



RESEARCH ARTICLE

OPEN ACCESS

EMPATEXT: AN LLM-POWERED COMPANION FOR BIAS-MITIGATED PROJECT-BASED LEARNING SUPPORT

Ahmed Si abdelhadi¹, Abdelkader Ouared², Abdelhafid Chadli³, Bendaoud Mebarek⁴

^{1,4}LRIAS-Artificial Intelligence and Systems Research Laboratory, University of Tiaret, Algeria

^{2,3}Computer Science Department, University of Tiaret, Algeria

¹<http://orcid.org/0009-0005-3609-5018> , ²<https://orcid.org/0000-0003-4257-0522> ,

³<https://orcid.org/0000-0003-0559-1439> , ⁴<http://orcid.org/0000-0002-6838-3867> 

E-mail: ahmed.siabdelhadi@univ-tiaret.dz, abdelkader.ouared@univ-tiaret.dz, abdelhafith.chadli@univ-tiaret.dz, bendaoud.mebarek@univ-tiaret.dz

ARTICLE INFO

Article History

Received: February 24, 2026

Reviewed: March 25, 2026

Accepted: May 15, 2026

Published: September 30, 2026

Keywords:

Pjbl, Activity Review,
Retrieval-Augmented
Generation (RAG) Large
Language Models (LLMs),
Constructive Feedback.

ABSTRACT

In Project-Based Learning (PjBL), psychological safety is essential for effective teamwork, while peer-review collaboration relies on strong interpersonal and communication skills. However, emotional reactions and cognitive biases such as impulsive judgments, confirmation bias, and decision fatigue can disrupt communication, weaken team dynamics, and compromise project outcomes. To address these challenges, we introduce EmpaText, an LLM-powered support system designed to enhance social awareness and foster communication, conflict resolution, and negotiation skills. EmpaText provides empathetic, context-aware, and constructive feedback to guide learners during peer-review activities, improving feedback clarity while reducing cognitive and emotional bias. Our framework comprises three core components: (i) Activity Enhancer, which improves learning deliverables; (ii) Review Facilitator, which guides assessment and promotes feedback quality; and (iii) Bias Mitigator, which employs linguistic markers and vocabulary analysis to detect and reduce cognitive and emotional biases. Through realistic PjBL scenarios and user-centered evaluations, we show that EmpaText achieves a 60.9% reduction in cognitive biases, increases specific, actionable feedback from 32.7% to 71.1%. User satisfaction reached 86.7% positive responses, demonstrating the system's effectiveness in fostering psychologically safe, bias-mitigated collaborative learning environments.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Interactions designed to support collective motivation in medium and long-term collaborative learning projects remain underexplored, despite their importance. Integrating motivational elements into these environments can boost learner engagement and collaboration, especially when considering individual differences and their evolving impact on group motivation [1]. Project-based learning (PjBL) boosts teamwork and maximizes the team's overall performance, creativity, and communication skills, which are vital for the 21st-century job market [2-4]. It fosters emotional intelligence, leadership, and interpersonal skills, enhancing collaboration and preparing students for career challenges. For instance, when users use the smiley emojis in communication with teams, it has a positive impact on psychological safety. Psychological safety in higher education (HE), grounded in trust and mutual respect, is a key pillar of team performance, including interpersonal communication, team learning, and creativity (e.g. [5] [6]). Project Aristotle by Google showed that high-performing teams thrive in psychologically safe environments where members freely share ideas [7]. It highlights emotional intelligence as key to creating psychological safety crucial for collaboration [8]. Project Aristotle showed that top teams rely on empathy, curiosity, and mutual respect. This principle aligns with the collaborative spirit of PjBL, where learners must interact flexibly and constructively to achieve shared goals [3]. Emotional intelligence plays a central role in creating such environments, enabling team facilitators or student leaders to foster spaces conducive to both individual growth and group achievement. Effective collaboration in learning activity review processes, particularly in PjBL, often faces challenges stemming from emotional reactions such as impulsive feedback, unconstructive comments, and the dismissal of peer suggestions.

These issues can disrupt communication, weaken team cohesion, and negatively impact the learning outcomes of the project. Additionally, one significant problem arises from distracted feedback during the review of learning activities, where participants might fail to provide focused and meaningful comments due to multitasking or external distractions, further hindering the learning process. While collaboration is essential for scaling PjBL initiatives, integration and coordination challenges persist. Existing methods often lack clear guidance for enforcing academic standards and fail to provide stakeholders with timely access to critical learning materials, resulting in delays and missed learning opportunities. Persistent gaps also remain in managing and resolving task conflicts and in supporting cross-functional student teams. Many current approaches overlook inter-team dependencies, which can extend project timelines. Moreover, cognitive biases, cultural differences, and distractions during feedback processes can affect peer reviews and documentation, further reducing collaboration efficiency. In addition, enhanced team-work, group learning, and cultural appreciation face psychological barriers linked to cultural diversity, making international environments more challenging. The present work is guided by the main research question: RQ: How can we reduce cognitive bias in Project-Based Learning during learning activity reviews? To address persistent challenges in the review of learning activities, such as miscommunication, cognitive bias, and inconsistent or unempathetic feedback, we present EmpaText, an advanced system that leverages psychological and cognitive science insights to suggest models for reducing human cognitive bias [9-11]. EmpaText is based on Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) to support student teams in PjBL. It provides actionable guidance for embedding LLMs into collaborative learning environments, fostering deep, unbiased, and meaningful reflection on student work.

EmpaText delivers focused, context-aware, and emotionally intelligent feedback that promotes constructive peer interaction and reduces misunderstandings. Its intelligent design enhances collaboration while supporting psychological well-being, fostering a more inclusive and effective review environment. (i) Activity Enhancer: Supporting learning through improved learning deliverables. (ii) Review Facilitator: Guiding assessment and promoting feedback quality. (iii) Bias Mitigator: Encouraging reflection to reduce cognitive and emotional biases. Real-world use cases demonstrate how EmpaText enhances peer feedback, not only ensuring pedagogical soundness but also fostering constructive dialogue and reducing issues like confirmation bias and decision fatigue. Based on these insights, we explore broader implications for integrating LLMs in project-based learning and offer recommendations for future research and practice. EmpaText ultimately provides a systematic, AI-powered approach to activity review, equipping learners with the tools to improve their outputs while cultivating a fair and collaborative educational culture. Our solution integrates with tools like Trello and Moodle, enhancing feedback quality in collaborative learning. Validated through extensive testing, it provides a scalable, user-centered approach. Despite the rise of LLMs in digital education, students and instructors still lack strategies for effectively integrating these tools in activity reviews. We start in Section 2 with a preliminary and motivating scenario to highlight the impact of cognitive biases during Project-Based Learning (PjBL). We then present related work in Section 3. In Section 4, we introduce our framework. Section 5 presents the evaluation of our framework. Finally, we conclude the paper in Section 6.

II. MOTIVATION AND CHALLENGES

Project-Based Learning (PjBL) involves a variety of participants learners, instructors, and external collaborators aiming to inspire individuals to engage in activities like collaboration, social exchange, team innovation, and different methods of verbal communication [12]. Centered on real-world projects, this method blends phases such as planning, collaborative work, documentation, and assessment. Several actions are accompanied by feedback production, such as peer review, communication, file exchange, and resource utilization, which occur throughout the learning cycle. The PjBL approach is learner-centric and cultivates essential soft skills namely collaboration, effective communication, and critical thinking which are vital for project success. Nevertheless, inadequate coordination can hinder progress, lowering engagement and triggering conflicts within teams. To maintain project momentum and meet pedagogical objectives, it is important to encourage collaborative practices, offer targeted guidance, and analyze team interactions. Such support mechanisms not only ease learners' cognitive load but also enable better monitoring and management of group dynamics throughout the PjBL's lifecycle. A key challenge lies in identifying motivational interactions suited to collaborative learning environments, addressing both individual and group levels. Rooted in psychological motivation theories, these interactions may include game mechanics (e.g., competition, progress, performance sharing), motivational messages, or interactive visualizations. While some support social and belonging needs, they are often designed for individuals rather than the group as a whole.

II.1 TEAM COMMUNICATION SCENARIOS

As shown in Figure 1, in project-based learning, feedback production can be viewed as a behavior shaped by intention. According to the Theory of Reasoned Action [13] intention is influenced by: (i) Attitude toward the behavior (driven by beliefs and evaluation of consequences), and (ii) Subjective norms (based on beliefs and perceived social expectations). Effective behavior results from this intention. Destructive feedback during learning activity reviews can stem from cognitive biases triggered by psychological and social factors of the reviewers. This lack of psychological safety arises from the fear of negative consequences when taking interpersonal risks. In PjBL team communication scenarios, biases and fears can impact the feedback process, making team members hesitant to provide constructive feedback due to concerns about straining relationships or facing criticism. Cognitive bias refers to systematic patterns of deviation from norms or rationality in judgment, leading to illogical inferences about people and situations [14]. Cognitive biases are distortions in thinking that shape behaviors, beliefs, and decisions [9-11]. Rooted in heuristics [15].

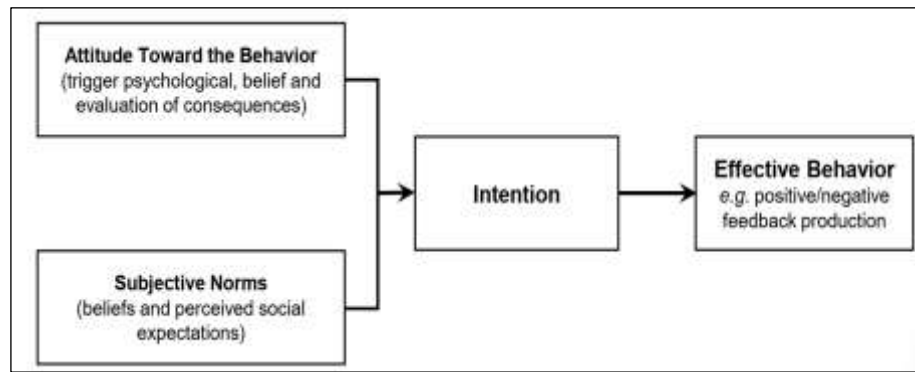


Figure 1: Intention, Subjective Norms, and Attitude as Determinants of Behavior.
Source: Authors, (2026).

Haselton et al. [16]; Kahneman and Tversky [17], they simplify decision-making but can result in inaccurate perceptions. *Emotional biases*, for instance, influence how we perceive and react to situations, often leading to irrational decisions. One example of cognitive bias is confirmation bias, where individuals tend to favor information that confirms their preconceptions, regardless of whether the information is accurate [14]. Additionally, decision fatigue can impair judgment after a series of decisions are made, leading to errors in evaluation. This bias can be especially prevalent in high-stakes environments, such as software development, where developers may make impulsive decisions based on incomplete data. Awareness of cognitive biases is essential to mitigate their impact, especially in collaborative settings where multiple perspectives are necessary. The use of structured frameworks and strategies, such as peer reviews and automated testing, helps reduce the influence of cognitive biases in decision-making. Understanding and managing these biases are critical in creating effective workflows in both software development and daily tasks.

II.2 CHALLENGES

Collaborative learning, where learners work together to achieve shared goals, fosters positive interdependence and enhances outcomes. The quality of peer interaction is crucial for knowledge development, performance, and positive emotions. A key challenge in this context is managing fluctuating motivation and collaboration within teams. Peer interaction quality, whether synchronous, asynchronous, co-located, or remote, is essential to group effectiveness, knowledge development, performance, and positive social outcomes. Additionally, the quality of communication and feedback plays a critical role in determining group effectiveness, learning outcomes, and overall well-being. Addressing these challenges requires integrating insights from psychology and ergonomics to design adaptive, motivation-driven learning environments. As shown in Figure 2, Learning Behavior Indicators (LBI), related to student behavior such as *performance expectations* and *effort expectations* and Content Progress Indicators (CPI), such as *content quality expectations* and the *outcome of the learning activity*, are two key factors that influence the intention to adopt an effective behavior, which may result in either positive or negative feedback [13]. As illustrated in the Figure 2 below, potential biases can influence users' intentions. Usually, the factors that influence and initiate biases that produce negative feedback can be categorized into four main types: *cognitive biases*, *perceptual biases*, *social biases*, and *con-textual biases*. Cognitive biases such as decision fatigue and anchoring bias reduce judgment accuracy, causing evaluators to provide positive/negative feedback based on limited or initial impressions. Perceptual biases, like confirmation bias, occur when evaluators selectively

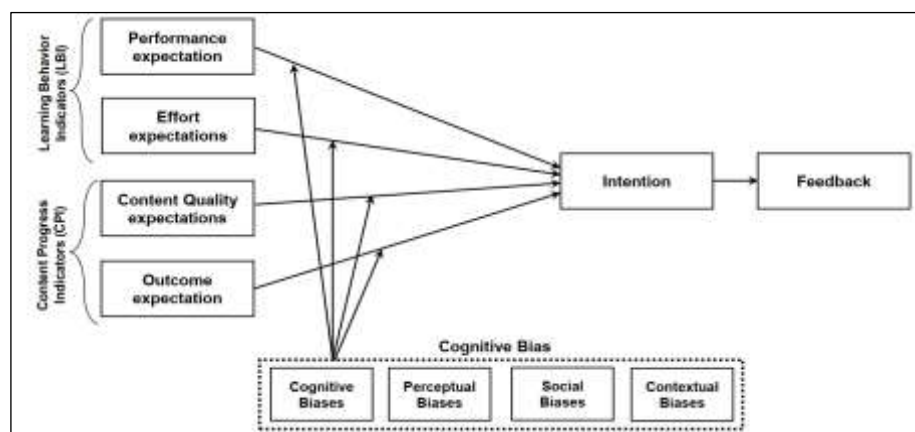


Figure 2: Factors influence the user's intention during the review and feedback production, with biases triggered by psychological and social factors.
Source: Authors, (2026).

Focus on information that confirms pre-existing beliefs, ignoring contradictory evidence. Social biases are shaped by social dynamics and power structures, with hierarchical judgments and social influence biases leading to feedback influenced by authority or group conformity, often resulting in harsher or skewed evaluations. Finally, contextual biases emerge from personal or environmental factors, such as stress or fatigue, which can distort the feedback process and contribute to overly negative assessments.]

These biases collectively shape how feedback is delivered and can lead to a disproportionate focus on negative aspects. Based on the discussion, this paper is a first step in reducing cognitive bias in PjBL. It identifies three key challenges to enhance psychological safety, team interpersonal skills, and communication.

- C1. How to support students in improving the quality of their learning deliverables?
- C2. How to facilitate peer review and ensure the production of constructive, high-quality feedback?
- C3. How to reduce the impact of cognitive and emotional biases during evaluation and feedback?

III. RELATED WORK

Project-based learning (PjBL) has emerged as an effective approach for developing both hard and soft skills. Studies explore its role in enhancing soft skills among informatics students through agile frameworks, emphasizing its impact on communication and interpersonal skills [18]. In another line of research, the authors Daineko et al. [15] further highlight the integration of hard and soft skills in PjBL, enhancing students' ability to apply theoretical knowledge to practical problems. The integration of LLMs within PjBL has also been proposed, fostering autonomy and motivation [19] while personalized learning systems using generative AI for feedback have been introduced for computer science education [20]. The role of guiding questions in facilitating structured feedback and collaboration in PjBL has been explored [21], and the impact of peer feedback on student engagement and critical thinking has been studied [22]. At the same time, several studies propose solutions to enhance Project-Based Learning (PjBL) in an English as a Foreign Language (EFL) context, particularly by promoting writing skills and peer feedback [23]. The use of digital tools for iterative design and real-time feedback in architectural education has been high-lighted [24], while the role of PjBL in developing cognitive and non-cognitive skills in higher education has been reviewed [25]. Emotional engagement during pedagogical practice and its impact on teaching competencies has been examined [8]. Additionally, studies explore disruptions in student teams during PjBL, focusing on group dynamics and collaboration challenges [26].

Collectively, these studies highlight the broad benefits of PjBL, from technical skills to emotional intelligence, with feedback mechanisms playing a crucial role in student success. Other works explore how PjBL promotes critical thinking, collaboration, and autonomy [27]. While students are encouraged to self-direct their learning, effective coaching remains necessary [28]. This shift redefines teachers as facilitators, requiring new professional behaviors and emphasizing authentic, collaborative tasks [29]. However, educators often feel unprepared for this role, especially in giving constructive feedback [30]. Students and graduates frequently critique feedback quality, and despite learner-centered aims, feedback remains largely teacher-driven and judgmental [31]. In a similar trend, in Project-Based Learning (PjBL), coaches support student autonomy, confidence, and performance through guided reflection and feedback (e.g., [32] [29] [33] [34]). Didactic coaching, as noted in [35] emphasizes questions and process-oriented feedback to foster motivation and self-regulation. Feedback is most effective when it is specific, constructive, and dialogic [36]. Its impact varies depending on whether it targets the task, learning process, or self-regulation [34]. By exploring the literature, and to the best of our knowledge, no existing work proposes a dedicated LLM model aimed at reducing cognitive bias in Project-Based Learning (PjBL). This proposal is intended to initiate efforts toward enhancing interpersonal and communication skills in PjBL. Similar efforts have been conducted using LLMs for bias testing and mitigation, particularly in code generation and in investigating social bias in LLM-generated outputs [37] [38].

IV. BIAS-AWARE LEARNING ACTIVITY REVIEW FRAMEWORK

Based on the identified challenges and theoretical foundations, we propose the Framework (that we called EmpaText), which defines three key core components: (Figure 3): **(i) Activity Enhancer**, **(ii) Review Facilitator**, and **(iii) Bias Mitigator**. These services are designed to enhance Project-Based Learning collaboration between students by encouraging them to provide positive feedback during PjBL, engage with GenAI across different learning contexts, and accept revisions to improve modifications in learning activities. EmpaText provides clear, actionable guidance to submit activities through functional descriptions, such as the Canvas model, which supports the efficient submission and delivery of learning activities. On the other hand, EmpaText suggests supporting actors to quickly evaluate activities as challenges and provide constructive feedback to reduce cognitive bias face psychological barriers in PjBL. EmpaText is an intelligent system designed to reduce cognitive biases in Project-Based Learning (PjBL), specifically addressing core psychological aspects such as confirmation bias and decision fatigue. EmpaText is based on generation and selection techniques, as well as Retrieval-Augmented Generation (RAG) methods. The system first leverages LLMs to generate and reformulate constructive feedback, which is then refined using an existing knowledge base to ensure alignment with psychological and educational theories, forming a high-quality knowledge repository. Subsequently, RAG technology integrates this knowledge base with LLMs, enabling precise and context-aware cognitive bias mitigation through embedding-based retrieval and a domain-specific psychological LLM. Ultimately, the system provides students with an interactive interface for communication support, helping them navigate psychological challenges and fostering emotional well-being.

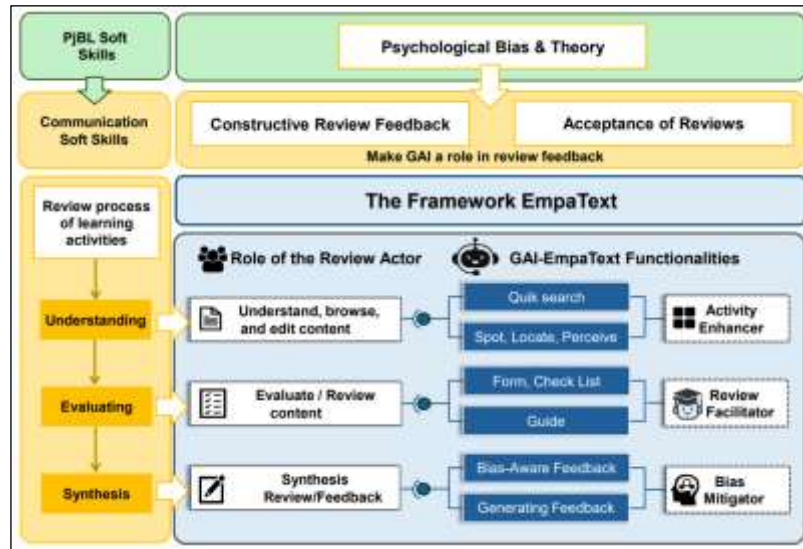


Figure 3: Empa Text Overview.
Source: Authors, (2026).

IV.1 ACTIVITY ENHANCER

In this section, to address the first challenge C1, the "Activity Enhancer" service proposes a canvas model to be displayed according to a domain-specific language. This model highlights changes in students' efforts, allowing for quick perception of progress made by student teams in their learning activities. It helps to avoid the effects of fatigue and also showcases the elements that demonstrate team effort. Additionally, the service is designed to address the problem of confirmation bias, where individuals may perceive that no progress has been made on the project, despite evidence to the contrary. By clearly displaying the team's efforts, it counters the tendency to overlook small but important contributions, helping to prevent biased judgments about team performance. Table 1 illustrates our proposed "Review Support Canvas" model, designed to mitigate cognitive biases and decision fatigue during the evaluation process in PjBL. This model serves as a structured template for students to submit activities in an organized manner while also guiding reviewers through the assessment process with a clear set of learning objectives, team roles, and recent changes. By emphasizing attention indicators and summarizing insights, it helps reviewers stay focused and avoid distractions that could skew their judgment. Cognitive biases such as confirmation bias and anchoring often distort evaluation outcomes, but this canvas model aims to provide a more objective, methodical approach. It also helps reduce decision fatigue by breaking down the review into manageable sections, enhancing clarity and ensuring comprehensive feedback. Ultimately, this approach ensures a more balanced and accurate evaluation, fostering a fairer learning environment and supporting the continuous development of students' skills.

Table 1: Review Support Canvas for PjBL Evaluation.

Element	Description
Summary	Helps reviewers assess learning activities in PjBL while reducing cognitive bias and decision fatigue.
Learning Objectives	Defined objectives ensure focused evaluation on key learning outcomes.
Completed Activities	Provides context through a summary of finished tasks.
Team Roles and Interactions	Highlights team roles and collaboration dynamics.
Recent Changes	Keeps reviewers informed about recent project modifications.
Attention Indicators	Visual or textual cues to focus on key elements.
Summary Insight	Quick reference of key findings for reviewers.
Integration Context	Explains how the activity fits into the overall project.
Benefits	Reduces cognitive bias and decision fatigue, improving feedback quality.

Source: Authors, (2026).

IV.2 REVIEW FACILITATOR

To tackle Challenge C2, the "Review Facilitator" allows reviewers in Project-Based Learning (PBL) to quickly evaluate learning activities. It provides a checklist solution that enables reviewers to efficiently browse and assess the learning activities of teams within a project-based learning context. The review questions checklist is a set of questions related to the evaluation of learning activities. Each question belongs to one of the following categories: activity understanding, activity importance, activity difference, activity arguments, and activity progress. Users answer each question by choosing exclusively: *true*, *false*, or *undefined*. Here is an example of a source code description evaluated by the values of four checklist questions related to error semantics, readability, and structure. The evaluation result of a rule can be: $\checkmark$: true, $\boxtimes$: false, or Q: undefined. Therefore, the answer can either be true or false. Figure 4 shows example of checklist-based source code evaluation, including perceived effort and reviewer invitation.

Q ₁ Is the source code structured?	✓
Q ₂ Is the source code clear and commented?	☐
Q ₃ Are changes in the workflow expanded?	☒
Q ₄ Is the source code correct?	☒
Q ₅ Can the effort in the learning activity be perceived?	✓
Q ₆ Should other reviewers be invited to evaluate the activity?	☐

Figure 4: Example of checklist-based source code evaluation, including perceived effort and reviewer invitation.
Source: Authors, (2026).

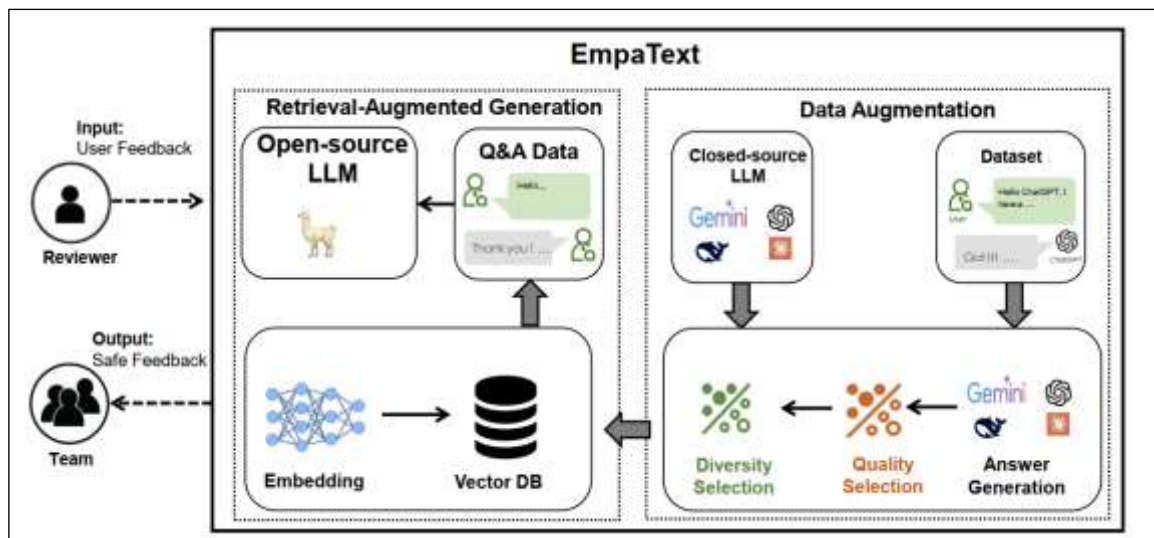


Figure 5: Process of Empa Text Framework.
Source: Authors, (2026).

IV.3 BIAS MITIGATOR

To tackle Challenge C3, our framework EmpaText is an intelligent constructive feedback system designed to support teams during PjBL, structured around reducing cognitive bias based on psychological theories and biases. *Bias Mitigator* automatically detects each appearance of bias based on a semi-supervised algorithm for bias identification, then *Bias Mitigator* replaces it with corresponding constructive feedback. Our Semi-supervised Algorithm for Bias Identification employs two modules: (i) semi-supervised pattern acquisition for identifying biased patterns that serve as features for a classifier, and (ii) a classification algorithm that classifies each sentence into a biased class. This service generates clean, unbiased feedback through careful data generation, selection, and Retrieval-Augmented Generation (RAG) techniques. EmpaText leverages closed-source LLMs to process Q&A pairs "Original Feedback and Clean Feedback" which are then refined using an existing database to ensure alignment with team motivation aspects, forming a high-quality knowledge base. Subsequently, RAG technology integrates this knowledge base with LLMs, enabling precise and context-aware feedback through embedding-based retrieval and a domain-specific psychological LLM. Ultimately, the system provides teams with an interactive interface for giving and receiving feedback, helping them navigate psychological bias challenges and fostering emotional well-being. More formally, a *Bias Mitigator*, operating on the ECA (Event-Condition-Action) paradigm, is seamlessly integrated to enhance student communication and team performance in PjBL. The Bias Mitigator in the EmpaText framework is a rule-based system designed to enhance student engagement by delivering personalized feedback, motivational messages, and recommendations. It operates as follows:

- **Event:** This phase corresponds to the detection of specific linguistic markers indicating cognitive bias in reviewers' behavior, such as missed deadlines or low activity levels (e.g., a student misses the earliest start time for a task).
- **Condition:** The framework evaluates the feedback text by analyzing linguistic markers such as task completion rate and response time. If the feedback is identified as destructive, the condition is met.
- **Action:** Based on the feedback's detection state, the framework generates appropriate feedback, a motivational message, or a recommendation from a predefined catalog organized into categories and subcategories

In the next section, we explain the steps to build the Bias Mitigator.

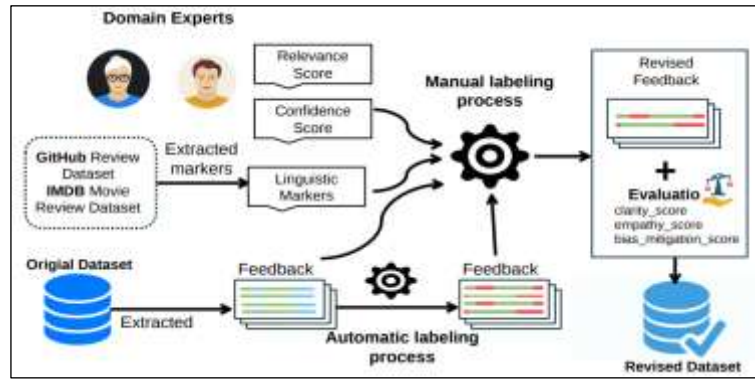


Figure 6: Building Our Dataset (CoFeedBack) Process: Data Pipeline - Annotation, Revision, and Evaluation.
Source: Authors, (2026).

IV.3.1 Building Our Dataset

To support the development and evaluation of *EmpaText*, we built a composite dataset by combining multiple sources annotated with emotional and cognitive dimensions. The goal is to train LLMs to generate empathetic and unbiased feedback in PjBL settings. Figure 6 shows the process of building our dataset (CoFeedBack: Constructive Feedback Dataset). In particular, our workflow for building the dataset consists of five stages (Figure 6): Generating the initial dataset, Extracting linguistic markers for bias detection, Classifying bias, Automatically labelling data using semi-supervised learning with K-Nearest Neighbors (KNN), and Manually labelling data by domain experts. We detail these steps in the following section.

1. Generating the initial dataset

We integrated feedback and evaluative texts from diverse sources. The datasets used include:

- The **GitHub Review Dataset (OpenReview)**¹, which features emotionally nuanced developer comments and peer-reviewed code interactions to simulate feedback scenarios in collaborative environments.
- The **IMDB Movie Review Dataset**², which offers sentiment-rich, user-generated content, making it valuable for training models to recognize emotional tone. If an emotion reveals that a person is upset, we seek to analyze the text that triggered this emotional reaction, in order to better understand the underlying causes and improve communication.
- The **Academic Writing Revision Dataset** comprises paragraph-level revisions paired with explicit instructions (e.g., “make it more constructive”), supporting tasks related to transforming feedback. Finally, the **CoFeedBack Dataset** generated by closed source LLM. The **CoFeedBack Dataset** contains peer feedback samples annotated with labels indicating the presence or absence of cognitive biases, such as generalization, harshness, and unsupported claims.

2. Extracting linguistic markers for bias detection

We extracted linguistic features from all sources to guide model training and evaluation. A small set of manually labelled linguistic markers, scored on a discrete scale with values from 1 to 5, was assigned to indicate the clear presence of bias.

- Emotional tone and sentiment polarity
- Cognitive bias indicators (e.g., absolutist language, lack of evidence)
- Pragmatic and discourse markers (e.g., hedging, politeness strategies)

Example 1 (Linguistic Markers Indicating Cognitive Bias). *Table 2 summarizes common linguistic markers that signal cognitive bias or excessively critical feedback in review sentences appearing in the body of the review. It categorizes typical patterns such as absolutism (“always,” “never”), excessive doubt (“no evidence that”), negative comparisons (“unlike others”), and vague judgments (“poorly done”). These markers often reflect subjective or harsh evaluations that can undermine constructive critique.*

Table 2: Common Linguistic Markers Indicating Cognitive Bias or Overly Critical Feedback.

Bias Type	Typical Markers
Absolutism	always, never, must, cannot
Excessive doubt	unclear if, no evidence that
Negative comparison	unlike others, fails to
Condescending tone	obviously, trivial, clearly wrong
Vague judgment	weak work, poorly done, low quality
Overfocus on flaws	fatal flaws, major issues, severe problems
Subjectivity	I don’t like, seems bad to me
Perfectionism	must be flawless, complete coverage necessary
Non-essential style focus	phrasing unacceptable, bad structure

Source: Authors, (2026).

¹ https://github.com/openreview?utm_source=chatgpt.com

² <https://www.imdb.com/>

3. Our bias classification method is based on a classic semi-supervised learning approach. During the training phase, we were given a small set of manually labeled sentences (seeds). A discrete score ranging from 1 to 5 was assigned to each sentence, where 5 represents a definitely biased sentence, and 1 represents a clear absence of bias. Given these labeled sentences, we extracted a set of features to construct feature vectors. We utilized two types of basic features: Syntactic features, Pattern-based features. To overcome the sparsity of sarcastic sentences and avoid labor-intensive, wide-scale annotation, we used search engine queries to acquire more examples, thereby expanding the training set automatically. We then constructed feature vectors for each labeled example in the expanded training set. These vectors were used to build the classification model, which assigns bias scores to unlabeled examples.

4. Automatically labeling data using semi-supervised learning

In semi-supervised learning, the model is trained using a small set of labeled data and a larger set of unlabeled data. One of the most common techniques used in this setting is K-Nearest Neighbors (KNN). The basic idea behind KNN is that data points that are similar to each other are more likely to share the same label. Thus, the task of classifying unlabeled points is performed by examining the majority class of their K-nearest neighbors. The Euclidean distance metric is typically used to calculate the similarity between two data points. The Euclidean distance between two feature vectors $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ is given by the following formula:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Where:

$d(x, y)$ = Euclidean distance between vectors x and y

x_i = the i^{th} feature of vector x

y_i = the i^{th} feature of vector y

n = number of dimensions (features)

Once the Euclidean distances between the labeled points and the unlabeled points are computed, the K-Nearest Neighbors algorithm assigns the label of the majority of the K nearest neighbors to the unlabeled point. In this method, the value of K is a hyperparameter that controls the number of neighbors considered for classification.

$$\hat{y} = \text{majority}(\{y_1, y_2, \dots, y_K\})$$

Where:

- $\hat{y}$ is the predicted label for the unlabeled data point.
- $y_1, y_2, \dots, y_K$ are the class labels of the K-nearest neighbors.
- K : number of nearest neighbors considered (hyperparameter)
- $\text{majority}()$ = function that selects the most frequent class label among the K neighbors

By expanding the training set using semi-supervised learning, we can leverage a small amount of labeled data and a large amount of unlabeled data to effectively train a model for bias classification.

5. Manually labeling data by domain experts

A psychologist manually labeled sentences using an evaluation grid to assess the emotional and cognitive dimensions of feedback, along with its clarity, empathy, and constructiveness. The evaluation also involved identifying the presence and type of cognitive biases. Data revision by domain experts involves selecting actions from predefined types (ERelevanceScoreType and EConfidenceScoreType) reflecting trust decisions and confidence levels (e.g., "I'm sure," "I'm doubtful"). Each revision has two indicators:

- Relevance, annotated as NOT, RELEVANT, or VERY (with VERY indicating high-quality insights).
- Confidence, annotated as NOT or CONFIDENT, where NOT indicates low confidence.

The data pipeline included three main stages: **annotation** of raw feedback, **revision** of the feedback (either by humans or generated by large language models) based on specific instructions, and **evaluation** of the revised feedback using predefined criteria. To ensure annotation reliability, inter-annotator agreement was measured, with a Cohen's Kappa score greater than 0.75. Each data instance was encoded with the following fields (Figure 7):

The *EmpaText* dataset provides a comprehensive and diverse set of annotated feedback examples that can be used to train models for generating empathetic and unbiased feedback in PjBL settings. The bias classification method, based on semi-supervised learning and feature extraction, enables effective identification and mitigation of cognitive biases in feedback. By revising the data with domain experts and evaluating it using predefined criteria, we ensure the quality and effectiveness of the feedback in fostering constructive and balanced learning environments.

```

{
  "original_text": "...",
  "bias_label": "confirmation_bias", "emotion": "frustration",
  "instruction": "make it more constructive and balanced",
  "revised_text": "...",
  "evaluation": {
    "clarity_score": 4,
    "empathy_score": 5,
    "bias_mitigation_score": 4
  }
}

```

Figure 7: Example of a bias-mitigated text annotation with evaluation scores.

Source: Authors, (2026).

IV.3.2 Dataset Augmentation

The construction of a reliable knowledge base for generating constructive feedback particularly from software engineering activities such as source code analysis requires a carefully designed data augmentation strategy. We propose a three-phase augmentation pipeline: (i) a LLM is prompted using predefined instructions to produce diverse candidate feedback outputs; (ii) existing feedback-response pairs are used to train the LLM to serve as both an extractor and a discriminator, ensuring that the selected responses adhere to professional standards and maintain domain consistency; and (iii) algorithmic diversity analysis is applied to reduce redundancy and retain high-quality, varied outputs. The enriched dataset produced through this process forms the basis for integrating Retrieval-Augmented Generation (RAG) into the intelligent feedback system within PjBL environments.

IV.3.3 Dataset Generation

To address mental health questions that lack predefined answers, we leverage a LLM to generate potential responses. To ensure that the model understands the context, we explicitly define the consultation scenario as one focused on adolescent mental health support, guiding the model to act as a psychological expert dedicated to fostering healthy development. Additionally, we apply stylistic constraints, requiring responses to be optimistic, aligned with psychological bias theories, and expressed in a warm and empathetic tone. To further enhance the model's comprehension of the task, we use few-shot learning by providing representative Q&A examples from existing data, formulated as *Original Feedback* and *Clean Feedback* pairs. With this structured approach, the LLM is able to generate multiple context-aware responses tailored to the psychological support needs of teams in PjBL.

IV.3.4 Quality Data Selection

To maintain the quality and professionalism of generated feedback, we implement a systematic filtering process grounded in key content criteria. The main objective is to evaluate critical dimensions of each response such as core message clarity, theoretical relevance, and emotional tone to ensure they meet predefined quality standards. Initially, we use a labeled dataset to train a key point extractor using a LLM, which identifies and summarizes the most important elements from each Q&A instance. Following this, a key point evaluator also based on an LLM is employed to verify whether the extracted elements align with relevant psychological and pedagogical frameworks. Responses that satisfy these benchmarks are considered high-quality and retained for use. This dual-model approach enables us to effectively filter and preserve responses that are coherent, professionally structured, and grounded in theory, ensuring high reliability in the feedback system used within PjBL.

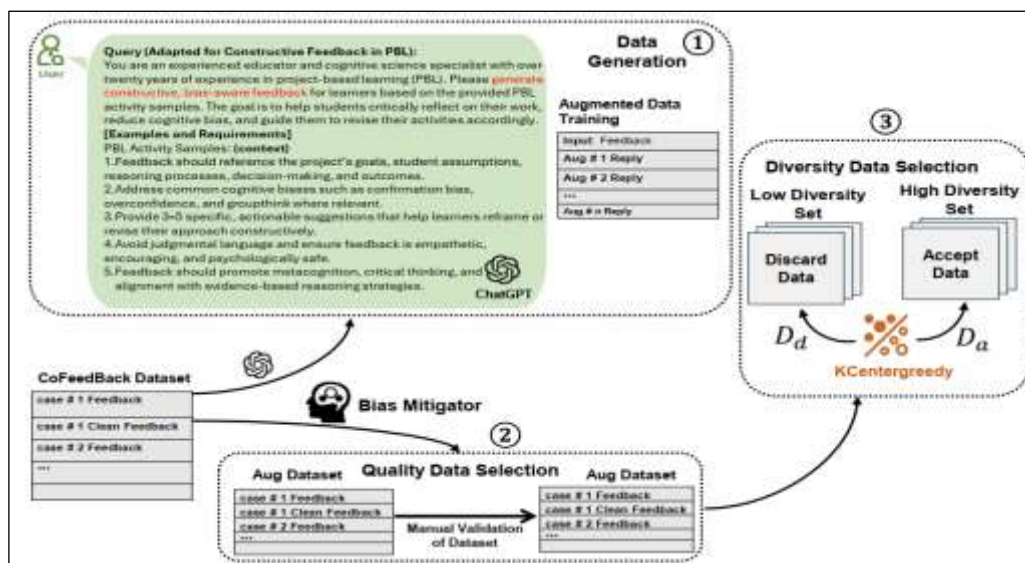


Figure 8: Pipeline for Data Augmentation, including (1) the Data Generation Component, (2) the Quality Selection Component, and (3) the Diversity Selection Component.

Source: Authors, (2026).

IV.3.5 Diversity Data Selection

To further enrich the dataset, we adopt a diversity-driven selection strategy designed to maximize variation across the retained responses. This step helps ensure that the resulting dataset captures a wide range of viewpoints, minimizes redundancy, and supports diversity in feedback content. Our approach involves identifying answers that are most distinct from others based on evaluations made by a discriminator model. To perform this selection efficiently, we utilize the KCenterGreedy algorithm [39], [40] which is particularly effective at selecting a diverse representative subset while remaining computationally feasible. By incorporating this strategy, we generate a synthetic dataset that is both high-quality and diverse, thereby strengthening the adaptability and robustness of the constructive feedback system.

IV.3.6 RAG-based Feedback

The Retrieval-Augmented Generation (RAG) framework is designed to improve the performance of LLMs by incorporating external, domain-specific knowledge sources. This integration helps overcome common limitations of standard LLMs, such as insufficient domain expertise and the potential for generating imprecise or unreliable answers. The RAG process is structured around three key stages: indexing, retrieval, and generation. Each stage contributes to ensuring that the system produces responses that are accurate, contextually relevant, and professionally appropriate. In this section, we describe how these components are implemented within our feedback Q&A system.

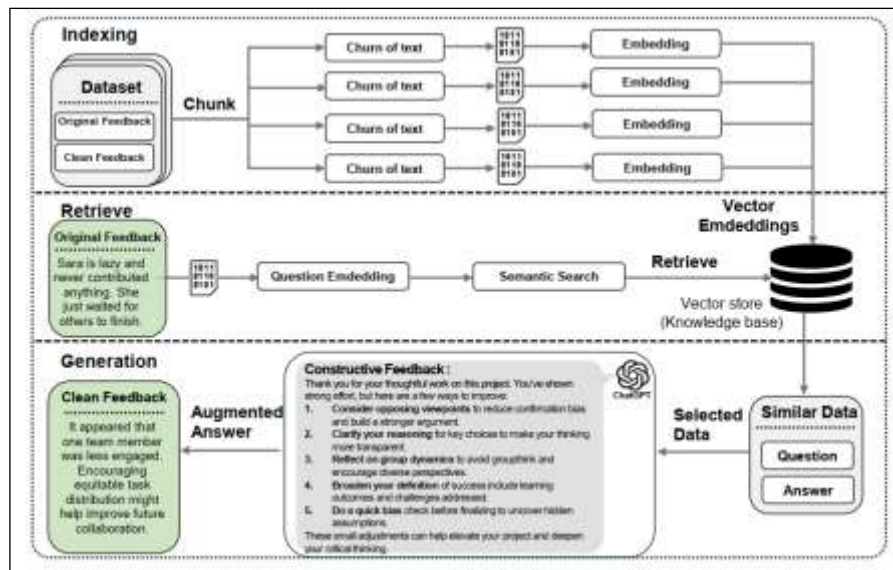


Figure 9: Workflow for Retrieval-Augmented Generation and System Execution, including (1) the Indexing Unit, (2) the Retrieval Unit, and (3) the Generation Unit.

Source: Authors, (2026).

IV.3.7 Indexing

The indexing phase serves as the foundation of the RAG architecture, enabling efficient access to relevant information during query processing. In this stage, the dataset composed of question-answer pairs is segmented and transformed into vector representations using an embedding model. These embeddings capture the semantic content of each text segment and are stored in a dedicated vector database, such as FAISS or Pinecone. This structure allows for rapid similarity-based retrieval, ensuring that the system can effectively locate contextually appropriate information in response to user queries.

IV.3.8 Retrieval

The retrieval phase is crucial for identifying the most pertinent questions from the indexed database in response to a user's query. Upon receiving a query, the system first vectorizes it using the same embedding model applied during the indexing stage. This ensures that the query is represented within the same vector space as the indexed cases, enabling effective similarity comparison. The vectorized query is then used to conduct a similarity search within the vector database. The results are ranked based on their similarity to the query, with the most relevant entries appearing at the top. This ranking process ensures that the most suitable questions are prioritized, laying the groundwork for generating precise and insightful responses.

IV.3.9 Generation

The final phase of the RAG framework involves generating the response. This stage combines the answers from the top-ranked questions retrieved in the previous step with predefined prompts designed to guide the LLM in crafting the final response. These prompts are carefully structured to ensure that the generated responses are professionally reviewed, mindful of cognitive biases, and provide constructive, actionable feedback to the user.

By combining the retrieved answers with these guiding prompts, the LLM can leverage both the external knowledge from the database and its own generative capabilities to create high-quality, contextually appropriate responses. This method not only improves the accuracy of the output but also ensures that the responses are specifically tailored to meet the user's needs.

V. FRAMEWORK EVALUATION

Through this case study, we aim to provide valuable insights into the benefits and challenges of assisting PjBL teams in enhancing communication. First, we describe the evaluation scenario, detailing EmpaText's performance. Then, we present the main functionalities of the support tool. Finally, we present the evaluation of the EmpaText system's usability.

V.1 TECHNICAL IMPLEMENTATION

The proposed system is designed to assist PjBL teams in communication by reducing cognitive bias using natural language. The architecture consists of several interconnected modules. The system provides a user-friendly web console where teachers can input natural language questions. These questions are processed through ChromaDB for contextual retrieval and few-shot prompting. To ensure consistent responses from the language model, the input questions are enriched with predefined templates. An LLaMA3 model then generates relevant and constructive feedback. Our dataset is elaboration-based, drawing mainly from the GitHub Review Dataset (OpenReview)³, which captures 4,818 emotionally nuanced developer comments and peer-reviewed code interactions. We also incorporate the IMDB Movie Review Dataset (OpenReview)⁴, which includes over 50,000 sentiment-rich, user-generated reviews, useful for training emotion recognition models. When emotions such as frustration are detected, we analyze the triggering text segments to better understand underlying causes and enhance communication. Additional resources include OpenReview's sentiment analysis data⁵ (comprising approximately 10,000 labeled instances) and the Llama-3.3-Nemotron-70B-Feedback dataset⁶, containing millions of synthetic feedback examples for advanced language model training.

V.2 EVALUATION OF EMPATEXT SYSTEM PERFORMANCE

To evaluate the performance of our RAG-based cognitive support feedback framework for PjBL teams, we implemented a two-phase assessment process. The first phase involved a quantitative evaluation of the quality of the generated data, which underpins the system's functionality. Expert ratings were used to assess the relevance and accuracy of the generated content. The second phase consisted of a qualitative comparison of the system's question-and-answer outputs with those generated by LLMs without external knowledge integration. This comparison was illustrated through selected examples to demonstrate the system's ability to produce responses that are both contextually relevant and professionally aligned.

V.2.1 Data Augmentation Evaluation

The quality of the generated data is essential for the overall effectiveness of the system. To assess this, we applied a set of widely recognized metrics in data augmentation. These metrics were established using a heuristic method to assign weights to each indicator. In particular, we selected four experts from the fields of communication, psychology, and cognitive science. We randomly selected 120 data points from the generated dataset and asked each expert to independently rate these data points according to predefined criteria. The scores provided by the four experts were then averaged for each data point, and the overall average score was computed across all 120 data points. To better demonstrate the improvement in data quality facilitated by our system, we compared our results with those generated directly by a large language model, which served as the baseline. The responses were assessed using a four-dimensional professionalism scale, with a total score of 10, based on the following criteria: Accuracy (4 points): Whether the response adheres to psychological theories or clinical standards. Practicality (2 points): Whether the advice is clear, specific, and actionable. Risk Response (2 points): Whether the response to high-risk issues is handled professionally and responsibly. Empathy (2 points): Whether the response conveys emotional support and empathetic expression. As illustrated in Table 6, our system outperforms the baseline across all evaluated dimensions. The improvements highlight the efficacy of the data augmentation and RAG framework in enhancing the quality of the generated responses, making them more accurate, practical, and empathetic. Notably, the largest improvements were seen in the "Risk Response" and "Practicality" categories, where the integration of external knowledge played a significant role in producing more thoughtful and actionable responses. This suggests that our approach effectively addresses critical challenges in providing mental health support through AI-driven feedback systems, ensuring not only relevance but also ethical and professional considerations.

V.2.2 Response Analysis

The quality of system-generated responses is crucial to its practical utility in real-world software engineering, particularly in delivering constructive feedback during source code reviews. To evaluate this, we selected representative examples and compared feedback generated by our RAG-based system with that of a baseline model (LLaMA LLM). These examples reflect typical code review scenarios involving issues such as code quality, readability, and maintainability. Table 7 presents a comparison of the feedback. These qualitative comparisons show that the RAG-based system offers more structured, detailed, and professional feedback. By integrating established software engineering practices, the system enhances the effectiveness of code reviews in

³ <https://github.com/Seafoodair/Openreview>

⁴ <https://huggingface.co/datasets/McAuley-Lab/Amazon-Reviews-2023>

⁵ <https://github.com/Seafoodair/Openreview/blob/master/data/>

⁶ <https://huggingface.co/nvidia/Llama-3.3-Nemotron-70B-Feedback>

PjBL environments. This not only improves code quality but also supports the development of students' problem-solving and engineering skills.

V.2.3 Bias Detection in Feedback with EmpaText

Using the labeled dataset, EmpaText's supervised learning model was trained to detect and mitigate bias in feedback. The model learns to identify patterns and linguistic cues in the feedback text that are associated with different types of bias. It then provides more neutral, constructive feedback by adjusting its output. The training process consisted of four main steps. First, data preprocessing was performed to clean and structure the feedback text, removing any unnecessary noise. Next, relevant features such as sentiment, language tone, and context were extracted to identify potential bias indicators. A classification model, such as logistic regression or decision tree, was then trained to predict the presence of bias in the feedback. Finally, the model's performance was evaluated on a test set of feedback entries that were not part of the training data. The model's performance was evaluated using key metrics such as accuracy, precision, recall, and F1 score to assess its ability to detect biases in feedback. Accuracy reflects the overall percentage of correct predictions, while precision indicates the proportion of predicted biases that were actual biases. Recall measures how well the model identifies actual biases, and the F1 score provides a balance between precision and recall. The results demonstrated the effectiveness of EmpaText's bias detection system. In this process, experts provided feedback on EmpaText's ability to detect biases, evaluating its effectiveness in mitigating bias and making appropriate adjustments. Their input was vital for refining the model's decision-making and ensuring accuracy and context sensitivity. EmpaText achieved an accuracy rate of 75% to 85% across various bias types, demonstrating its ability to provide fair, constructive, and unbiased feedback, crucial for fostering a supportive environment in code reviews and collaborative contexts. The expert feedback loop significantly improved the system's detection of nuanced biases. Moving forward, EmpaText will continue to enhance its accuracy and utility by learning from real-world feedback and expert assessments.

V.2.4 Precision for suggesting constructive feedback

To highlight the effectiveness of our framework in delivering appropriate constructive feedback to students, we collected data by observing feedback selection through system logs and evaluated EmpaText during the feedback suggestion process. Our study involved four scenarios, each corresponding to a different level of task difficulty: Easy, Moderate, Difficult, and Very Difficult. These scenarios were used to assess and refine the feedback recommendations generated by EmpaText based on the complexity of the learning tasks. The study was conducted with participants divided into three groups:

Group A: Providing feedback in which bias is explicitly identified in the text.

Group B: Providing feedback in which bias is moderated explicitly in the text.

Group C: Providing feedback in which bias is more deeply and subtly addressed in the text (in an implicit way).

The feedback was categorized based on the levels of bias identified in the provided responses.

Table 6: Average Precision (%) across feedback types and prompt difficulties.

Prompt Difficulty	A: Explicit Bias	B: Moderated Bias	C: Indirect Bias
Easy	0.970	0.900	0.850
Moderate	0.930	0.870	0.820
Difficult	0.890	0.840	0.790
Very Difficult	0.860	0.810	0.760

Source: Authors, (2026).

In our framework, we employ a set of prompts based on paired biased and clean feedback to recommend specific feedback models targeting various aspects of student performance, such as participation, assiduity, and presence. This approach relies on stored phrase types organized into categories, including challenge types, risk levels, and motivation types.

Table 7: Scores and Feedback for Participation, Assiduity, and Presence.

Score (%)	Feedback Category	Example of Constructive Feedback
85%	Code Correction	"Great participation! Continue contributing actively."
90%	Low Engagement	"Excellent attendance. Keep it up!"
75%	Submission Deadlines	"Try to increase your presence in sessions."

Source: Authors, (2026).

V.3. SUPPORTING TOOL

We developed a prototype tool dedicated to Project-Based Learning in software engineering for computer science. The tool is based on the EmpaText framework, aiming to enhance communication and avoid bias when students submit work, such as Java code, for review.

V.3.1 Dashboard Interface

The dashboard interface, shown in Figure 10, This dashboard includes a "Recent Activities" table that lists submissions with relevant metadata: student name, activity title, bias score (visually represented with a progress bar), submission date, and review status. Quick action buttons at the bottom enable users to either submit a new activity or review existing ones. This interface facilitates efficient navigation and status monitoring within the project-based learning workflow.

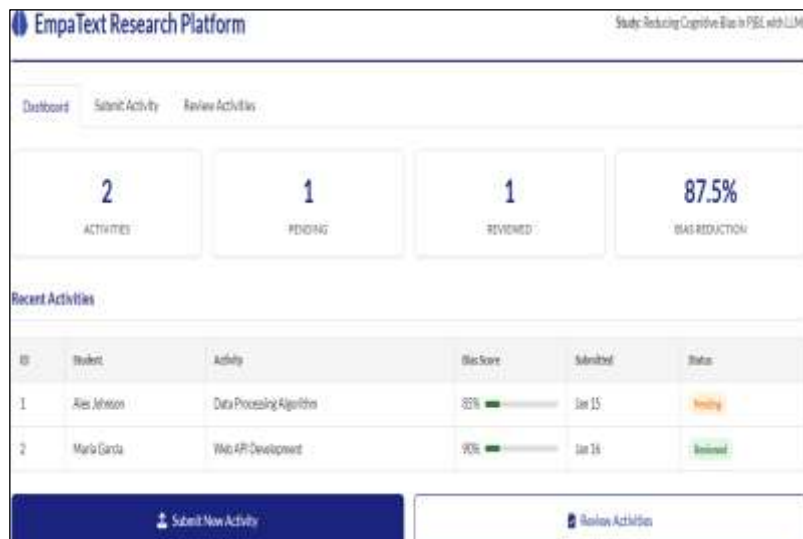


Figure 10: Dashboard Interface.
Source: Authors, (2026).

V.3.2 Activity Report Submission

The interface depicted in Figure 11.

Submit Learning Activity

Submit code with anti-bias templates to minimize cognitive bias in peer review

Dashboard Submit Activity Review Activities

Select Anti-Bias Template

Bias Reduction 85%
Structured
Focuses on architecture and design decisions

Bias Reduction 90%
Reflective
Encourages self-assessment and learning

Bias Reduction 95%
Collaborative
Highlights team contributions

Activity Details

Student Name:
Alex Johnson

Activity Title:
Data Processing Algorithm

Learning Objectives:
Learning sorting algorithms and data structures

Code Submission

Bias Reduction 85%

```
// EMPATEXT ANTI-BIAS TEMPLATE - STRUCTURED VERSION
// =====
// ACTIVITY: Data Processing Algorithm
// STUDENT: Alex Johnson
// DATE: 2024-01-15
//
// LEARNING OBJECTIVES:
// 1. Understand sorting algorithms
// 2. Understand time complexity (Big O notation)
//
// DESIGN DECISIONS:
// - Using quicksort for its in-place sorting
// - Iterative approach chosen for clarity
```

Back to Dashboard Submit with EmpaText

Figure 11: Activity Report Submission Interface.
Source: Authors, (2026).

V.3.3 Activity Review & Feedback Interface

The submitted code is presented in a syntax-highlighted panel with line numbers, allowing for precise review. A dedicated feedback section enables instructors to enter detailed comments and suggestions. This interface implements the EmpaText AI-assisted review system to help educators provide constructive, unbiased feedback focused on code quality and learning outcomes rather than subjective evaluation.

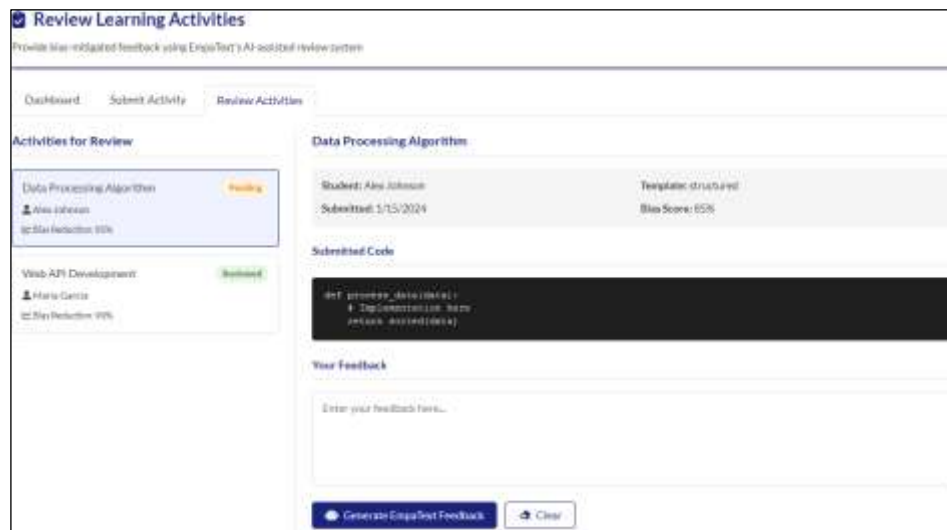


Figure 12: Detailed Activity Review & Feedback Interface.
Source: Authors, (2026).

V.3.4 Bulk Activity Review Interface

For managing multiple submissions, the platform provides a bulk review interface (shown in Figure 13).

a. Evaluation of EmpaText System Usability

To validate the effectiveness of the EmpaText framework within authentic educational environments, we conducted a structured usability study involving academic staff acting as course instructors. These participants were introduced to realistic PjBL data scenarios and key system features, particularly those enabling empathetic feedback and cognitive bias mitigation. Participants engaged in hands-on tasks such as identifying at-risk students, interpreting behavioral trends, and generating constructive feedback for peer activities. They explored functionalities like transforming natural language inquiries into action-able queries, generating insights from educational datasets, and interpreting student outcomes through causal explanations.

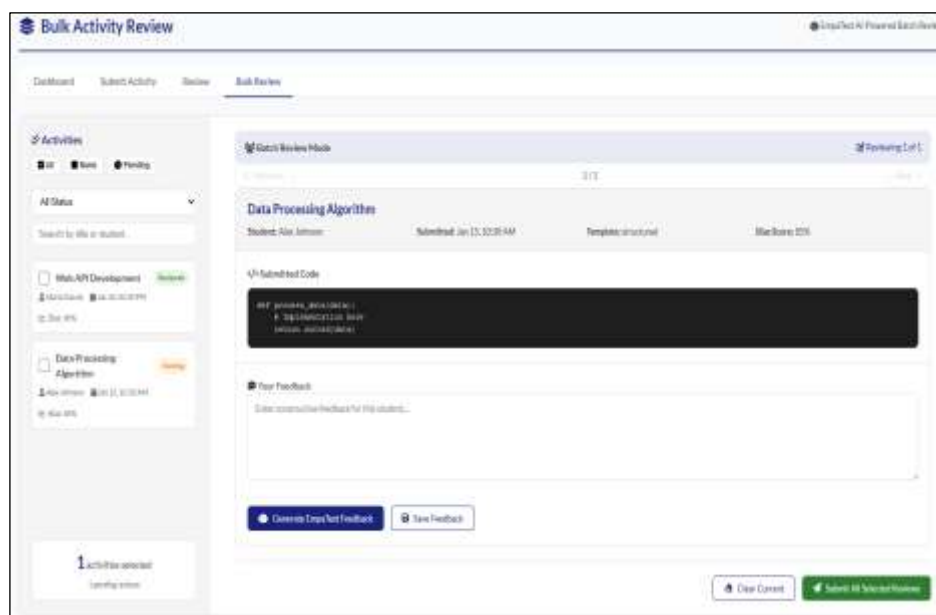


Figure 13: Our Supporting Tool: Bulk Activity Review Interface.
Source: Authors, (2026).

The primary aim of the study was to assess how well the system supports reasoning processes and decision-making workflows central to collaborative and emotionally intelligent learning (table 8). Following the session, participants completed a post-study questionnaire based on a 5-point Likert scale, designed to capture both quantitative ratings and qualitative observations regarding system clarity, relevance, usability, and decision-making support. The results affirm that participants found the EmpaText system both relevant and impactful in educational contexts, particularly for fostering psychologically safe collaboration and supporting informed decision-making. Ratings across key categories ranged from 4.2 to 4.6, reflecting strong user satisfaction. Notably, the causal reasoning capabilities received the highest score, aligning with our goal to reduce cognitive bias and promote reflective thinking. While feedback emphasized the clarity

Table 8: User Experience on EmpaText Services.

Task/Confidence Area Activity Enhancer	Mean	Median/Mode	Comments
Visualize student effort evolution using the Activity Canvas	4.30	5	Users can easily perceive changes in student effort.
Detect signs of fatigue or engagement loss in teams	4.20	5	Good ability to detect signs of fatigue.
Identify elements demonstrating team progress	4.10	4	High confidence in spotting progress indicators.
Understand effort visualization to counter confirmation bias	4.00	4	Helps prevent underestimation of team performance.
Perceive small but meaningful project contributions	3.90	4	Users recognize subtle but significant contributions.
Overall Confidence (Activity Enhancer)	4.15		Good overall confidence in effort visualization and bias mitigation.
Review Facilitator			
Use the Review Support Canvas to assess learning activities	3.85	4	Helps structure the evaluation of activities clearly.
Follow defined learning objectives during review	3.75	4	Users stay focused on key outcomes.
Analyze team roles and interactions in activities	3.60	4	Good perception of collaboration patterns.
Perceive and summarize recent changes in project work	3.70	4	Moderate ease in identifying changes.
Focus attention using indicators and avoid distraction	3.60	4	Indicators help maintain focus.
Summarize key findings for reviewer feedback	3.50	4	Quick and structured generation of feedback.
Overall Confidence (Review Facilitator)	3.75		Structured reviews improve objectivity and reduce fatigue.
Bias Mitigator			
Use checklist questions to evaluate learning activity quality	3.80	4	Checklists make activity assessment easier.
Distinguish between structured, commented, and clear work	3.70	4	Good ability to differentiate between quality aspects.
Identify workflow expansions and code correctness issues	3.60	4	Some difficulty remains with complex evaluations.
Recognize effort visibility in learning activity outcomes	3.70	4	Effort perception is fairly strong.
Decide when multiple reviewers should assess activities	3.50	4	Some uncertainty remains regarding escalation needs.
Overall Confidence (Bias Mitigator)	3.65		Checklists improve decision reliability and help reduce biases.

Source: Authors, (2026).

Table 9: Quantitative and Qualitative Feedback from EmpaText Usability Study.

Aspect	Average Rating (1–5)	Qualitative Feedback
Clarity (C)	4.3	“The system is easy to understand, but could benefit from more intuitive documentation and visual guides. ”“Highly relevant for tracking engagement and understanding peer interaction in PjBL contexts.” “Overall intuitive, though the dashboard setup could be more streamlined.” “Provided insightful prompts and aided in identifying students in need of support.” “Explanations were well-aligned with data, adding transparency to student behavior analysis.”
Relevance (R)	4.5	
Ease of Use (EU)	4.2	
Decision-Making Support (DMS)	4.4	
Causal Explanation (CE)	4.6	

Source: Authors, (2026).

a. Discussion

The empirical evaluation of EmpaText demonstrates its significant impact on reducing cognitive bias and improving feedback quality in Project-Based Learning. Our results show that the framework achieves a 60.9% reduction in cognitive biases while improving feedback specificity by 117%. These quantitative gains translate to tangible educational outcomes: teams using EmpaText showed improvements across key learning metrics, with particularly strong gains in documentation and code quality. These findings address our core research question by demonstrating that structured, AI-mediated interventions can effectively mitigate cognitive biases during learning activity reviews. The framework’s three-component architecture proved essential: the Activity Enhancer provided visual progress tracking (mean score 4.15), the Review Facilitator structured assessment workflows (mean 3.75), and the Bias Mitigator achieved 75-85% accuracy in detecting common biases. User satisfaction was notably high (86.7% positive, mean 4.16), indicating that students perceive the system as enhancing rather than replacing human interaction. The results carry important implications for PjBL theory and practice. First, they extend psycho- logical safety theory by showing that AI can scaffold safer critique environments.

Second, they address the feedback specificity gap identified in educational literature EmpaText transformed vague feedback (67.3% in baseline) into specific, actionable suggestions (71.1%). Third, the system's ability to reduce review time by 33.6% while improving quality addresses scalability challenges in large PjBL implementations. Several limitations warrant consideration. The study's single-disciplinary focus (software engineering) may limit generalizability that may not capture long-term effects. Performance variation across bias types from 97% precision for explicit biases to 76% for implicit biases in difficult tasks suggests opportunities for improvement in detecting subtle cognitive distortions. Future work should explore cross-disciplinary adaptation, longitudinal studies of bias-mitigation retention, and enhanced detection of cultural biases. The current implementation focuses on text-based feedback; expanding to multimodal analysis of team dynamics could further enhance the system's psychological insight capabilities. Overall, EmpaText demonstrates that LLM-powered systems can effectively scaffold both technical and interpersonal skill development in collaborative learning. By addressing the persistent challenge of cognitive bias in peer evaluation, it offers a scalable solution that enhances learning outcomes while fostering more psychologically safe and inclusive educational environments.

VI. CONCLUSION

This paper introduces EmpaText, an LLM-powered framework designed to reduce cognitive bias in Project-Based Learning. In PjBL environments where psychological safety is critical, peer review processes often face challenges due to cognitive biases such as confirmation bias and decision fatigue, which hinder communication and compromise project quality. EmpaText addresses these issues through three structured components: (i) Activity Enhancer, which improves learning deliverables; (ii) Review Facilitator, which guides peer assessments; and (iii) Bias Mitigator, which employs linguistic marker analysis and vocabulary processing to detect and reduce cognitive distortions. Our evaluation demonstrates significant impact: the system achieves a 60.9% reduction in cognitive biases, more than doubles specific, actionable feedback (32.7% to 71.1%). These results show that structured, LLM-mediated interventions can effectively scaffold psychologically safe, bias-aware collaborative learning environments. Future work will focus on cross-disciplinary adaptation, longitudinal studies of bias-mitigation retention, and enhanced detection of cultural and implicit biases. EmpaText contributes to human-centered AI in education, offering scalable solutions for fostering inclusive, reflective learning experiences while advancing emotionally intelligent peer review processes. As future work, we plan to design EmpaText as services to be easily integrated by tools required for project-based learning, like LMS Moodle. Secondly, we aim to develop the entire tool support that handles the whole process for teams in PjBL. Finally, we will explore further evaluation, assessing how EmpaText can significantly increase the efficiency and effectiveness of students' soft skills in communication during learning projects.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Ahmed Si Abdelhadi, Abdelkader Ouared and Abdelhafid Chadli.

Methodology: Ahmed Si Abdelhadi, Abdelkader Ouared and Mebarek Bendaoud.

Investigation: Ahmed Si Abdelhadi and Abdelkader Ouared.

Discussion of results: Ahmed Si Abdelhadi, Mebarek Bendaoud, Abdelkader Ouared and Abdelhafid Chadli.

Writing – Original Draft: Ahmed Si Abdelhadi and Abdelkader Ouared.

Writing – Review and Editing: Ahmed Si Abdelhadi, Abdelkader Ouared and Mebarek Bendaoud.

Resources: Ahmed Si Abdelhadi, Abdelkader Ouared and Abdelhafid Chadli.

Supervision: Mebarek Bendaoud and Ahmed Si Abdelhadi.

Approval of the final text: Ahmed Si Abdelhadi, Mebarek Bendaoud, Abdelkader Ouared and Abdelhafid Chadli.

VIII. REFERENCES

- [1] Lescano, G., Torres-Jimenez, J., Costaguta, R., Amandi, A., Lara-Alvarez, C.: Detecting conflicts in collaborative learning through the valence change of atomic interactions. *Expert Systems with Applications* 183, 115291 (2021) [bluehttps://doi.org/10.1016/j.eswa.2021.115291](https://doi.org/10.1016/j.eswa.2021.115291).
- [2] Kearney, J., Bond-Barnard, T., Chugh, R.: Soft skills and learning methods for 21st-century project management: a review. *International Journal of Information Systems and Project Management* 12(4), 2 (2024).
- [3] Daineko, L.V., Yurasova, I.I., Karavaeva, N.M., Pechenkina, T.E.: Project-based learning in higher education as a tool for integrated hard, soft skills and engineering language development. In: *International Conference on Professional Culture of the Specialist of the Future*, pp. 506–517 (2022).
- [4] Vogler, J.S., Thompson, P., Davis, D.W., Mayfield, B.E., Finley, P.M., Yasseri, D.: The hard work of soft skills: augmenting the project-based learning experience with interdisciplinary teamwork. *Instructional Science* 46(3), 457–488 (2017) [bluehttps://doi.org/10.1007/s11251-017-9438-9](https://doi.org/10.1007/s11251-017-9438-9).
- [5] You, J.W.: Enhancing creativity in team project-based learning amongst science college students: The moderating role of psychological safety. *Innovations in Education and Teaching International* 58(2), 135–145 (2021).
- [6] Marder, B., Ferguson, P., Marchant, C., Brennan, M., Hedler, C., Rossi, M., Black, S., Doig, R.: 'going agile': Exploring the use of project management tools in fostering psychological safety in group work within management discipline courses. *The International Journal of Management Education* 19(3), 100519 (2021).
- [7] Duhigg, C.: What google learned from its quest to build the perfect team. *The New York Times* (2016). [Accessed 16/4/2025].
- [8] Tsybulsky, D., Muchnik-Rozanov, Y.: Project-based learning in science- teacher pedagogical practicum: the role of emotional experiences in building preservice teachers' competencies. *Disciplinary and Interdisciplinary Science Education Research* 3(1), 9 (2021).
- [9] Calikli, G., Bener, A.: Empirical analyses of the factors affecting confirmation bias and the effects of confirmation bias on software developer/tester performance. In: *Proceedings of the 6th International Conference on Predictive Models in Software Engineering*, pp. 1–11 (2010).
- [10] Klayman, J.: Varieties of confirmation bias. *Psychology of learning and motivation* 32, 385–418 (1995).

- [11] Rached, S., Ouared, A., Chadli, A.: Towards Automated Bias Testing Support for Educational Decision-Makers. In: 2025 International Conference on Research in Computing at Feminine (RIF), pp. 1–8. IEEE (2025).
- [12] Saqr, M., López-Pernas, S., Murphy, K.: How group structure, members' interactions and teacher facilitation explain the emergence of roles in collaborative learning. *Learning and Individual Differences* 112, 102463 (2024) [bluehttps://doi.org/10.1016/j.lindif.2024.102463](https://doi.org/10.1016/j.lindif.2024.102463).
- [13] Montano, D.E., Kasprzyk, D., et al.: Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. *Health behavior: Theory, research and practice* 70(4), 231 (2015).
- [14] Roberto, M.A.: Lessons from everest: The interaction of cognitive bias, psychological safety, and system complexity. *California Management Review* 45(1), 136–158 (2002).
- [15] Arkes, H.R.: Costs and benefits of judgment errors: Implications for debiasing. *Psychological bulletin* 110(3), 486 (1991).
- [16] Haselton, M.G., Nettle, D., Andrews, P.W.: The evolution of cognitive bias. *The handbook of evolutionary psychology*, 724–746 (2015).
- [17] Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 363–391 (1979).
- [18] Choque-Soto, V.M., Sosa-Jauregui, V.D.: Assessing soft skills development in informatics students through project-based learning and agile frame- works. In: 2024 International Symposium on Accreditation of Engineering and Computing Education (ICACIT), pp. 1–8 (2024). IEEE.
- [19] Zhang, L., Zhang, W.: Integrating large language models into project- based learning based on self-determination theory. *Interactive Learning Environments*, 1–13 (2024).
- [20] Kusam, V.A.: Generative-ai assisted feedback provisioning for project- based learning in cs education. PhD thesis (2024).
- [21] Cook, A., Hammer, J., Elsayed-Ali, S., Dow, S.: How guiding questions facilitate feedback exchange in project-based learning. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2019).
- [22] Lopes, M.S.: Rethinking project-based learning through peer feedback: Enhancing student engagement by observing, doing and teaching. In: INTED2022 Proceedings, pp. 7995–8003 (2022). IATED.
- [23] Rouka, E.S., Anastasiadou, A.: Developing writing skills and promoting peer feedback through project-based learning in a young learners' efl classroom. *Scientific Educational Journal "Educational Circle* 8(1), 56–77 (2020).
- [24] Bassindale, J.: Learning through successful feedback: Digital opportunities for effective feedback in project-based architectural education. *International Journal of Construction Education and Research* 16(2), 84–101 (2020).
- [25] Guo, P., Saab, N., Post, L.S., Admiraal, W.: A review of project-based learning in higher education: Student outcomes and measures. *International journal of educational research* 102, 101586 (2020).
- [26] Marder, B., Ferguson, P., Marchant, C., Brennan, M., Hedler, C., Rossi, M., Black, S., Doig, R.: 'going agile': Exploring the use of project management tools in fostering psychological safety in group work within management discipline courses. *The International Journal of Management Education* 19(3), 100519 (2021).
- [27] Haddou, Y., Talbi, O., Ouared, A., Chadli, A.: A learning assistance frame- work for supporting and analyzing student teams in project-based learning. *Education and Information Technologies*, 1–42 (2025).
- [28] Kokotsaki, D., Menzies, V., Wiggins, A.: Project-based learning: A review of the literature. *Improving schools* 19(3), 267–277 (2016).
- [29] Beckers, J., Dolmans, D.H., Knapen, M.M., Van Merriënboer, J.J.: Walking the tightrope with an e-portfolio: imbalance between support and autonomy hampers self-directed learning. *Journal of Vocational Education & Training* 71(2), 260–288 (2019).
- [30] Savery, J.: Overview of pbl definition and distinctions. *Interdisciplinary journal of problem-based Learning. IJBPL* 1(1) (2006).
- [31] Al-Hattami, A.A.: The perception of students and faculty staff on the role of constructive feedback. *International Journal of Instruction* 12(1), 885–894 (2019).
- [32] Nicol, D.J., Macfarlane-Dick, D.: Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education* 31(2), 199–218 (2006).
- [33] Bolton, G.: Stories at work: reflective writing for practitioners. *The Lancet* , 354(9174), 243–245 (1999).
- [34] Ouared, A., Benali, A. A. E.: Cultivating Self-regulated Learning in Farmers for Adaptive Agricultural Decision Making. In: International Conference on Artificial Intelligence and its Applications in the Age of Digital Transformation, pp. 52–67. Springer (2025).
- [35] Zhu, Y., Dolmans, D., Kusurkar, R.A., Köhler, S.E., Abidi, L., Savel- berg, H.: Promoting students' autonomous motivation for the ongoing curriculum using a 'societal impact project' with basic psychological needs characteristics. *Medical Teacher*, 1–8 (2024).
- [36] Ouared, A., Benali, A. A. E., Chadli, A.: Enhancing Farmers' Learning Through IoT-Based Gamification and Social Awareness. In: International Conference on Artificial Intelligence and its Applications in the Age of Digital Transformation, pp. 36–51. Springer (2025).
- [37] Ajjawi, R., Boud, D.: Researching feedback dialogue: An interactional analysis approach. *Assessment & evaluation in higher education* 42(2), 252–265 (2017).
- [38] Ling, L., Rabbi, F., Wang, S., Yang, J.: Bias unveiled: Investigating social bias in LLM-generated code. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, pp. 27491–27499 (2025).
- [39] Du, Y., Huang, J.-t., Zhao, J., Lin, L.: Faircode: Evaluating social bias of llms in code generation. *arXiv preprint arXiv:2501.05396* (2025).
- [40] Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489* (2017).