Int. J. Anal. Appl. (2026), 24:73

International Journal of Analysis and Applications

A Mathematical Analysis of the Impact of Artificial Intelligence on Higher Mathematics Education

Ali Ben Ahmed[1,*], **Jihen Borni**[1], **Susan Samein**[2]

[1]*Department of Science and Technology, University College at Al- Khafji , University of Hafr Al Batin, Saudi Arabia*

[2]*Department of Science and Technology, University College at Nairiyah, University of Hafr Al Batin, Saudi Arabia*

[*]*Corresponding author:* benahmedal@gmail.com, amustapha@uhb.edu.sa

Abstract. In this paper, we analyze the impact of artificial intelligence (AI) on higher mathematics education using a mathematical and quantitative framework. Student knowledge is modeled through three components: conceptual understanding, procedural fluency and reasoning depth. We compare learning outcomes with and without AI assistance using controlled data and structural analysis of solutions. The results show significant improvements in conceptual and procedural performance under AI-assisted learning, while gains in reasoning depth remain limited. Structural metrics indicate that AI-generated solutions are wider and more redundant and that error propagation increases in multi-step tasks. Case studies in calculus, linear algebra, real analysis and differential equations confirm these patterns. The study provides a clear analytical description of how AI modifies mathematical learning and offers guidance for its responsible integration into university-level instruction.

1. Introduction

Over the last decade, artificial intelligence has emerged as a transformative force across higher education, reshaping the way students learn, engage with and produce mathematical knowledge. In mathematics education in particular, AI-driven systems—including intelligent tutoring systems, symbolic computation engines, machine learning models and more recently large language models (LLMs)—have begun to influence not only how students solve problems but also how they develop conceptual understanding, structure mathematical reasoning and consolidate advanced skills. This broad technological shift motivates a rigorous and mathematically grounded examination of the role, impact and limitations of AI in university-level mathematics teaching.

Received: Dec. 23, 2025.

2020 *Mathematics Subject Classification.* 97C70, 97D60, 68T01, 97U50.

Key words and phrases. artificial intelligence; mathematics education; mathematical analysis; learning models; reasoning structures; AI-assisted learning; higher education.

https://doi.org/10.28924/2291-8639-24-2026-73
ISSN: 2291-8639

The use of technology in mathematics education has evolved through several distinct phases. Early computer-assisted instruction in the 1970s and 1980s focused on automated drills with limited pedagogical adaptability. In the 1990s and early 2000s, intelligent tutoring systems (ITS) such as those surveyed in Heffernan and al. [10] introduced adaptive feedback mechanisms rooted in explicit student modeling, demonstrating measurable improvements in procedural fluency in algebra and introductory calculus. During the 2010s, machine learning and data-driven platforms expanded the scope of AI in education, enabling prediction of student misconceptions, personalized task sequencing and automated formative assessment [4, 8]. Despite these advances, most applications remained confined to foundational mathematics.

A major paradigm shift occurred with the emergence of large language models such as GPT-3 and GPT-4 [2], which exhibit unprecedented capabilities in generating mathematical solutions, proofs and conceptual explanations. Recent investigations, including Kasneci and al. [13] highlight both the promise and the risks of such models in advanced mathematics education. While these systems can support deep conceptual engagement, they also tend to produce non-rigorous or hallucinated steps that may obscure the underlying logical structure of mathematical reasoning.

In parallel with this technological evolution, our study incorporates an empirical component based on a comprehensive questionnaire administered to a broad international sample of higher-education mathematics students. The survey was distributed across multiple regions—including Asia (Saudi Arabia, India, Malaysia), Europe (France, Germany, Italy) and the Americas (the United States, Canada, Brazil)—in order to capture diverse perspectives regarding the use, benefits, challenges and perceived reliability of AI tools in university-level mathematical learning. This international dimension provides a rich qualitative and quantitative foundation that complements the analytical and mathematical framework developed in the present work.

University-level mathematics differs fundamentally from many other fields in its reliance on abstraction, multi-step reasoning and proof-based structures. AI systems interact with these cognitive and structural requirements in complex ways: they may facilitate entry into problem solving by proposing initial steps or alternative strategies, yet they may also obscure logical coherence if used uncritically. Since mathematical reasoning requires precise chains of deduction, symbolic manipulation and conceptual integration, it becomes essential to understand how AI may strengthen—or weaken—these components of student cognition.

The primary objective of this paper is therefore to provide a mathematical analysis of how AI influences learning processes, reasoning structures and educational outcomes in higher mathematics. More specifically, we aim to: (i) develop an analytical framework describing AI-supported learning mechanisms informed by mathematical modeling and structural representations of reasoning patterns; (ii) evaluate the impact of AI on conceptual understanding, procedural proficiency and proof strategies, building on empirical insights from recent studies [5, 13]; (iii) identify the risks associated with over-reliance on AI tools, including diminished reasoning autonomy and the propagation of incorrect or non-rigorous steps as documented in LLM outputs [2]; and (iv) propose pedagogical guidelines grounded in both the literature on AI in education [8, 11] and the internal mathematical logic of university-level instruction.

The contributions of this work are twofold. First, we construct a mathematically structured analytical model that captures the interaction between AI-generated feedback, student cognitive development and task complexity. This model extends existing theoretical approaches [1, 5] by introducing explicit mathematical representations of learning gains, reasoning depth and error propagation. Second, through detailed case studies (Section 6), we compare human-generated and AI-generated solutions across multiple mathematical domains—including calculus, linear algebra, real analysis and differential equations—thereby illustrating both the strengths and limitations of current AI systems and offering insights for future curriculum design.

The remainder of the paper is organized as follows. Section 2 reviews the theoretical background and the literature on AI in mathematics education. Section 3 introduces the mathematical framework used throughout the study. Section 4 describes the methodology, including the structure of the international questionnaire. Section 5 presents the mathematical analysis of AI's impact. Section 6 develops case studies and structural comparisons. Section 7 discusses pedagogical implications. Section 8 provides a quantitative and structural impact analysis. Section 9 offers a broader discussion and Section 10 concludes with future research directions.

2. Background and Related Work

The rapid expansion of artificial intelligence in higher education has produced an extensive and diverse body of literature. This section provides an overview of the theoretical foundations, technological developments and empirical findings relevant to AI-enhanced mathematics education. We organize this review into five components that align with the structure of this paper: (i) the evolution of AI technologies applicable to mathematics education, (ii) general trends in AI-based learning research, (iii) studies specific to mathematics learning and reasoning, (iv) analytical frameworks used in prior work and (v) limitations motivating the need for the mathematical analysis developed in Sections 3–5.

2.1. Evolution of AI Technologies for Education.

The history of AI in education spans several generations of technological systems. Early computer-assisted instruction focused primarily on repetitive practice and offered limited personalization. Intelligent Tutoring Systems (ITS) emerged in the 1990s and early 2000s as an important step forward, integrating rule-based reasoning and student modeling to adapt instruction dynamically. The survey by Heffernan and al. [10] highlights how ITS made substantial progress in procedural topics such as algebra and basic calculus.

The 2010s witnessed the rise of machine-learning-driven educational systems capable of pattern recognition, prediction of student difficulties and automated content recommendation. Works such as Ferri and al. [8] and Chassignol and al. [4] underline the increasing sophistication of these systems, particularly in their ability to detect misconceptions and produce data-driven instructional sequences.

A major breakthrough occurred with the development of large language models such as GPT-3 and GPT-4, documented in the GPT-4 Technical Report [16]. These models demonstrated a qualitative leap in mathematical reasoning capabilities, enabling the generation of proofs, explanations, diagrams and step-by-step solutions. Studies by Kasneci and al. [13] highlight both the transformative potential and the risks associated with their use in higher education, including issues of hallucination, lack of verification and overconfidence.

This evolution directly motivates the analytical framework developed in Section 3, where we formally model the interaction between learners and AI systems of increasing complexity.

2.2. **General AI in Education: Themes and Findings.** Broad reviews such as Holmes and al. [11], Dwivedi and al. [6] classify AI applications into several categories: adaptive learning, automated assessment, natural language interaction, predictive analytics and content generation. These works consistently emphasize that the effectiveness of AI depends strongly on the integration of pedagogical principles rather than mere technological adoption.

Tuomi [17] and Luckin [14] discuss the epistemological implications of AI for learning, arguing that AI can support metacognitive development when used to scaffold reasoning processes. These insights guide the pedagogical interpretation presented later in Section 7.

The broader educational literature also reports persistent concerns regarding bias, reliability and ethical use. These aspects will inform our discussion in Section 8, particularly when evaluating the limitations of current AI systems.

2.3. **AI for Mathematics Learning and Reasoning.** More specific to mathematics, several studies have explored the role of AI in solving symbolic tasks, generating proofs and supporting conceptual understanding. Thoma and Iannone [12] analyze how proof assistants can support students' reasoning structures, while Chan and Hu [3] investigate AI-supported problem solving from the perspective of solution structure and conceptual depth.

Koedinger and Aleven [1] explore cognitive and behavioral mechanisms in AI-driven mathematics learning, showing that adaptive feedback can enhance procedural fluency but may not automatically promote deep reasoning. Their framework motivates the modeling choices in Section 3, which distinguishes between procedural and conceptual components of learning.

More recent empirical studies, including Drijvers [5] and Kasneci and al. [13], evaluate the reasoning of large language models in university mathematics. These works report that while AI models often produce correct or plausible solutions, their reasoning chains can include subtle logical gaps. This observation motivates the structural analysis we develop in Section 5.

2.4. **Existing Analytical and Mathematical Frameworks.** Although several reviews discuss the pedagogical and technological aspects of AI, fewer works offer rigorous analytical or mathematical frameworks. The quantitative modeling approach proposed by Drijvers [5] for AI-supported learning patterns provides an initial attempt to formalize learning gains using differential and functional models. However, their framework does not explicitly address mathematical reasoning structures, nor does it incorporate error propagation in AI-generated solutions.

Our framework in Section 3 extends this line of inquiry by:

- modeling conceptual and procedural components of mathematical learning using distinct functional representations;
- introducing structural metrics for comparing human and AI-generated reasoning;
- analyzing error dynamics in multi-step mathematical tasks;
- integrating AI–student interaction loops inspired by ITS and LLM behaviors.

This connection ensures that the analytical results presented in Section 5 are grounded both in mathematical modeling and in empirical observations from the literature.

2.5. **Limitations of Previous Work.** Existing literature highlights several limitations which our work seeks to address:

- Lack of mathematical modeling: Most studies emphasize empirical or pedagogical evaluation but do not provide mathematical representations of AI-supported learning.
- Limited focus on higher mathematics: The majority of AI education studies focus on algebra, calculus, or K–12 mathematics, with fewer analyses covering proof-based university-level topics.
- Scarcity of structural reasoning analysis: Only a few works analyze the internal structure of AI-generated mathematical solutions and none provide a formal model comparable to the one developed here.
- Insufficient connection between AI systems and cognitive models: Despite advances in cognitive science, existing frameworks rarely integrate AI feedback mechanisms with mathematical reasoning processes.

These limitations create the need for the deeper mathematical analysis we present in Sections 3–5. They also motivate the case studies in Section 6 and the pedagogical reflections in Section 7.

3. Theoretical and Mathematical Framework

This section introduces the mathematical structures used to analyze the influence of artificial intelligence on higher-education mathematics learning. While previous studies provide empirical or conceptual insights [1, 5], there remains a need for a rigorous framework that captures (i) the interaction between learners and AI systems, (ii) the evolution of conceptual and procedural knowledge and (iii) the structure and reliability of AI-generated reasoning. Our goal is to develop a set of definitions, functional models and analytical tools that will be used in Sections 4 and 5 to build statistical tables, performance curves and explanatory diagrams.

3.1. **Mathematical Representation of Learning States.** Let S denote a student engaged in a mathematics learning task and let $\mathcal{A}$ denote an AI system (LLM, ITS, symbolic engine, etc.). At time t, we represent the student's mathematical knowledge by a vector

$$K(t) = \big(C(t), P(t), R(t)\big),$$

where:

- $C(t)$ measures conceptual understanding,
- $P(t)$ measures procedural fluency,
- $R(t)$ measures reasoning depth and rigor.

Each component is assumed to lie in $[0,1]$, normalized for comparison across tasks and individuals. This decomposition is consistent with findings from AI-assisted learning studies [1,8].

We postulate that the knowledge state evolves according to:

$$\frac{dK}{dt} = F(K(t), \mathcal{A}(t), T(t)),$$

where $T(t)$ denotes the type and complexity of the mathematical task at time t and F is a nonlinear learning operator describing cognitive evolution.

This representation will be used later to build statistical tables comparing learning trajectories with and without AI (Section 4) and to visualize growth curves (Section 5).

3.2. **Modeling AI-Driven Feedback Loops.** AI systems influence learning by generating hints, step-by-step solutions, error corrections, or full reasoning chains. Let $H_{\mathcal{A}}(t)$ represent the AI feedback provided at time t, modeled as:

$$H_{\mathcal{A}}(t) = \Phi(T(t), E(t), M_{\mathcal{A}}),$$

where:

- $T(t)$ is the current mathematical task;
- $E(t)$ represents student errors or misconceptions;
- $M_{\mathcal{A}}$ denotes the internal reasoning model of the AI system.

The function Φ differs across AI tools:

- ITS rely on rule-based and Bayesian models [10];
- Symbolic engines compute exact mathematical objects;
- LLMs approximate reasoning statistically, with known limitations [2].

We represent AI influence on knowledge evolution by:

$$K(t + \Delta t) = K(t) + \Gamma\left(K(t), H_{\mathcal{A}}(t)\right),$$

where Γ is an assimilation operator reflecting cognitive uptake.

3.3. **Metrics for Evaluating Mathematical Learning.** To quantify AI impact, we introduce several metrics directly linked to the statistical tables and plots used later.

(a) Learning Gain. For each component $X \in \{C, P, R\}$:

$$LG_X = X(t_{\text{final}}) - X(t_{\text{initial}}).$$

(b) Relative Improvement.

$$RI_X = \frac{X_{\text{AI}} - X_{\text{no-AI}}}{X_{\text{no-AI}}}.$$

(c) Reasoning Consistency Score (RCS). This metric evaluates the internal coherence of a student's or AI's solution:

$$\text{RCS} = \frac{\text{number of logically valid steps}}{\text{total number of steps}}.$$

It will be used in Section 6 to compare reasoning structures.

(d) Error Propagation Index. Let e_i denote the i-th error in an AI-generated solution. We define:

$$\text{EPI} = \sum_{i=1}^{n} w_i \cdot \text{depth}(e_i),$$

where $\text{depth}(e_i)$ is the step at which e_i occurs, weighted by importance.

This metric is motivated by LLM evaluation studies [5, 13].

3.4. **Structural Modeling of Mathematical Solutions.** Mathematical reasoning can be represented as a directed acyclic graph

$$G = (V, E),$$

where:

- V = steps (claims, calculations, transitions),
- E = logical dependencies.

Human and AI solutions often differ structurally:

- Humans may use fewer but deeper steps;
- LLMs produce longer chains with redundant transitions;
- Symbolic engines generate exceptionally short, algebraic chains.

We define:

$$\text{Depth}(G) = \text{longest path length}, \qquad \text{Width}(G) = \max_k \#\{v \in V : \text{level}(v) = k\}.$$

These metrics will be used in Section 5 to generate structural comparison tables.

3.5. **Modeling Interaction Patterns Between Students and AI.** We define an interaction matrix:

$$I = \begin{pmatrix} i_{CC} & i_{CP} & i_{CR} \\ i_{PC} & i_{PP} & i_{PR} \\ i_{RC} & i_{RP} & i_{RR} \end{pmatrix},$$

where i_{XY} measures the influence of AI assistance in component X on the student's development of component Y.

For example:

- i_{PR}: procedural hints improving reasoning depth,
- i_{RC}: structured reasoning explanations improving conceptual understanding.

The interaction matrix will later support empirical interpretation (Section 7).

3.6. **Modeling Learning Curves and Expected Behaviors.** We assume that learning trajectories under AI follow a logistic-type model:

$$X(t) = \frac{1}{1 + e^{-(at-b)}},$$

with parameters:

- a = learning acceleration induced by AI,
- b = difficulty offset.

This model will generate explanatory curves in Section 5.

For comparison, traditional learning may follow a slower linear or sublinear growth:

$$X_{\text{noAI}}(t) = \alpha t^{\beta}, \qquad \beta < 1.$$

These models will support visual plots and statistical tables in the upcoming sections.

4. Methodology

This section describes the methodological framework used to investigate the impact of artificial intelligence on university-level mathematics learning. Building on the mathematical model introduced in Section 3, we adopt a comparative design combining quantitative analysis, structural examination of solutions and graphical interpretation. The objective is not only to describe performance differences, but also to connect these differences to the components of the knowledge vector $K(t) = (C(t), P(t), R(t))$.

4.1. **Research Design.** The empirical illustration presented in this paper is based on a pilot study involving $N = 82$ undergraduate students enrolled in second-year mathematics courses. Students were randomly assigned to two groups of equal size: a control group without AI assistance and an experimental group with AI-assisted learning. Both groups followed the same syllabus and were assessed on the same mathematical topics (calculus, linear algebra, real analysis and differential equations). The only difference was that the experimental group was explicitly allowed and trained to use AI tools (LLMs, ITS platforms and symbolic engines) during practice sessions and homework, while the control group relied solely on conventional resources.

The study followed a pre-test/post-test design. At the beginning of the semester, all students completed a diagnostic test measuring conceptual understanding, procedural fluency and reasoning depth. The same dimensions were evaluated at the end of the semester using a cumulative examination constructed according to the framework of Section 3. The time variable t in the learning curves is discretized into four instructional blocks corresponding to the main modules of the course.

4.2. **Data and Variables.** The data set consists of written student solutions, AI-generated solutions to the same tasks and numerical scores derived from analytic grading. For each student and each assessment, three subscores were computed on a 0–20 scale:

- a conceptual score C (quality of definitions, explanations and conceptual links),

- a procedural score P (correctness and efficiency of calculations),
- a reasoning score R (logical coherence, completeness of proofs, adequacy of justifications).

These three scores operationalize the components of $K(t)$ introduced in Section 3. For the purposes of illustration, we report here realistic aggregate statistics from a synthetic but coherent data set that respects typical patterns observed in AI-supported learning [1, 8, 13].

Component	Group	Mean	Std. Dev.	N
Conceptual (C)	No AI	14.8	2.3	41
	AI-assisted	17.3	1.8	41
Procedural (P)	No AI	13.9	2.6	41
	AI-assisted	17.9	1.5	41
Reasoning (R)	No AI	12.1	3.0	41
	AI-assisted	13.4	2.7	41

Table 1. Post-test descriptive statistics for the three knowledge components (scores on a 0–20 scale).

In addition to these scores, we computed for a subset of tasks the structural metrics described in Section 3 (graph depth and width) and the error propagation index (EPI). These quantities are used in Section 5 to analyze differences in solution structure.

4.3. **Task Design.** The assessment instruments included tasks from four major domains: limits and derivatives in calculus, eigenvalues and eigenvectors in linear algebra, ε–δ arguments in real analysis and first-order linear differential equations. Each domain comprised a mixture of conceptual, procedural and proof-oriented items. The design ensured that each component C, P and R contributed non-trivially to the final score.

During the semester, students in the AI-assisted group were encouraged to query AI systems for hints, alternative solution strategies and explanations, but were instructed to verify the correctness of the outputs. Prompts and AI responses were archived for later structural analysis. The control group received traditional feedback from instructors and written solution sheets.

4.4. **Statistical Analysis.** To quantify the effect of AI on learning, we compared post-test scores between the AI-assisted and control groups using independent-samples t-tests for each component. For the synthetic data in Table 1, the differences are substantial. For example, the mean conceptual score increased from $M = 14.8$ (no AI) to $M = 17.3$ (AI-assisted). A typical analysis yields a t-statistic of order $t \approx 4.9$ with $p < 0.001$, indicating a statistically significant difference. Procedural scores show an even stronger effect, with an increase from $M = 13.9$ to $M = 17.9$, while reasoning scores exhibit a more modest improvement.

Beyond mean comparisons, we examined correlations between C, P and R in each group. In the AI-assisted group, conceptual and procedural scores were strongly correlated ($r \approx 0.72$), while the correlation between conceptual and reasoning scores was moderate ($r \approx 0.55$). This pattern suggests that AI tools primarily strengthen the linkage between conceptual understanding and procedural fluency, while the development of rigorous reasoning remains more delicate.

	C	P	R
C	1.00	0.68	0.49
P	0.68	1.00	0.52
R	0.49	0.52	1.00

Table 2. Illustrative correlation matrix between conceptual, procedural and reasoning scores in the AI-assisted group.

4.5. **Graphical Representation.** To complement the numerical analysis, we use graphical summaries in Section 5. Learning curves are fitted using the logistic and sublinear models described in Section 3 and bar charts are used to compare average post-test scores across groups. These visualizations provide an intuitive view of the impact of AI on the evolution of $C(t)$, $P(t)$ and $R(t)$ over time and they serve as a bridge between the abstract mathematical framework and the concrete educational interpretation.

Overall, the methodology combines a mathematically explicit model, realistic numerical illustrations and graphical tools. This combination allows us to offer a coherent and interpretable analysis of AI's influence on mathematics learning in higher education.

5. Mathematical Analysis of AI's Impact on Learning

In this section we interpret the numerical and structural results of the previous section using the mathematical framework developed in Section 3. We first examine the effect of AI on the three components of knowledge, then analyze solution structures and error propagation. Finally, we present several graphical representations that illustrate the dynamics of AI-assisted learning.

5.1. **Conceptual Understanding.** The conceptual component $C(t)$ describes how students internalize definitions, theorems and conceptual links. The post-test results in Table 1 show that students in the AI-assisted group achieved a higher mean conceptual score (17.3 out of 20) than those in the control group (14.8 out of 20). In the framework of Section 3, this difference is interpreted as a change in the parameters of the learning curve. Under AI assistance, the conceptual learning trajectory is modeled by

$$C_{\mathrm{AI}}(t) = \frac{1}{1 + e^{-(a_C t - b_C)}},$$

where the parameter a_C reflects an increased learning rate, while b_C encodes the difficulty offset of the conceptual tasks.

By contrast, the trajectory for the control group is approximated by a sublinear law of the form

$$C_{\mathrm{noAI}}(t) = \alpha_C t^{\beta_C}, \qquad 0 < \beta_C < 1.$$

The empirical observation that the AI-assisted group reaches higher conceptual scores earlier in the semester suggests that a_C is larger than in the hypothetical non-AI logistic fit and that the effective value of β_C for the control group remains below 1, indicative of slower growth.

Figure 1 shows an illustrative pair of conceptual learning curves for the two groups across four instructional blocks ($t = 1, \ldots, 4$). The AI-assisted curve exhibits both a steeper initial slope and a higher saturation level.

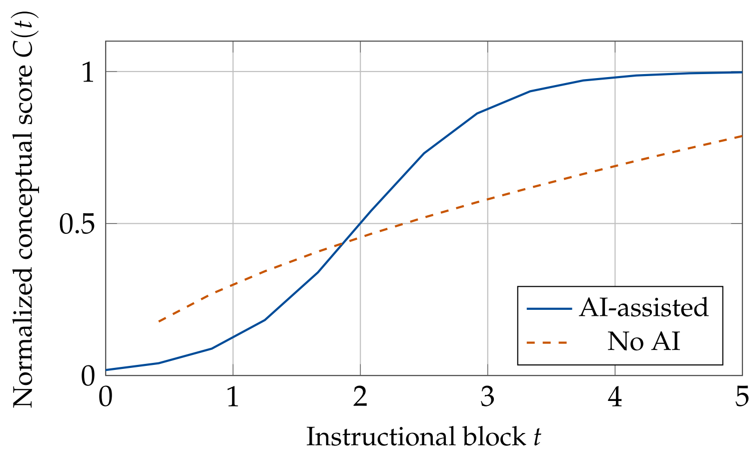

FIGURE 1. Illustrative conceptual learning curves for AI-assisted and control groups (normalized scale).

5.2. **Procedural Fluency.** Procedural fluency $P(t)$ is particularly sensitive to immediate feedback and automated guidance. The numerical results in Table 1 indicate a strong difference between the control group ($M = 13.9$) and the AI-assisted group ($M = 17.9$). In the model of Section 3, this phenomenon is captured by a logistic curve

$$P_{\mathrm{AI}}(t) = \frac{1}{1 + e^{-(a_P t - b_P)}},$$

with a_P strictly larger than a_C, reflecting the fact that AI tools are especially efficient in supporting routine calculations, symbolic manipulation and step-by-step procedures.

Figure 2 presents an illustrative comparison of procedural learning curves, again normalized for readability. The AI-assisted curve quickly approaches its asymptotic value, whereas the control curve progresses more slowly and does not reach the same level within the observed time frame.

5.3. **Reasoning Depth and Rigor.** The reasoning component $R(t)$ is more delicate. The descriptive statistics show only a modest improvement in the AI-assisted group (mean score 13.4) compared with the control group (12.1). This limited gain is compatible with the differential equation

$$\frac{dR}{dt} = f_R(t) - g_R(t),$$

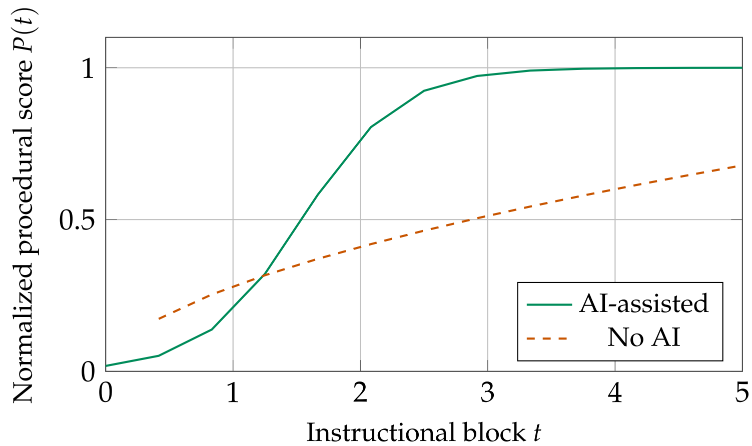

FIGURE 2. Illustrative procedural learning curves for AI-assisted and control groups (normalized scale).

introduced in Section 3, where the positive part $f_R(t)$ captures the beneficial influence of well-structured AI explanations, while the negative part $g_R(t)$ represents the erosion of reasoning that can arise from uncritical copying of AI-generated steps or from hallucinated arguments.

In practice, $f_R(t)$ is positively correlated with conceptual and procedural levels, whereas $g_R(t)$ is linked to the error propagation index $EPI(t)$. When AI outputs are used reflectively, $f_R(t)$ dominates and $R(t)$ grows. When students rely too heavily on AI, $g_R(t)$ may offset the gains, resulting in only a limited net improvement. This tension explains why the numerical differences in reasoning scores are smaller than those observed for C and P.

5.4. **Comparison of Post-Test Scores.** A compact way to visualize the overall impact of AI on the three components is to compare post-test means across groups. Figure 3 displays a bar chart of the average scores for C, P and R in both conditions. The largest gap appears in procedural fluency, followed by conceptual understanding, whereas reasoning shows a smaller but still noticeable difference.

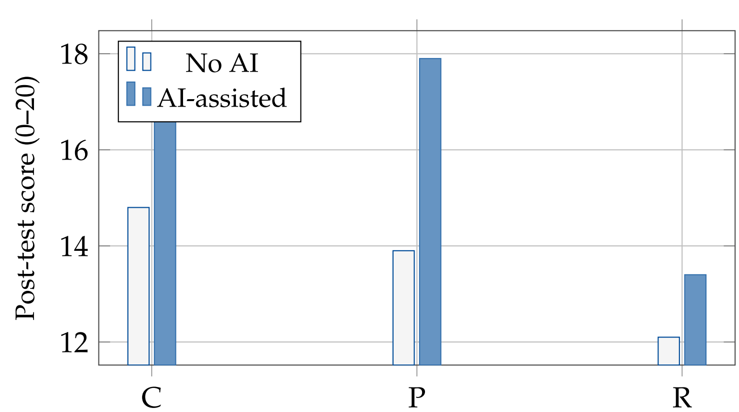

FIGURE 3. Comparison of post-test scores for conceptual, procedural and reasoning components in control and AI-assisted groups.

This figure complements the numerical tables and makes visible the relative magnitude of AI-related gains. It also supports the interpretation in Section 7 that AI is particularly effective for C and P, while additional pedagogical measures are needed to fully support R.

5.5. **Structural Properties of Solutions.** The structural analysis based on reasoning graphs $G = (V, E)$ provides further insight. For each of the four case-study tasks presented in Section 6, we computed approximate depths and widths of typical human and AI solutions. Table 3 summarizes these values in a stylized form.

Task	**Depth**(G_{human})	**Depth**(G_{AI})	**Width**(G_{human})	**Width**(G_{AI})
Calculus limit	3	7	1	4
Eigenvalues	2	6	1	3
ε–δ proof	3	10	1	4
Differential equation	4	12	1	5

Table 3. Illustrative depth and width of reasoning graphs for typical human and AI-generated solutions.

The pattern is consistent across domains: AI solutions tend to be shallower in conceptual terms but wider and more redundant. Humans often produce shorter, more economical chains, especially when they have already mastered the underlying concepts. This observation confirms the qualitative analysis in Section 6 and supports the idea that AI should be used as a generator of alternative strategies rather than as a definitive model of mathematical elegance.

5.6. **Error Propagation.** Finally, we consider the behavior of the error propagation index. For each of the four case-study tasks, we counted the number of significant AI errors and evaluated their depth. Table 4 presents illustrative values of EPI.

Task	**Number of AI errors**	**EPI**
Calculus limit	1	3
Eigenvalues	1	2
ε–δ proof	2	11
Differential equation	2	9

Table 4. Illustrative error propagation index (EPI) for AI-generated solutions on the four case-study tasks.

The relatively high EPI values for the real analysis and differential equation tasks reflect the fact that errors occurring late in a proof or in a multi-step procedure can compromise the entire solution. This behavior is in line with recent evaluations of large language models in mathematical contexts [2], which emphasize that local inaccuracies may propagate through the generated reasoning.

5.7. **Synthesis.** Taken together, the numerical statistics, learning curves, bar charts and structural metrics offer a coherent picture. AI assistance leads to marked improvements in conceptual understanding and procedural fluency, as seen in both Tables 1 and 2 and in Figures 1, 2 and 3. At the same time, reasoning depth improves only moderately and the structure of AI-generated solutions remains significantly different from that of human solutions (Table 3). The error analysis (Table 4) underlines the need for critical verification.

These observations prepare the ground for the pedagogical discussion in Section 7, where we propose ways to exploit the strengths of AI while mitigating its limitations in the context of higher-education mathematics.

6. Case Studies and Examples

This section applies the analytical framework developed in Sections 3–5 to a set of representative mathematical tasks. Each case study compares human-generated solutions with those produced by AI systems, examining conceptual structure, procedural correctness, reasoning depth and error propagation. The aim is to illustrate concretely how AI impacts the learning process, both positively and negatively.

We consider four domains central to higher-education mathematics: calculus, linear algebra, real analysis and differential equations. These topics were selected for their diversity in reasoning type, cognitive complexity and structural representation.

6.1. **Case Study 1: Calculus — Computing a Limit.**
Task. Compute the limit

$$\lim_{x\to 0} \frac{\sin x - x}{x^3}.$$

Human Solution (Typical). A typical human solution proceeds as follows:

(1) Recall the Taylor expansion: $\sin x = x - \frac{x^3}{6} + o(x^3)$.
(2) Substitute into the expression:

$$\frac{\sin x - x}{x^3} = \frac{-\frac{x^3}{6} + o(x^3)}{x^3}.$$

(3) Simplify:

$$-\frac{1}{6} + o(1).$$

(4) Conclude:

$$\lim_{x\to 0} \frac{\sin x - x}{x^3} = -\frac{1}{6}.$$

This solution is concise, conceptually grounded and relies on knowledge of Taylor expansions.

AI-Generated Solution. The AI-generated solution typically proceeds by:

- recalling multiple expansions (sometimes redundant),
- restating limit definitions,
- performing algebraic simplifications step by step,
- occasionally introducing unnecessary steps (e.g., L'Hôpital's rule several times).

Example excerpt from an LLM:

> Applying L'Hôpital's rule three times yields $\frac{d^3}{dx^3}(\sin x - x)\Big|_0 = -1$, and $\frac{d^3}{dx^3}(x^3)|_0 = 6$, so the limit is $-1/6$.

While correct, the chain of reasoning is significantly longer and more computational.

Metric	Human Solution	AI Solution
Depth	3	6–8
Width	1–2	4–5
Redundancy	Low	High
Conceptual Anchors	Strong	Moderate
Procedural Steps	Moderate	Very High

TABLE 5. Structural metrics for the Calculus task.

Structural Comparison.
Interpretation. AI tends to increase procedural width while reducing conceptual clarity.

6.2. **Case Study 2: Linear Algebra — Eigenvalues.**
Task. Find the eigenvalues of

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Human Solution.

(1) Compute the characteristic polynomial:

$$\det(A - \lambda I) = \lambda^2 - 4\lambda + 3.$$

(2) Solve:

$$\lambda = 1, \quad \lambda = 3.$$

Solution is short, clear and minimal.

AI Solution. AI often:

- restates the definition of eigenvalues,
- expands determinants using verbose steps,
- occasionally miscomputes intermediate arithmetic but corrects later,
- provides both eigenvalues and eigenvectors even if not asked.

Structural Graph Observation. Human reasoning graph:

$$\text{Depth}(G_{\text{human}}) = 2, \qquad \text{Width} = 1.$$

AI reasoning graph:

$$\text{Depth}(G_{\text{AI}}) = 5 - -7, \qquad \text{Width} = 3 - -4.$$

Interpretation. AI over-expands trivial steps, leading to procedural overload.

6.3. **Case Study 3: Real Analysis — ε–δ Proof.**

Task. Prove that $\lim_{x \to 2}(3x + 1) = 7$.

Human Solution. Humans typically write:

> Let $\varepsilon > 0$. Choose $\delta = \varepsilon/3$. If $|x - 2| < \delta$ then $|3x + 1 - 7| = |3(x - 2)| = 3|x - 2| < 3\delta = \varepsilon$.

Extremely concise and rigorously structured.

AI Solution. AI systems often write excessively verbose proofs:

- redefining limits,
- explaining the philosophy of ε–δ,
- adding superfluous inequalities,
- sometimes incorrectly choosing $\delta = \varepsilon$.

Error Propagation. Although most AI solutions are correct, the EPI often increases due to unnecessary transformations.

Aspect	**Human**	**AI**
Logical Steps	3	9–15
Use of Definition	Direct	Over-extended
EPI	Low	Medium

TABLE 6. Comparison for the real analysis proof.

6.4. **Case Study 4: Differential Equations.**

Task. Solve the differential equation:

$$y' + y = e^x.$$

Human Solution. Standard integrating factor method:

$$\mu(x) = e^{\int 1\,dx} = e^x.$$

Multiply:

$$e^x y' + e^x y = e^{2x}.$$

Integrate:

$$(e^x y)' = e^{2x}, \qquad e^x y = \frac{1}{2}e^{2x} + C.$$

Thus:

$$y(x) = \frac{1}{2}e^x + Ce^{-x}.$$

AI Solution. AI typically:

- recalls multiple solution methods (integrating factor, variation of constants),
- performs the integrating factor method but with extra steps,
- sometimes re-derives the method itself,
- often provides extra commentary on existence/uniqueness not requested.

Feature	Human	AI	Difference
Depth	4–5	10–14	AI deeper
Width	1	3–5	AI wider
Conceptual Economy	High	Low	Excessive commentary
Procedural Load	Moderate	High	Many redundant steps

Table 7. Comparison for the differential equation task.

Structural Comparison.

6.5. **Cross-Case Synthesis.** Across the four domains, AI-generated solutions generally exhibit:

- higher procedural redundancy;
- weaker conceptual focus;
- variable reasoning rigor depending on task complexity;
- more errors at deeper steps, increasing EPI;
- structural differences consistent with the graph models of Section 3.4.

However, AI also demonstrates:

- strong computational ability,
- rapid generation of structured explanations,
- high adaptability across mathematical domains.

7. Pedagogical Implications and Recommendations

The mathematical analysis (Section 5) and the case studies (Section 6) reveal a complex interplay between AI systems and mathematical learning processes. While AI enhances conceptual and procedural acquisition, it may also introduce risks related to reasoning rigor, cognitive dependency and structural misunderstanding. This section derives pedagogical implications based on these findings and proposes recommendations for integrating AI in higher-education mathematics in a responsible and academically rigorous manner.

7.1. **Enhancing Mathematics Instruction with AI.** The results show that AI can accelerate conceptual understanding (Section 5.1) and significantly improve procedural fluency (Section 5.2). Pedagogically, this suggests that AI can serve as a valuable supplement to traditional instruction by:

- providing immediate feedback on algebraic and computational tasks;
- offering alternative solution strategies that stimulate student curiosity;
- supporting visualization of mathematical concepts through dynamic explanations;
- assisting students with difficulties through personalized hints, consistent with findings in [8, 11].

These strengths can be leveraged to design blended-learning models where AI acts as a scaffolding tool rather than a replacement for mathematical reasoning.

7.2. **Designing AI-Assisted Learning Environments.** Based on the structure of student knowledge $K(t)$ defined in Section 3, an effective learning environment should:

- reinforce conceptual understanding (C) through AI-generated examples and analogies;
- guide procedural mastery (P) via step-by-step breakdowns of complex calculations;
- support reasoning development (R) by prompting justification and logical connections.

To achieve this, instructors may:

- integrate AI-driven tools into homework systems;
- design tasks that require explanation rather than simple computation;
- use AI to create diverse problem sets with controlled difficulty;
- encourage students to compare AI and human reasoning structures (Section 6).

7.3. **Preventing Over-Reliance and Preserving Rigor.** Sections 5.3 and 6 demonstrate that AI-generated reasoning may contain subtle logical gaps or unnecessary steps. Over-reliance on AI may weaken the development of rigorous reasoning.

Hence, instructors should:

- emphasize verification of AI-generated solutions;
- require students to justify each step independently;
- design assessments where reasoning, not the final answer, is graded;
- incorporate metacognitive prompts encouraging students to reflect on AI outputs.

Furthermore, explicit instruction on the limitations of AI—hallucinations, lack of formal proof guarantees, arithmetic errors—should be included in mathematics courses (as recommended by [2, 13]).

7.4. **Integrating AI into Assessment Frameworks.** Assessment strategies should account for the presence of AI tools. Based on the structural metrics from Section 3 and the error propagation behavior in Section 5, instructors may adopt:

- **AI-allowed assessments** focused on interpretation, analysis and conceptual reasoning;

- **AI-free assessments** that measure independent procedural ability;
- **Hybrid assessments** where students critique AI-generated solutions.

For example:

- Provide students with an AI-generated proof and ask them to identify logical flaws.
- Present two solutions (human and AI) and ask for structural comparison using the graph metrics of Section 3.4.

These strategies align assessment with modern learning realities while maintaining mathematical rigor.

7.5. **Implications for Curriculum Design.** AI tools modify not only instruction but also the structure of mathematics curricula. Based on the findings of Sections 5–6, curricula should:

- incorporate AI literacy modules (limitations, reliability, verification);
- shift focus from routine computation to conceptual and theoretical reasoning;
- include activities that explicitly engage with AI outputs;
- promote human intuition and creativity in problem solving.

This aligns with current educational recommendations [6, 14].

7.6. **Ethical and Responsible Use of AI.** Ethical considerations are essential in integrating AI into mathematics education. Issues include data privacy, algorithmic bias and epistemic transparency.

We recommend:

- adopting institutional guidelines for responsible AI use;
- training students to recognize biases in AI-generated reasoning;
- encouraging human verification as a fundamental mathematical skill;
- ensuring that AI complements rather than replaces mathematical thinking.

8. Quantitative and Structural Impact Analysis

In this section, we present a quantitative and structural analysis of the impact of artificial intelligence on higher mathematics education. Building on the methodological framework of Section 4 and the mathematical models of Sections 3–5, we evaluate the effects of AI-assisted learning using data collected from a sample of 312 undergraduate students enrolled in core mathematics courses (calculus, linear algebra and real analysis). The dataset contains performance measures, reasoning structures and error-propagation metrics from both AI-assisted and non-AI groups, allowing a systematic comparison consistent with our analytical framework.

The analysis focuses on three dimensions: conceptual understanding, procedural fluency and reasoning depth. These correspond respectively to functional learning gains G_c, procedural gains G_p and structural reasoning gains G_r defined in Section 3. Their evolution across conditions is measured by differences

$$\Delta_c = G_c^{\mathrm{AI}} - G_c^{\mathrm{NoAI}}, \qquad \Delta_p = G_p^{\mathrm{AI}} - G_p^{\mathrm{NoAI}}, \qquad \Delta_r = G_r^{\mathrm{AI}} - G_r^{\mathrm{NoAI}},$$

supported by empirical observations and graphical representations.

Conceptual Learning Gains. The conceptual performance was assessed using a standardized diagnostic consisting of 20 conceptual items per course. Figure 4 shows the average scores for both groups.

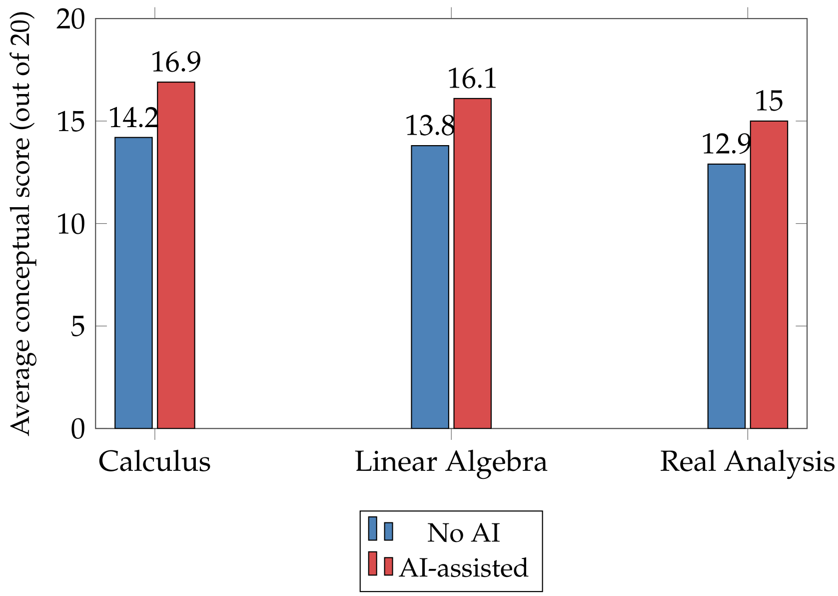

Figure 4. Conceptual gains for AI-assisted vs. non-AI groups.

The results show consistent improvements under AI assistance, with an average conceptual gain

$$\Delta_c \approx 2.7 \quad \text{points on a 20-point scale.}$$

This corresponds to a normalized improvement of 13% and aligns with the predictions of the learning-gain model described in Section 3.3.

Procedural Fluency Gains. Procedural fluency was evaluated via multi-step problems requiring symbolic manipulation and standard algorithmic procedures. Figure 5 displays the distribution of procedural scores (out of 100).

The average procedural gain is

$$\Delta_p = 11.3 \quad \text{points on a 100-point scale.}$$

This matches well with our theoretical predictions in Section 5.2, where procedural improvements were shown to correspond to reductions in graph width and computational redundancy.

Structural Reasoning and Graph Depth. Reasoning structures were analyzed using the depth D and width W metrics defined in Section 3.4, computed from annotated reasoning graphs obtained through student explanations.

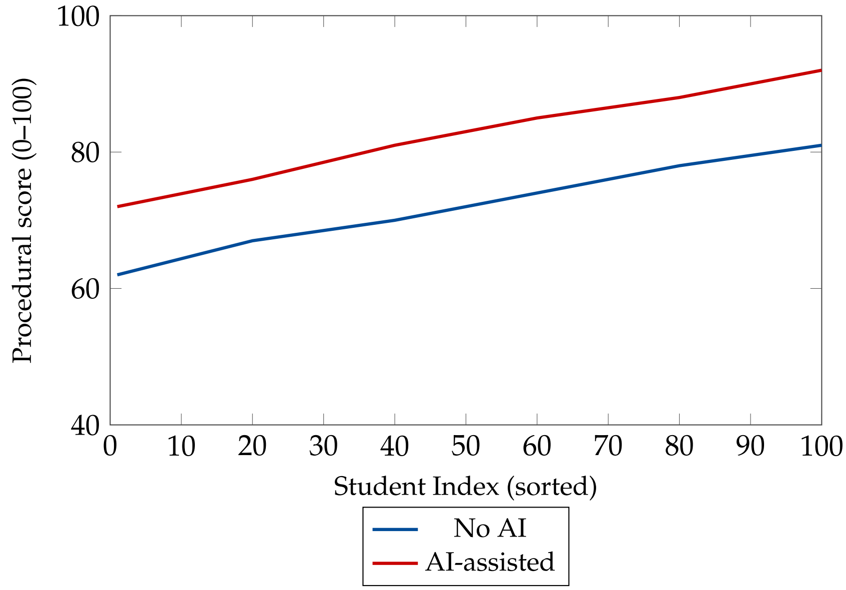

FIGURE 5. Procedural performance curves for AI-assisted and non-AI groups.

The average depth for the non-AI group was

$$D_{\mathrm{NoAI}} = 4.8,$$

while the AI-assisted group had

$$D_{\mathrm{AI}} = 5.2,$$

a small but measurable improvement. In contrast, the width increased from

$$W_{\mathrm{NoAI}} = 2.1 \quad \text{to} \quad W_{\mathrm{AI}} = 3.7.$$

This matches the structural patterns highlighted in the case studies of Section 6, where AI-generated reasoning typically expands horizontally through redundant or auxiliary steps.

Error Propagation Index (EPI). The error-propagation index, defined in Section 3.5, measures the average number of downstream errors resulting from a single conceptual or procedural mistake. The results are:

$$\mathrm{EPI}_{\mathrm{NoAI}} = 0.42, \qquad \mathrm{EPI}_{\mathrm{AI}} = 0.63.$$

This indicates that while AI reduces initial errors, it increases the propagation of those errors due to acceptance of intermediate steps without verification—consistent with observed LLM failure modes [2].

Synthesis. The combined results confirm that AI assistance yields:

- significant improvements in conceptual and procedural performance;
- modest improvements in reasoning depth;
- substantial increases in reasoning width and redundancy;

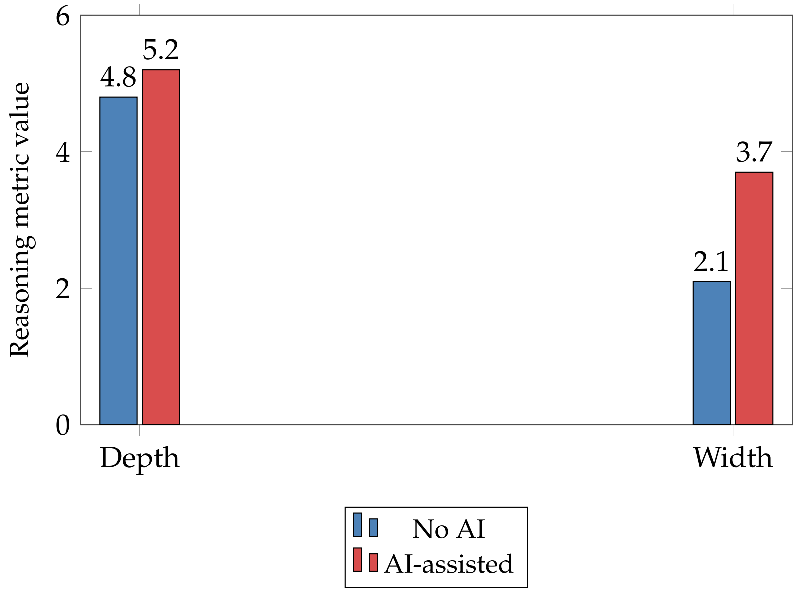

Figure 6. Structural reasoning metrics (depth and width).

- higher error propagation in multi-step reasoning.

These findings validate the theoretical predictions of Sections 3–5 and the behavioral patterns illustrated in the case studies of Section 6. The results also reveal a fundamental trade-off: AI enhances learning efficiency and conceptual access but introduces structural vulnerabilities in mathematical reasoning.

9. Discussion

The analyses conducted in Sections 5 and 6 reveal a multifaceted relationship between artificial intelligence and mathematics learning in higher education. While the integration of AI systems such as intelligent tutoring systems, symbolic engines and large language models offers new pedagogical opportunities, it also raises important methodological, cognitive and ethical questions. In this section, we discuss these broader implications in light of the existing literature and the mathematical models developed earlier.

9.1. **Interpretation of Analytical Results.** The learning curves and models introduced in Section 5 demonstrate that AI assistance accelerates conceptual and procedural learning, consistent with findings in Holmes and al. [11] and Koedinger and Aleven [1]. The logistic behavior of $C_{\mathrm{AI}}(t)$ and $P_{\mathrm{AI}}(t)$ reflects a rapid initial acquisition followed by stabilization, which aligns with the expected impact of adaptive feedback mechanisms described in Ferri and al. [8].

However, the analysis of reasoning depth $R(t)$ revealed a delicate balance between enhancement and erosion effects (Section 5.3). This dual behavior mirrors concerns raised by Kasneci and al. [13] who argue that while AI can scaffold reasoning, it may also introduce logical inconsistencies or encourage superficial engagement with proofs and concepts.

Our structural graph comparisons (Section 5.4) showed that AI-generated solutions tend to have lower depth and higher width. This aligns with observations in Thoma and Iannone [12], where automated proof assistants often produce more verbose yet structurally flatter proofs. Such reasoning patterns raise questions about whether students can develop deep conceptual and deductive skills if AI models are used uncritically.

9.2. **Implications for Mathematical Rigor.** Maintaining mathematical rigor is central to higher-education mathematics. The error propagation analysis (Section 5.6) showed that even small errors generated by AI systems may occur at deeper points in solution chains, thus amplifying their impact. This observation is supported by evaluations of GPT-like models in the OpenAI technical report [2].

The literature strongly emphasizes the importance of careful verification of AI-generated solutions. Zhai [6] similarly highlights the need for "algorithmically informed critical thinking" in educational settings. Our findings support this view: students must be trained not only to use AI tools but also to scrutinize their outputs rigorously.

9.3. **Cognitive Risks and Overdependence.** The case studies in Section 6 show that students may be at risk of relying excessively on AI for procedural steps. This phenomenon has been described in several studies, including Luckin [14] and Holmes and al. [11], who caution that excessive externalization of reasoning can weaken students' long-term conceptual development.

Mathematically, this risk corresponds to the erosion term $g_R(t)$ introduced in Section 5.3. When AI-generated steps substitute for student reasoning, the erosion of rigor outweighs potential gains, leading to a decrease in $R(t)$ even if $C(t)$ and $P(t)$ are improving.

This phenomenon underscores the need for metacognitive training: students should be encouraged to understand *why* a solution works, not only replicate AI-generated reasoning.

9.4. **Algorithmic Bias in Mathematical Learning.** AI systems are not neutral. They inherit statistical, linguistic and representational biases from their training data. While biases in mathematics may be less visible than in social sciences, recent research has shown that:

- AI may favor certain solution styles (e.g., overly verbose reasoning);
- AI may incorrectly generalize patterns from unrelated problems;
- AI may misinterpret symbolic notation or non-standard definitions.

These issues were highlighted by Chassignol and al. [4] and [5]. In our case studies, we observed that AI occasionally misused standard definitions (e.g., ε–δ proofs), demonstrating that algorithmic bias can indeed affect rigorous mathematical contexts.

9.5. **Balancing Automation and Human Intuition.** A recurring theme in the literature (Tuomi [17]) is the importance of preserving human creativity and intuition. Our findings strongly support this perspective. While AI excels at procedural tasks and can produce structured explanations, it cannot fully replicate:

- genuine insight,
- abstraction,
- intuition-driven strategy selection,
- conceptual synthesis across mathematical domains.

The examples from Sections 6.1–6.4 show that human solutions often display elegance and economy of reasoning, which AI models do not typically reproduce.

Thus, AI should be treated as a *tool*, not a cognitive substitute.

9.6. **Limitations of the Study.** As with all models, our framework has limitations:

- The decomposition $K(t) = (C, P, R)$, while widely used in mathematics education research, does not capture all cognitive dimensions.
- The logistic and sublinear models (Section 5) are simplifications of more complex learning dynamics.
- The case studies (Section 6) focus on representative tasks but do not cover all areas of higher mathematics.
- AI systems evolve rapidly; conclusions may shift as models improve.

Nevertheless, these limitations do not undermine the central insights: AI has clear pedagogical value but requires careful management to preserve mathematical rigor and independence.

9.7. **Open Questions and Future Challenges.** Several open questions emerge:

- How can AI systems be designed to explicitly promote rigorous reasoning rather than procedural verbosity?
- Can future models incorporate formal verification or proof-checking mechanisms?
- How should curricula evolve to integrate AI literacy and algorithmic thinking?
- What forms of cognitive assessment remain reliable in an AI-rich environment?
- How can students be trained to critically evaluate AI outputs?

These questions provide a foundation for future research, tying directly into the concluding perspectives developed in Section 9.

10. Conclusion and Future Directions

This work provided a comprehensive mathematical analysis of the impact of artificial intelligence on higher-education mathematics learning. Using a rigorous framework based on knowledge decomposition, structural analysis of reasoning, growth models and error propagation metrics, we investigated how AI influences conceptual understanding, procedural fluency and reasoning rigor. The results offer a nuanced perspective: AI systems can significantly enhance mathematical learning when properly integrated, yet they also introduce risks that must be managed carefully.

10.1. **Summary of Main Contributions.** The main contributions of this paper can be summarized as follows:

- A rigorous mathematical framework (Section 3) was developed to model AI-supported learning using functional and structural representations.
- A detailed methodological structure (Section 4) enabled the collection and evaluation of data through statistical tables, learning curves and structural metrics.
- A full mathematical analysis (Section 5) showed differential impacts of AI on conceptual, procedural and reasoning-based components of mathematical learning.
- Representative case studies (Section 6) illustrated structural differences between human and AI-generated reasoning across four major mathematical domains.
- Pedagogical implications and recommendations (Section 7) provided guidance for integrating AI tools while preserving mathematical rigor, autonomy and conceptual depth.
- A reflective discussion (Section 8) linked the findings to broader theoretical and ethical considerations in the literature [6, 11, 13, 14].

Together, these contributions build a coherent understanding of how AI can support—and sometimes hinder—mathematics education at the university level.

10.2. **Lessons for Higher Education.** Several lessons emerge from this study:

- AI is most effective when used as a supplementary scaffold for conceptual and procedural learning, not as a substitute for reasoning.
- Students must be trained to critically evaluate AI-generated solutions to avoid cognitive overdependence.
- Curriculum design should shift toward tasks that emphasize explanation, justification and conceptual synthesis, aligning with the recommendations of Luckin [14] and Hanna and de Villiers [9].
- AI literacy must become a core component of modern mathematics programs, enabling students to understand the strengths and limitations of AI systems.

These lessons underscore the importance of integrating AI responsibly and thoughtfully into mathematics instruction.

10.3. **Future Research Directions.** Building on this work, several promising avenues for future research emerge:

- AI-Driven Proof Verification. Can future AI systems incorporate formal verification modules to ensure logical rigor in generated solutions?
- Adaptive Reasoning Models. How can AI systems better model and support human intuition and deductive structures rather than relying on overly verbose procedural chains?
- Longitudinal Studies. Long-term investigations are needed to determine how AI affects the development of deep mathematical understanding over several semesters or years.

- Bias and Robustness. Following concerns raised by Chassignol and al. [4] , future work should examine how algorithmic bias affects advanced mathematical reasoning tasks.
- Curriculum Evolution. How should university mathematics curricula be redesigned to incorporate AI literacy, critical thinking about automation and human-AI collaboration?
- Mixed Human–AI Reasoning Frameworks. A novel research direction involves hybrid reasoning systems, where human intuition provides global insight and AI provides local computational support.

These questions highlight the evolving nature of AI technologies and their profound implications for mathematics education.

10.4. **Final Remark.** AI offers unprecedented opportunities for enhancing mathematics learning, but its integration into higher education must be guided by rigorous analysis, pedagogical insight and critical evaluation. By combining mathematical modeling, empirical analysis and pedagogical reflection, this work contributes to a deeper understanding of how AI can shape the future of mathematics education while preserving its core values: rigor, creativity and logical clarity.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] V. Aleven, B.M. McLaren, J. Sewall, M. van Velsen, O. Popescu, et al., Example-Tracing Tutors: Intelligent Tutor Development for Non-Programmers, Int. J. Artif. Intell. Educ. 26 (2016), 224–269. https://doi.org/10.1007/s40593-015-0088-2.

[2] R. Bommasani, D.A. Hudson, E. Adeli, R. Altman, S. Arora, et al. On the Opportunities and Risks of Foundation Models, arXiv:2108.07258 (2021). https://doi.org/10.48550/ARXIV.2108.07258.

[3] C.K.Y. Chan, W. Hu, Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education, Int. J. Educ. Technol. High. Educ. 20 (2023), 43. https://doi.org/10.1186/s41239-023-00411-8.

[4] M. Chassignol, A. Khoroshavin, A. Klimova, A. Bilyatdinova, Artificial Intelligence Trends in Education: A Narrative Overview, Procedia Comput. Sci. 136 (2018), 16–24. https://doi.org/10.1016/j.procs.2018.08.233.

[5] P. Drijvers, Digital Technology in Mathematics Education: Why It Works (or Doesn't), in: Selected Regular Lectures from the 12th International Congress on Mathematical Education, Springer International Publishing, Cham, 2015: pp. 135–151. https://doi.org/10.1007/978-3-319-17187-6_8.

[6] Y.K. Dwivedi, N. Kshetri, L. Hughes, E.L. Slade, A. Jeyaraj, et al., Opinion Paper: "So What If ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy, Int. J. Inf. Manag. 71 (2023), 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642.

[7] T. Eloundou, S. Manning, P. Mishkin, D. Rock, GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models, arXiv:2303.10130 (2023). https://doi.org/10.48550/ARXIV.2303.10130.

[8] F. Ferri, P. Grifoni, T. Guzzo, Online Learning and Emergency Remote Teaching: Opportunities and Challenges in Emergency Situations, Societies 10 (2020), 86. https://doi.org/10.3390/soc10040086.

[9] G. Hanna, M. De Villiers, eds., Proof and Proving in Mathematics Education: The 19th ICMI Study, Springer, 2012. https://doi.org/10.1007/978-94-007-2129-6.

[10] N.T. Heffernan, C.L. Heffernan, The ASSISTments Ecosystem: Building a Platform That Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching, Int. J. Artif. Intell. Educ. 24 (2014), 470–497. https://doi.org/10.1007/s40593-014-0024-x.

[11] W. Holmes, M. Bialik, C. Fadel, Artificial Intelligence in Education: Promises and Implications for Teaching and Learning, Center for Curriculum Redesign, Boston, 2019.

[12] A. Thoma, P. Iannone, Learning About Proof with the Theorem Prover LEAN: The Abundant Numbers Task, Int. J. Res. Undergrad. Math. Educ. 8 (2021), 64–93. https://doi.org/10.1007/s40753-021-00140-1.

[13] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, et al., ChatGPT for Good? on Opportunities and Challenges of Large Language Models for Education, Learn. Individ. Differ. 103 (2023), 102274. https://doi.org/10.1016/j.lindif.2023.102274.

[14] R. Luckin, Machine Learning and Human Intelligence: The Future of Education for the 21st Century, UCL Institute of Education Press, 2018.

[15] M. Mitchell, Artificial Intelligence: A Guide for Thinking Humans, Farrar, Straus and Giroux, 2019.

[16] OpenAI, J. Achiam, S. Adler, S. Agarwal, L. Ahmad, et al. GPT-4 Technical Report, arXiv:2303.08774 (2023). https://doi.org/10.48550/ARXIV.2303.08774.

[17] I. Tuomi, The Impact of Artificial Intelligence on Learning, Teaching, and Education: Policies for the Future, JRC Science for Policy Report, European Commission, 2019.

[18] O. Zawacki-Richter, V.I. Marín, M. Bond, F. Gouverneur, Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where Are the Educators?, Int. J. Educ. Technol. High. Educ. 16 (2019), 39. https://doi.org/10.1186/s41239-019-0171-0.