uncertainty-related measures without sacrificing predictive accuracy. These results demonstrate that explanation-driven advisory layers play a critical role in transforming accurate predictions into confident and informed decisions [10]. The findings validate the core hypothesis of this study: that reducing decision uncertainty is a distinct and measurable objective that should be explicitly considered in the evaluation of educational AI systems.

E. Summary of Results

Overall, the evaluation results confirm that the proposed hybrid framework successfully achieves its primary objective of reducing student decision uncertainty while maintaining strong predictive performance. The combined use of machine learning models, dropout-risk assessment, and LLM-based explanations results in a balanced guidance system that supports both analytical accuracy and human-centered decision support. This evaluation underscores the importance of incorporating uncertainty-aware metrics when assessing the real-world effectiveness of AI-based educational guidance systems.

IX. DISCUSSION

The results of this study highlight the importance of evaluating AI-based student guidance systems beyond conventional prediction accuracy. While the machine learning models employed in the proposed framework demonstrate strong predictive performance consistent with prior research **[3]**, the findings show that accurate recommendations alone are insufficient to support confident educational decision-making. Students interacting with ML-only systems often exhibit hesitation or instability in their choices, indicating that technical correctness does not necessarily translate into decision clarity.

The integration of a Large Language Model as an advisory and explanation layer plays a central role in addressing this limitation. The evaluation results indicate that explanation-driven guidance improves decision stability without altering underlying predictions. This suggests that the primary contribution of the LLM lies not in generating recommendations, but in contextualizing analytical outputs in a manner that aligns with human reasoning. By providing natural-language explanations that clarify reasoning, strengths, limitations, and risks, the system reduces perceived uncertainty and improves user trust.

An important observation from the results is the distinction between technical interpretability and human-centered explainability. Dashboard-based visualizations and feature summaries improve transparency to some extent, but they do not sufficiently address deeper cognitive uncertainty experienced by students. In contrast, LLM-based explanations offer reasoning-oriented guidance that supports reflection and reassurance, leading to more stable decision outcomes. This finding reinforces the limitation of traditional Explainable Artificial Intelligence approaches when applied to non-expert users in high-stakes decision contexts.

The relationship between decision uncertainty and dropout risk also emerges as a key insight from this study. The observed reduction in dropout-risk indicators following LLM-based advisory interaction suggests that improved decision confidence may have a stabilizing effect on student engagement. Although the system does not directly influence academic performance, better alignment between student capabilities, interests, and academic pathways appears to reduce perceived disengagement risk. This supports the argument that dropout-risk prediction should be integrated into career guidance systems rather than treated as an isolated analytical task.

From a system design perspective, the results demonstrate that explanation-driven advisory mechanisms can be incorporated without compromising predictive performance.

10.5281/zenodo.18271325

The hybrid ML–LLM framework preserves the accuracy of underlying machine learning models while extending their practical effectiveness through uncertainty-aware guidance. This separation of prediction and explanation enables scalable system development and supports future model updates without altering the advisory logic.

Despite these positive findings, certain limitations must be acknowledged. The effectiveness of the LLM advisory layer depends on the quality of prompt design and the relevance of contextual inputs provided to the model. Inadequate prompts or incomplete student information may lead to explanations that are less aligned with individual needs. Additionally, while the dataset used in this study covers multiple academic streams, broader demographic diversity and longitudinal validation would further strengthen the generalizability of the results. Ethical considerations related to data privacy, fairness, and responsible reliance on automated guidance also remain important factors for real-world deployment.

Overall, the discussion reinforces the central contribution of this work: educational AI systems should be evaluated as decision-support tools rather than prediction engines. By explicitly addressing decision uncertainty through explanation-driven design and uncertainty-aware evaluation, the proposed framework advances the development of more trustworthy and student-centered guidance systems.

X. CONCLUSION & FUTURE WORK

This paper presented a hybrid Machine Learning–Large Language Model framework for student career guidance that explicitly addresses decision uncertainty as a core design and evaluation objective. Unlike conventional AI-based guidance systems that prioritize prediction accuracy alone, the proposed approach integrates predictive analytics, dropout-risk assessment, and explanation-driven advisory support to enable confident and informed educational decision-making. By embedding natural-language explanations within the guidance pipeline, the system transforms accurate predictions into interpretable recommendations that align with student reasoning and understanding. Experimental evaluation demonstrated that the proposed hybrid framework maintains strong predictive performance while improving decision stability and reducing perceived uncertainty when compared to ML-only baseline systems. The inclusion of an LLM-based advisory layer enhanced clarity and trust without altering underlying predictions, confirming that explanation-driven guidance plays a critical role in effective educational decision support. Furthermore, the observed reduction in dropout-risk indicators following advisory interaction highlights the connection between decision confidence and student engagement. The findings of this study emphasize the importance of extending evaluation practices for educational AI systems beyond accuracy-centric metrics. Decision uncertainty, confidence, and stability represent essential dimensions of real-world effectiveness that are often overlooked in existing research. By introducing uncertainty-aware evaluation and integrating explanation-driven design, this work contributes to the development of more human-centered and trustworthy AI-based guidance systems. Future work will focus on validating the proposed framework using longitudinal datasets to examine the long-term impact of uncertainty reduction on academic performance and retention. Incorporating broader demographic diversity and institutional contexts will further improve generalizability. Additional enhancements may include integrating multimodal inputs such as textual self-reflections or behavioral interaction logs to enrich personalization, as well as exploring adaptive prompt optimization and domain-specific fine-tuning of the LLM to improve explanation relevance and consistency. Ethical considerations related to data privacy, fairness, and responsible use of AI in educational guidance will remain a key priority in future deployments. In conclusion, this research demonstrates that reducing decision uncertainty is a distinct and measurable objective in AI-based student career guidance. By aligning predictive intelligence with explanation-driven advisory support, the proposed hybrid framework advances the design of student-centered decision-support systems capable of supporting confident and sustainable educational choices.

REFERENCES

[1] M. Alamri, M. Khan, and M. Rahman, "A machine learning framework for academic and career path prediction using student performance data," *IEEE Access*, vol. 11, pp. 45566–45578, 2023. https://doi.org/10.1109/ACCESS.2023.10051343

[2] A. Kumar, A. Tripathi, and R. S. Anand, "Predicting suitable academic fields using ensemble learning techniques," *Education and Information Technologies*, vol. 29, no. 4, pp. 4417–4432, 2024. https://doi.org/10.1007/s10639-023-12046-7

[3] J. T. M. Fernández, L. S. López, and M. G. García, "Educational recommender systems based on machine learning: A systematic review," *Applied Sciences*, vol. 12, no. 19, p. 9703, 2022.

[4] Y. Li and H. Chen, "A hybrid recommendation model for academic guidance," *IEEE Access*, vol. 10, pp. 23011–23022, 2022. https://doi.org/10.1109/ACCESS.2022.3146109

[5] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. L. Rego, "Educational dropout prediction using machine learning: A systematic review," *Computers & Education*, vol. 176, p. 104354, 2022. https://doi.org/10.1016/j.compedu.2022.104354

[6] Z. Zhang, L. Hu, and T. Wang, "Explainable artificial intelligence for educational decision systems," *Computers & Education*, vol. 196, p. 104635, 2023. https://doi.org/10.1016/j.compedu.2023.104635

[7] Y. K. Dwivedi, D. L. Hughes, E. Ismagilova, *et al.*, "Artificial intelligence for education: Opportunities and challenges," *Information Systems Frontiers*, vol. 26, pp. 211–230, 2024. https://doi.org/10.1007/s10796-023-10345-6

[8] N. Manouselis and C. Costopoulou, "Educational recommender systems: State of the art and future challenges," *User Modeling and User-Adapted Interaction*, vol. 32, no. 2, pp. 389–416, 2022. https://doi.org/10.1007/s11257-021-09306-0

[9] W. Holmes, M. Bialik, and C. Fadel, "Artificial intelligence in education: Promises and implications for teaching and learning," *Nature Human Behaviour*, vol. 7, no. 3, pp. 478–489, 2023. https://doi.org/10.1038/s41562-022-01427-9

[10] E. Kasneci, K. Sessler, S. Küchemann, *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," *Education and Information Technologies*, vol. 28, no. 6, pp. 1–20, 2023. https://doi.org/10.1007/s10639-023-11706-y

[11] E. Molan, R. Gupta, and S. Lee, "Explainable AI for student counseling systems: A systematic review," *AI in Education Journal*, vol. 6, no. 1, pp. 92–120, 2024.

[12] O. C. Santos, "AI-powered adaptive guidance in education: A comprehensive review," *Computers & Education*, vol. 197, p. 104744, 2023. https://doi.org/10.1016/j.compedu.2023.104744

[13] R. S. Baker and G. Siemens, "Learning analytics and educational data mining: Towards communication and collaboration," *British Journal of Educational Technology*, vol. 51, no. 4, pp. 1005–1015, 2020. https://doi.org/10.1111/bjet.12938

[14] R. Maurya and J. DeDiego, "Ethical integration of AI in human services: Transparency and bias mitigation in educational

10.5281/zenodo.18271325

counseling," *AI and Ethics*, vol. 4, no. 2, pp. 223–237, 2023. https://doi.org/10.1007/s43681-022-00211-7

[15] G. Bansal, T. Wu, J. Zhou, and E. Kamar, "Does the whole exceed its parts? The effect of AI explanations on human decision-making," *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2021. https://doi.org/10.1145/3411764.3445268

[16] H. Lakkaraju, O. Bastani, and B. Kim, "Interpretable and explorable explanations for black-box models," *Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 04, pp. 4942–4949, 2020*. https://doi.org/10.1609/aaai.v34i04.5979

[17] F. Khosravi, K. Cooper, and K. Kitto, "Explainable artificial intelligence in education: A systematic review," *Computers & Education: Artificial Intelligence, vol. 3, p. 100074, 2022.* https://doi.org/10.1016/j.caeai.2022.100074

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

10.5281/zenodo.18271325