doi.org/10.51891/rease.v12i5.27351

EPISTEMIC RISKS OF UNSUPERVISED GENERATIVE AI USE IN AEC EDUCATION: AN EMPIRICAL STUDY

RISCOS EPISTÊMICOS DO USO NÃO SUPERVISIONADO DE IA GENERATIVA NA EDUCAÇÃO EM AEC: UM ESTUDO EMPÍRICO

RIESGOS EPISTÉMICOS DEL USO NO SUPERVISADO DE LA IA GENERATIVA EN LA EDUCACIÓN EN AEC: UN ESTUDIO EMPÍRICO

Matheus Goulart Mena Barreto[1]
Clara Vaqueiro Escosteguy[2]
Vitor de Oliveira Benvegnú[3]
André Lübeck[4]
Fabricio Longhi Bolina[5]
Débora Bretas Silva[6]
Eduardo Cesar Pachla[7]

ABSTRACT: The adoption of generative artificial intelligence (AI) in higher education has intensified concerns regarding learning quality, epistemic responsibility, and professional judgment in Architecture, Engineering, and Construction (AEC) education. This study investigates the consequences of unsupervised generative AI use in a design activity, focusing on performance, error propagation, and epistemic engagement. A two-phase design was adopted. Phase 1 applied a diagnostic survey (n = 244) on AI use and disclosure practices. Phase 2 conducted a quasi-experimental activity (n = 24), in which students solved a normative design task using ChatGPT (GPT-Vanilla) without prior instruction. Responses and interaction logs were evaluated through a rubric and an error taxonomy. Results indicate nearly universal and frequently undeclared AI use. In the GPT-Vanilla condition, performance was low (mean = 1.60/10), with errors concentrated in normative verification and multistep consistency. Of the 127 coded errors, 95.28% resulted from unverified acceptance of AI outputs, evidencing systematic error propagation and epistemic passivity. Findings reinforce the need for pedagogical mediation and explicit regulation, proposing "cognitive governance" as a core competency for responsible AI use in AEC education.

Keywords: Generative AI. AEC Education. Epistemic Responsibility. Metacognition. Cognitive Governance.

[1]Graduado em Engenharia Civil pela Universidade Federal de Santa Maria (UFSM) e mestrando em Arquitetura e Urbanismo pela mesma instituição.
[2] Graduada em Arquitetura e Urbanismo pela Universidade Federal de Santa Maria (UFSM) e mestranda em Arquitetura e Urbanismo pela mesma instituição.
[3]Graduando em Engenharia Civil pela Universidade Federal de Santa Maria (UFSM).
[4] Doutor em Engenharia Civil com ênfase em alvenaria estrutural e comportamento mecânico dos materiais, mestre em Engenharia Civil com foco em tecnologia do concreto e durabilidade de materiais, e graduado em Engenharia Civil, todos pela Universidade Federal de Santa Maria (UFSM). Atualmente, é professor adjunto do Departamento de Estruturas e Construção Civil da UFSM.
[5] Pós-doutor em Engenharia Civil pela Universidade Federal do Rio Grande do Sul (UFRGS), doutorado em Engenharia de Segurança ao Incêndio pela Universidade de Coimbra (Univ. de Coimbra – Portugal) e graduação em Engenharia Civil pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). É professor do Departamento de Engenharia Civil da Escola de Engenharia da UFRGS.
[6]Doutora em Engenharia Civil com ênfase em construção e infraestrutura pela Universidade Federal do Rio Grande do Sul (UFRGS), mestre em Engenharia com atuação em compósitos cerâmicos refratários pela Universidade Federal do Pampa (Unipampa) e graduada em Engenharia Civil pela mesma instituição. Atualmente, é Professora Assistente do Departamento de Arquitetura e Urbanismo da Universidade Federal de Santa Maria (UFSM).
[7]Doutor em Engenharia Civil com ênfase em construção e infraestrutura pela Universidade Federal do Rio Grande do Sul (UFRGS), mestre em Engenharia com foco em tecnologia dos materiais e compósitos cimentícios e graduado em Engenharia Civil pela Universidade Federal do Pampa (UNIPAMPA). Atualmente, é professor adjunto do Departamento de Estruturas e Construção Civil da Universidade Federal de Santa Maria (UFSM).

RESUMO: A adoção da IA generativa no ensino superior ampliou preocupações sobre aprendizagem, responsabilidade epistêmica e julgamento profissional na educação em Arquitetura, Engenharia e Construção (AEC). Este estudo investiga as consequências do uso não supervisionado de IA generativa em uma atividade de projeto, com foco no desempenho, na propagação de erros e no engajamento epistêmico. Adotou-se um delineamento em duas fases. A Fase 1 aplicou um questionário diagnóstico (n = 244) sobre uso e declaração de IA. A Fase 2 realizou uma atividade quase experimental (n = 24), em que estudantes resolveram uma tarefa normativa com apoio do ChatGPT (GPT-Vanilla), sem instrução prévia. As respostas e interações foram avaliadas por rubrica e taxonomia de erros. Os resultados indicam uso quase universal e frequentemente não declarado. Na condição GPT-Vanilla, o desempenho foi baixo (média = 1,60/10), com erros concentrados em verificação normativa e consistência multietapas. Dos 127 erros codificados, 95,28% resultaram da aceitação não verificada das respostas da IA, evidenciando propagação sistemática de erros e passividade epistêmica. Os achados reforçam a necessidade de mediação pedagógica e regulação explícita, propondo a "governança cognitiva" como competência central para o uso responsável da IA na educação em AEC.

Palavras-chave: Inteligência Artificial Generativa. Educação em AEC. Responsabilidade Epistêmica. Metacognição. Governança Cognitiva.

RESUMEN: La adopción de IA generativa en la educación superior ha generado preocupaciones sobre aprendizaje, responsabilidad epistémica y juicio profesional en Arquitectura, Ingeniería y Construcción (AEC). Este estudio analiza las consecuencias del uso no supervisado de IA generativa en una actividad de diseño, enfocándose en el desempeño, la propagación de errores y el compromiso epistémico. Se adoptó un diseño en dos fases. La Fase 1 aplicó una encuesta diagnóstica (n = 244) sobre uso y declaración de IA. La Fase 2 realizó una actividad cuasi experimental (n = 24), en la que estudiantes resolvieron una tarea normativa con apoyo de ChatGPT (GPT-Vanilla) sin instrucción previa. Las respuestas y registros de interacción fueron evaluados mediante una rúbrica y una taxonomía de errores. Los resultados indican un uso casi universal y frecuentemente no declarado de IA. En la condición GPT-Vanilla, el desempeño fue bajo (media = 1,60/10), con errores concentrados en la verificación normativa y la consistencia multietapa. De los 127 errores codificados, el 95,28% derivó de la aceptación no verificada de respuestas de IA, evidenciando propagación de errores y pasividad epistémica. Los hallazgos refuerzan la necesidad de mediación pedagógica, regulación explícita y gobernanza cognitiva para el uso responsable de IA en la educación en AEC.

Palabras clave: Inteligencia Artificial Generativa. Educación en AEC. Responsabilidad Epistémica. Metacognición. Gobernanza Cognitiva.

1. INTRODUCTION

Engineering education for architecture plays a central role in preparing professionals capable of acting responsibly, critically, and with technical rigor in contexts marked by high complexity and social impact. In Architecture, Engineering, and Construction (AEC) education, learning quality is directly associated with the development of critical thinking, deep conceptual understanding, and sound decision-making—core competencies for ethical professional practice (CLARIS; RILEY, 2012). The United Nations Sustainable Development Goal 4 (SDG 4 – Quality Education) reinforces this perspective by emphasizing not only access to higher education, but also the effectiveness and epistemic depth of learning outcomes (ONU, 2015). Research in engineering education consistently indicates that such quality requires moving beyond procedural knowledge toward higher-order cognitive processes, including analysis, evaluation, and technical judgment, as described in Bloom's taxonomy (ASOK et al., 2017).

Within this context, the rapid expansion of artificial intelligence (AI) in higher education has been particularly evident through the widespread adoption of generative language models such as ChatGPT. Empirical evidence shows that student use of these tools is increasingly common and often occurs spontaneously, without pedagogical mediation and prior to the consolidation of disciplinary conceptual mastery (SHARD et al., 2024). This accelerated adoption has created an asymmetry between technological accessibility and students' understanding of the limitations, assumptions, and vulnerabilities of generative AI systems. Although such tools may assist in organizing information and suggesting reasoning pathways, their outputs are generated through statistical pattern recognition rather than semantic understanding or internally verifiable reasoning, rendering them epistemically opaque (Yan et al., 2025).

Recent literature has raised ethical and epistemic concerns regarding unsupervised AI use in educational contexts. Identified risks include cognitive dependency, superficial reasoning, and the uncritical delegation of intellectual decision-making to algorithmic systems (DAKAKNI; SAFA, 2023). From a metacognitive perspective, effective learning in AEC education depends on students' ability to monitor, evaluate, and regulate their own problem-solving processes—capacities that may be weakened when reasoning is delegated to opaque generative systems. Issues of transparency, traceability, and responsibility further compromise learning integrity, as linguistically coherent responses may conceal conceptual errors and hinder meaningful validation of solution procedures (KISELEVA; KOTZINOS; DE HERT, 2022). In response to these concerns, international organizations such as the European Union, UNESCO, and the OECD have emphasized ethical principles for AI use, including governance, proportionality, transparency, and human-centered design (OECD LEGAL INSTRUMENTS, 2019; UE, 2018; UNESCO, 2021).

These risks are particularly salient in AEC education, where many learning activities involve normative reasoning, cumulative verification, and implications for the safety and reliability of the built environment. Educational tasks associated with structural conception and verification require not only formal procedures but also conceptual understanding, normative interpretation, and contextualized judgment (PETERSON et al., 2011). Errors in

ISSN: 2675-3375

such contexts extend beyond academic consequences and carry broader professional and social implications, aligning these challenges with Sustainable Development Goals related to infrastructure and urban sustainability (ONU, 2015).

Despite the growing body of research on AI in education, most studies focus on students' perceptions, attitudes, and general uses of AI tools (CHEN et al., 2023; HUSSAIN; AMERI SIANAKI; ABABNEH, 2019). Empirical investigations examining AI-supported performance in high-responsibility AEC learning tasks and technically grounded analyses of error patterns under unsupervised conditions remain limited (BASKARA, 2023; KOTSIS, 2024). Addressing this gap, the present study investigates the pedagogical and ethical risks associated with unsupervised use of ChatGPT in an AEC educational context, using a normative design task as an analytical artifact to examine performance, error propagation, and epistemic engagement.

2. THEORETICAL AND CONCEPTUAL FRAMEWORK

2.1. LEARNING, METACOGNITION, AND COGNITIVE GOVERNANCE IN AI-MEDIATED PROBLEM SOLVING

Engineering and AEC education aim to foster higher-order cognitive skills—such as analysis, synthesis, evaluation, and judgment—which are essential for technically responsible decision-making. Achieving these outcomes depends not merely on obtaining correct results, but on students' ability to structure their reasoning, verify assumptions, identify inconsistencies, and manage successive stages of problem solving. According to Bloom's Taxonomy, these competencies emerge at higher cognitive levels and are particularly relevant in cumulative and normative tasks, where conceptual or procedural errors in early stages tend to propagate and compromise subsequent results (SWART; DANETI, 2019).

When generative AI systems are used as cognitive support, this metacognitive layer becomes a central condition for learning quality. Large language models can produce fluent step-by-step narratives, but they remain epistemically opaque and do not provide reliable guarantees of internal verification or conceptual coherence (Yan et al., 2025). In this setting, students may shift from active evaluation to passive acceptance, delegating verification and decision-making to algorithmic outputs. Such delegation is a core epistemic risk in education, because it weakens the learner's obligation to justify and validate knowledge claims, undermining both learning and accountability (JOHRI, 2020).

ISSN: 2675-3375

To capture this challenge, this study adopts the concept of cognitive governance, defined as a situated competence through which learners integrate disciplinary knowledge, awareness of AI limitations, and metacognitive control over algorithmic support. Cognitive governance implies actively interrogating AI outputs, verifying intermediate steps, cross-checking results with normative constraints and problem conditions, and recognizing when AI-generated responses require external validation. Rather than rejecting AI, cognitive governance frames responsible use as an instructional and ethical competence that must be explicitly cultivated in AEC education, particularly in tasks that demand traceability and normative consistency.

2.2. RESEARCH GAP AND CONCEPTUAL CONTRIBUTION

Research on AI in higher education has expanded rapidly, but empirical studies still concentrate predominantly on student perceptions, attitudes, and self-reported uses of generative tools, offering limited evidence on what happens when these systems are used in technically sensitive, high responsibility learning tasks. In AEC education, this gap is consequential because many activities require cumulative reasoning, explicit verification, and alignment with normative design constraints, where errors are not isolated and may propagate systematically across steps.

Accordingly, there remains a need for empirical work that (i) examines performance under minimally constrained, unsupervised AI use; (ii) characterizes recurring technical and conceptual errors; and (iii) links these patterns to epistemic engagement and accountability mechanisms. This study addresses this gap by analyzing student performance and error propagation during a normative, step-based structural design exercise completed under an unsupervised AI condition, using rubric-based evaluation and error taxonomy. Conceptually, it contributes by operationalizing cognitive governance as an explanatory lens for distinguishing AI use that supports higher-order cognition from use that fosters epistemic passivity in high-responsibility AEC learning contexts.

3. POSITIONALITY STATEMENT

The authors are educators involved in undergraduate teaching and assessment within Architecture, Engineering, and Construction (AEC) programs. Their professional roles include curriculum delivery, academic supervision, and evaluation of student learning in normatively structured and high-responsibility educational contexts.

4. METHODS

4.1. OVERALL RESEARCH DESIGN

This study was conducted in two sequential and complementary phases, integrating a diagnostic survey with an applied quasi-experimental investigation. Together, these phases were designed to examine both the prevalence of generative artificial intelligence (AI) use in engineering-related education and the pedagogical and ethical implications of its unsupervised application in a technically sensitive learning task.

Phase 1 consisted of a descriptive survey administered to 244 students enrolled in engineering and architecture programs. The questionnaire was designed to characterize students' self-reported use of AI for academic purposes, including the tools employed, the types of courses in which AI was used, prior use of AI in graded assessments, and practices related to explicit acknowledgment of AI use in written academic work. This phase functioned as a contextual and diagnostic assessment of AI adoption patterns in higher education and provided the empirical basis for the design choices adopted in Phase 2. The survey protocol was approved by the Brazilian Research Ethics Committee system (CAAE: 91021025.8.0000.5346), in accordance with national ethical guidelines for research involving human participants.

Informed by the findings of Phase 1, Phase 2 was designed as an applied, quasi-experimental study focusing on the unsupervised use of generative AI in a reinforced concrete structures course. This phase involved a single cohort of 24 undergraduate students enrolled in Architecture and Urbanism, who completed a structural design task using ChatGPT without prior conceptual mastery of the topic, without training in prompt formulation, and without instructional supervision regarding AI use. This experimental condition, referred to as GPT-Vanilla, was intentionally defined to reflect spontaneous student behavior commonly observed in contemporary academic settings.

Phase 2 relied exclusively on fully anonymized and aggregated data derived from student submissions and AI interaction logs, with no possibility of individual identification. In accordance with Brazilian ethical regulations for social science and educational research (Resolution No. 510/2016, Art. 1, Item V), this phase did not require additional ethics committee approval.

Across both phases, the study employed a mixed-methods approach. Quantitative data were used to describe AI adoption patterns and task performance, while qualitative analyses focused on identifying conceptual errors, procedural omissions, and epistemic fragilities in student–AI interactions. This two-phase design enabled the study to progress from a broad characterization of AI use in engineering education to a focused examination of its consequences within a high-responsibility instructional context.

Generative artificial intelligence tools were not used for data generation, data analysis, or interpretation in this study; AI use was limited to language revision during manuscript preparation, as detailed in the Statement on Artificial Intelligence (Section 8).

4.2. PHASE 1: DIAGNOSTIC SURVEY ON AI USE IN ENGINEERING AND ARCHITECTURE EDUCATION

Phase 1 consisted of a diagnostic survey designed to characterize patterns of generative artificial intelligence (AI) use among students enrolled in engineering and architecture programs. The primary purpose of this phase was not hypothesis testing, but rather to establish an empirical baseline regarding the prevalence, context, and disclosure practices associated with AI use in higher education, thereby informing the design and justification of the applied study conducted in Phase 2.

4.2.1. PARTICIPANTS

The survey was conducted to a total of 244 undergraduate students enrolled in engineering and architecture programs. Participation was voluntary and anonymous, and no identifying personal or academic information was collected. Responses were aggregated for analysis to ensure confidentiality and prevent individual identification.

The sample included students at different stages of their undergraduate programs, allowing for a broad characterization of AI adoption across disciplinary and curricular contexts. Table 1 summarizes the main characteristics of the sample, including participants' age, and the academic period in which they were enrolled at the time of data collection.

Table 1 – Sample Characteristics of Phase 1 Survey Participants.

Characteristic	Category	N	Percentage (%)
Age (years)	18-25	215	88.11
	26-35	23	9.43
	36-45	5	2.05
	46+	1	0.41
Program	Engineering	98	40.20

	Architecture and Urbanism	**146**	**59.80**
Academic Stage	**Early stage (1st–3rd semester)**	**39**	**15.98**
	Mid stage (4th–6th semester)	**124**	**50.82**
	Advanced stage (7th semester or later)	**81**	**33.20**

Source: Authors.

4.2.2. SURVEY INSTRUMENT

The survey consisted of a structured questionnaire composed of closed-ended questions addressing five central dimensions:

1. prior use of artificial intelligence tools for academic purposes;
2. specific generative AI tools employed;
3. types of courses in which AI was used as learning support;
4. prior use of AI in graded assessments, such as assignments or exams; and
5. practices related to explicitly acknowledging AI use in written academic work.

The instrument was designed to capture students' self-reported behaviors rather than perceptions or attitudes, with emphasis on actual usage patterns.

4.2.3. DATA COLLECTION AND ANALYSIS

The survey was administered electronically during the academic term and analyzed descriptively using response frequencies and distributions. Phase 1 results were used diagnostically to justify the relevance of unsupervised AI use and to inform the design assumptions adopted in Phase 2.

4.3. PHASE 2: APPLIED QUASI-EXPERIMENTAL STUDY (GPT-VANILLA)

Phase 2 was designed as an applied, quasi-experimental study aimed at examining the pedagogical and ethical implications of unsupervised use of generative artificial intelligence in a technically sensitive learning context. Building on the diagnostic findings of Phase 1, this phase focused on a realistic instructional scenario in which students engaged with AI tools in the absence of prior conceptual mastery, structured guidance, or instructional supervision. The experimental design intentionally reflected spontaneous patterns of AI use commonly observed in contemporary higher education.

4.3.1. PARTICIPANTS AND EDUCATIONAL CONTEXT

Phase 2 involved 24 undergraduate students enrolled in a Reinforced Concrete Structures course within an Architecture and Urbanism program. The activity occurred before formal instruction on flexural design of reinforced concrete beams, so participants did not yet have consolidated conceptual mastery of the target procedure. To preserve anonymity, no individual identifiers were collected, and results were analyzed at an aggregated level; findings are interpreted through analytical rather than statistical generalization.

4.4. LEARNING TASKT AND EXPERIMENTAL CONDITION

This section describes the learning task assigned to students in Phase 2 and formally defines the experimental condition under which generative artificial intelligence was used. Together, these elements constitute the core of the applied investigation and provide the basis for the subsequent analysis of pedagogical and ethical implications.

4.4.1. LEARNING TASK DESCRIPTION

The learning task consisted of the structural design of a reinforced concrete beam subjected to simple bending, with the scope restricted to the determination of the required longitudinal reinforcement area. The task required students to follow standard design steps, including the determination of the design bending moment, calculation of design material strengths, verification of the neutral axis position, and computation of the required steel reinforcement area.

The problem statement included predefined design variables, such as applied loading, span length, cross-sectional geometry, and concrete strength class. Students were instructed to present their solutions in a step-by-step manner, clearly documenting intermediate calculations and results. The complete problem statement and its reference solution are provided in Supplementary Material.

4.4.2. PEDAGOGICAL RATIONALE FOR TASK SIMPLIFICATION

The task was limited to a fundamental flexural design problem aligned with the cohort's curricular stage. This scope enabled analysis of cumulative reasoning and verification behaviors without introducing unnecessary procedural complexity.

4.4.3. DEFINITION OF THE GPT-VANILLA CONDITION

The experimental condition adopted in this study, referred to as GPT-Vanilla, was defined to represent spontaneous and minimally constrained use of generative AI by students. Under this condition, students were allowed to use ChatGPT-4 as a support tool during task completion, without any prior training in prompt formulation, without guidance on effective AI interaction strategies, and without instructional supervision regarding the interpretation or validation of AI-generated outputs.

Additionally, students did not possess prior conceptual mastery of reinforced concrete beam design, as the topic had not yet been formally introduced in the course. The GPT-Vanilla condition therefore combined four defining characteristics: (1) spontaneous AI use, (2) absence of domain-specific conceptual knowledge, (3) lack of prompt engineering instruction, and (4) absence of real-time instructional mediation.

This condition was intentionally constructed to mirror common patterns of AI use identified in Phase 1 and to allow examination of student–AI interactions under conditions of limited epistemic scaffolding. By formalizing GPT-Vanilla as a methodological construct, the study enhances internal validity, supports analytical replication, and provides a clear foundation for subsequent ethical and pedagogical analysis.

4.4.4. TIMEFRAME AND AVAILABLE RESOURCES

Students completed the task in a 50-minute class session. They had access to the problem statement, calculators, notes, textbooks, design codes, and internet resources, and were allowed to use ChatGPT as support. The defining constraint of the GPT-Vanilla condition was the absence of instructor mediation during task completion.

4.5. DATA COLLECTION (PHASE 2)

Two data sources were collected: (i) students' submitted solutions, including intermediate steps and results, and (ii) complete ChatGPT interaction logs used during task completion. Data were anonymized before analysis, with submissions coded numerically and analyzed in aggregate. The combined dataset supported alignment between final outcomes and AI-mediated reasoning trajectories, including omissions, unverified steps, and propagation of incorrect intermediate values.

4.6. ASSESSMENT INSTRUMENT: RUBRIC-BASED EVALUATION

Student performance in Phase 2 was evaluated using a rubric-based assessment instrument specifically designed for reinforced concrete beam design tasks. The rubric operationalized key procedural and conceptual steps required for correct structural dimensioning and enabled systematic identification of errors across successive stages of the design process.

The rubric comprised seven criteria: (1) determination of the design bending moment; (2) calculation of design material strengths (fcd and fyd); (3) definition of cross-sectional geometry and effective depth; (4) determination of the neutral axis position and domain verification; (5) calculation of the required steel reinforcement area; (6) verification of the minimum reinforcement area; and (7) clarity, organization, and logical sequencing of the solution. Each criterion was assigned a predefined weight reflecting its relative importance within the overall design process, as presented in Table 2.

Table 2 – Rubric-Based Assessment Instrument for Reinforced Concrete Beam Design.

Criterion	Assessed Dimension	Maximum Score	Scoring Rule
C1	Determination of design bending moment (Md)	0.5	Full credit awarded when Md is correctly determined with proper application of safety factors; zero credit assigned when Md is incorrect.
C2	Calculation of design material strengths (fcd and fyd)	0.5	Full credit awarded when fcd and fyd are correctly calculated using appropriate safety coefficients; zero credit assigned when one or both values are incorrect.
C3	Definition of cross-sectional geometry and effective depth	1.0	Full credit awarded when b_w, h, and effective depth (d) are correctly defined and determined; zero credit assigned when geometric parameters are missing or incorrect.
C4	Neutral axis position and domain verification	3.0	Full credit awarded when the neutral axis depth is correctly determined and domain conditions are properly verified; zero credit assigned when verification is absent or incorrect.
C5	Calculation of required longitudinal reinforcement area	2.5	Full credit awarded when the required steel reinforcement area is correctly calculated; zero credit assigned when the calculated value is incorrect.
C6	Verification of minimum reinforcement area	2.0	Full credit awarded when the minimum reinforcement area is correctly verified and satisfied; zero credit assigned when the verification is incorrect.
C7	Organization and traceability of solution steps	0.5	Full credit awarded when results are presented in a clear, logical, and sequential manner; zero credit assigned when solutions are disorganized or lack traceable steps.

Source: Authors.

Scoring followed a binary logic: full credit was awarded when a criterion was met correctly, and zero credit was assigned when errors or omissions were identified. No partial credit was granted. This decision was intentional, as partial scoring could obscure the distinction between validated reasoning and unverified acceptance of generated outputs, particularly in contexts where generative AI is used as cognitive support.

In addition to technical correctness, transparency in the presentation of results was treated as a mandatory condition. Submissions that omitted intermediate steps, failed to document assumptions, or presented numerically correct results without traceable reasoning were penalized accordingly. This assessment logic reinforced expectations of accountability in problem-solving and aligned the evaluation process with higher-order cognitive activities associated with analysis, evaluation, and self-monitoring, as described in Bloom's taxonomy.

4.7. DATA ANALYSIS

Data analysis followed a mixed-methods approach, integrating quantitative and qualitative procedures to examine both task outcomes and AI-mediated reasoning processes.

Quantitative analysis focused on rubric scores obtained across the seven evaluation criteria. Descriptive statistics were used to examine overall performance patterns, frequencies of correct and incorrect solutions, and the distribution of errors across design stages. These analyses provided an overview of task performance under the GPT-Vanilla condition.

Qualitative analysis was conducted on both student submissions and ChatGPT interaction logs. This analysis aimed to identify recurrent types of conceptual errors, procedural omissions, and patterns of error propagation across successive design steps. Attention was given to instances in which AI-generated responses were accepted without verification, leading to compounding errors in subsequent calculations.

Integration of quantitative and qualitative findings enabled identification of relationships between performance outcomes and underlying reasoning behaviors. This analytical strategy supported examination of epistemic passivity, characterized by limited metacognitive monitoring and uncritical reliance on AI-generated outputs, as contrasted with active verification and conceptual reasoning.

4.8. VALIDITY, RELIABILITY, AND TRUSTWORTHINESS

Several strategies were adopted to enhance the validity, reliability, and trustworthiness of the study. Content validity was supported through alignment between the learning task, the assessment rubric, and the stated research objectives. The rubric criteria reflected canonical steps in reinforced concrete beam design and were consistent with instructional expectations for the course level.

Reliability was strengthened using clearly defined evaluation criteria and binary scoring rules, which minimized subjective interpretation. Evaluation was conducted collaboratively by the course instructor and teaching assistants, with discrepancies discussed until consensus was reached.

Trustworthiness of qualitative analysis was supported through triangulation of data sources, combining student submissions with AI interaction logs. This approach reduced reliance on self-reported reasoning and enabled reconstruction of student–AI interaction trajectories. Limitations related to sample size and contextual specificity are acknowledged, and findings are interpreted through an analytical generalization lens rather than statistical inference.

4.9. ETHICAL PRINCIPLES GUIDING THE STUDY

The study was guided by five ethical principles adopted to support the responsible use of generative AI in educational research contexts: governance and oversight, inclusion, proportionality, transparency, and human-centeredness. These principles were defined based on the seven core ethical principles proposed by Nguyen et al. (2023). From this broader framework, the principles most directly aligned with the objectives, methodological design, and scope of the present research were selected, ensuring both ethical rigor and practical applicability to the empirical context under investigation.

Governance and oversight were ensured through clear definition of experimental conditions and instructor supervision of study design, even in the absence of real-time instructional intervention. Inclusion was addressed by allowing equal access to AI tools and learning resources for all participants. Proportionality guided the decision to limit task complexity while maintaining technical relevance.

Transparency was operationalized through mandatory disclosure of AI interaction histories and explicit documentation of reasoning steps. Finally, human-centeredness was preserved by positioning AI as a supportive tool rather than an authoritative source, maintaining students and instructors as primary agents of decision-making and responsibility.

Together, these principles provided an ethical and pedagogical framework for examining generative AI use in engineering education and informed the interpretation of findings related to epistemic responsibility and cognitive engagement.

5. RESULTS

5.1. RESULTS OF PHASE 1: DIAGNOSTIC SURVEY ON AI USE IN ENGINEERING AND ARCHTECTURE EDUCATION

Phase 1 aimed to establish an empirical baseline regarding the prevalence, scope, and disclosure practices associated with the use of generative artificial intelligence (AI) in engineering and architecture education. The results confirm that AI tools are already deeply embedded in students' academic routines.

Responses to Question 1 (prior use of artificial intelligence tools for academic purposes) indicate that 94.7% of participants reported using AI tools for academic purposes, demonstrating that generative AI is no longer marginal but constitutes a routine cognitive support mechanism in higher education. Only a small minority reported no use of AI in academic activities.

Regarding Question 2 (specific generative AI tools employed), ChatGPT emerged as the dominant AI tool, being used by 97.4% of respondents who reported AI use. Other tools, such as Gemini (53.2%), DeepSeek (16.3%), and Copilot (12.4%), were also mentioned, but with substantially lower frequency. The prevalence of ChatGPT suggests a strong centralization of AI-mediated academic practices around a single generative model.

Results from Question 3 (types of courses in which AI was used as learning support) reveal that AI use occurs across multiple types of courses, rather than being confined to theoretically oriented subjects. The highest incidence was observed in theoretical courses (69.3%), followed by technological and construction-related courses (44.3%), design studios (43.4%), and calculation-based courses (42.2%). Planning (32.4%) and interdisciplinary courses (29.1%) also showed substantial AI use, while more specialized domains—such as transportation infrastructure and water systems—presented lower adoption rates. Notably, only 5.3% of respondents reported not using AI in any course, reinforcing the transversal nature of AI integration.

Question 4 (prior use of AI in graded assessments, such as assignments or exams) addressed AI use in graded academic activities. The results indicate that 82.8% of students had already used AI tools in assessed tasks, including assignments or exams, while 17.2% reported not doing so. This finding demonstrates that AI use has extended beyond informal study support and now directly mediates evaluative academic processes, often in the absence of explicit institutional regulation.

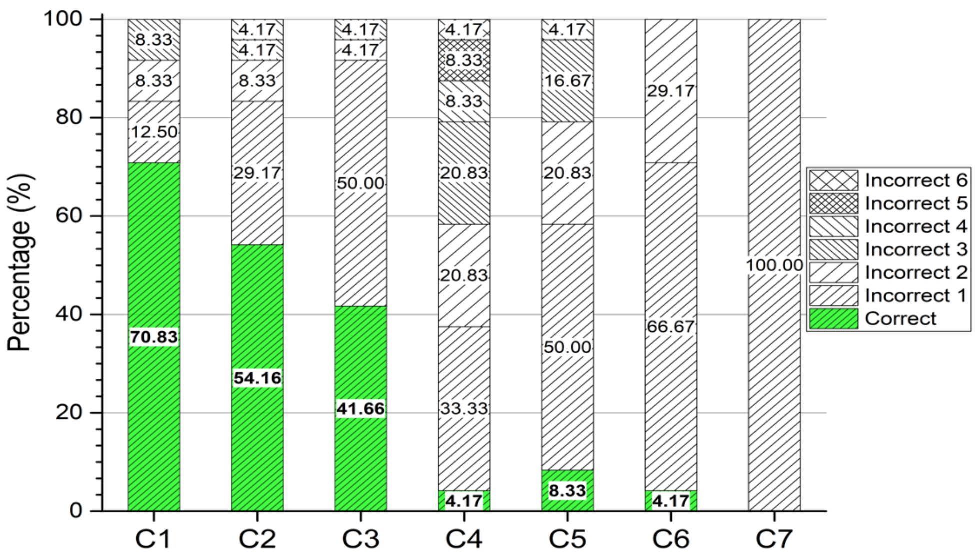
Finally, responses to Question 5 (practices related to explicitly acknowledging AI use in written academic work) reveal a marked discrepancy between AI use and formal disclosure. Only 29.5% of respondents reported explicitly acknowledging AI use in their academic work, while 63.5% used AI without formal reference, and 7.0% reported no AI use. This gap indicates that, although AI is widely used as a cognitive support tool, it remains largely invisible within formal academic documentation.

Taken together, the results of Phase 1 substantiate the assumption that unsupervised and spontaneous AI use is already normalized in engineering and architecture education. These findings provided empirical motivation for Phase 2, which investigates the consequences of such use in a technically sensitive instructional context.

5.2. RESULTS OF PHASE 2: OVERALL PERFORMANCE UNDER THE GPT-VANILLA CONDITION

Phase 2 examined student performance in a reinforced concrete beam design task conducted under the GPT-Vanilla condition. Figure 1 synthesizes the distribution of correct and incorrect outcomes across the seven criteria defined in the assessment rubric.

Figure 1 - Distribution of correct solutions and error types across the rubric criteria under the GPT-Vanilla condition (Phase 2).



Source: Authors.

As shown in Figure 1, student performance varied substantially across criteria. Higher success rates were observed in criteria associated with lower conceptual and procedural complexity (C1, C2, and C3), whereas a sharp decline in correct responses occurred in criteria requiring deeper structural reasoning, normative verification, and multi-step mathematical consistency (C4, C5, and C6). Criterion C7, related to the organization and traceability of solution steps, exhibited no correct responses, indicating a complete absence of structured reasoning across all submissions.

The progressive deterioration in performance across criteria suggests that errors introduced in early stages were not isolated but instead accumulated and propagated throughout the design process. Even in initial stages, however, accuracy remained below expectations for the academic level of the participants, indicating that the combination of limited conceptual knowledge and unsupervised AI use compromised task reliability from the outset.

5.3. DISTRIBUTION AND CLASSIFICATION OF ERRORS ACROSS DESIGN CRITERIA

To complement the aggregated view provided in Figure 1, Table 3 presents a detailed classification of errors identified in student submissions. Errors were coded using the format Cx–Iy, where "Cx" denotes the rubric criterion and "Iy" identifies a specific incorrect type within that criterion.

Table 3 – Detailed classification of errors identified in student submissions, coded by rubric criterion (Cx) and specific incorrect type (Iy), with corresponding frequency and percentage.

Code	Description	Frequency	Percentage (%)
C1-I1	Did not demonstrate how "d" was determined	3	12.50
C1-I2	Used the wrong equation	2	8.33
C1-I3	Provided incorrect concrete cover in the prompt	2	8.33
C2-I1	Used the 0.85 coefficient twice	7	29.17
C2-I2	Did not calculate f_{cd}	2	8.33
C2-I3	Did not specify f_{ck} in the prompt	1	4.17
C2-I4	Did not demonstrate how f_{cd} was calculated	1	4.17
C3-I1	Assumed $M_k = M_d$	12	50.00

C3-I2	Entered an incorrect M_k value in the prompt	1	4.17
C3-I3	Did not specify M_k in the prompt	1	4.17
C4-I1	Incorrectly applied the equation for the resultant compressive stresses in concrete	8	33.33
C4-I2	Did not explain the procedures or present the result for the neutral axis depth (*LN*)	5	20.83
C4-I3	Did not mention or determine *LN*, only determining the lever arm	5	20.83
C4-I4	Did not mention or determine *LN*, only determining the compression block depth (*a*)	2	8.33
C4-I5	Used the dimensionless parameter *KMD*, but with conceptual errors and incorrectly calculated variables	2	8.33
C4-I6	Correctly used the equation for the required steel area, but with an incorrectly calculated lever arm	1	4.17
C5-I1	Correct *As* equation, but incorrect variables were used	12	50.00
C5-I2	Did not indicate which equation was used to obtain the result	8	20.83
C5-I3	Incorrectly applied the equation for the resultant compressive stresses in concrete	4	16.67
C5-I4	*Mk* was not specified in the prompt, making it impossible to determine	1	4.17
C6-I1	Did not calculate $A_{s,min}$	16	66.67
C6-I2	Used an incorrect equation to determine *As,min*	7	29.17
C7-I1	Activity performed in non-numbered steps	24	100.00

Source: Authors.

As shown in Table 3, incorrect type frequency varied considerably across criteria. In Criterion 1, the most frequent incorrect type (C1-I1) involved the absence of explicit determination of the effective depth (d). In Criterion 2, the dominant incorrect type (C2-I1) resulted from the repeated application of the reduction coefficient 0.85, leading to incorrect calculation of design material strength. Criterion 3 exhibited a high incidence of incorrect types related to confusion between characteristic and design bending moments.

The most complex incorrect type patterns were observed in Criterion 4, which involved determination of the neutral axis position. Multiple incorrect types were identified, including incorrect application of equilibrium equations, omission of intermediate steps, and reliance on incomplete parameter definitions. Similar patterns of compounded incorrect types were observed in Criterion 5, where incorrect intermediate variables directly affected the calculation of the required reinforcement area.

Criterion 6 showed the highest concentration of errors related to omission, with 66.67% of students failing to verify the minimum reinforcement area, and an additional 29.17% applying incorrect equations. Criterion 7 exhibited uniform failure, as all submissions lacked numbered or logically structured solution steps.

5.4. ERROR PROPAGATION ACROSS THE STRUCTURAL DESIGN PROCESS

Analysis of the error distribution reveals a clear pattern of error propagation across the structural design workflow. Errors introduced in earlier criteria—such as incorrect geometric parameters, miscalculated material strengths, or incomplete prompts—systematically affected subsequent stages of the dimensioning process.

Because the determination of reinforcement area depends directly on previously calculated variables (e.g., neutral axis depth, lever arm, design moment, and material strengths), inconsistencies or omissions in early stages resulted in compounded inaccuracies. This propagation effect explains the sharp decline in performance observed in Criteria 4 through 6 and highlights the interdependence of procedural steps in reinforced concrete design.

5.5. SOURCES OF ERRORS: HUMAN-AI INTERACTION PATTERNS

Across all submissions, a total of 127 errors were identified. Of these, only 6 errors (4.72%) were attributable to inadequate prompt formulation or incorrect initial input by students. In contrast, 121 errors (95.28%) were associated with insufficient supervision of AI-generated outputs, including unverified calculations, omitted reasoning steps, and acceptance of incorrect intermediate results.

Further categorization of error sources indicates that:

62.20% of errors originated from conceptual misunderstandings in the interpretation and application of technical parameters;

19.69% resulted from incomplete or absent AI responses to required calculations;

13.39% were related to lack of transparency, characterized by missing equations, undocumented assumptions, or non-traceable solution paths.

These results demonstrate that performance limitations under the GPT-Vanilla condition were driven predominantly by interaction dynamics and lack of verification, rather than by isolated prompt errors. The findings underscore the vulnerability of unsupervised AI used in tasks that require cumulative reasoning and procedural consistency.

6. DISCUSSION

6.1. CONCEPTUAL FRAGILITY AND TECHNICAL LIMITATIONS OF CHATGPT IN NORMATIVE DESIGN TASKS

The results demonstrate that unsupervised use of ChatGPT in a normative structural design task led to performance substantially below expectations for the participants' academic level. The low mean score (1.60/10.00) reflects not isolated mistakes, but systematic conceptual fragilities that emerged across successive stages of the design process. Errors were concentrated in criteria requiring equilibrium reasoning, normative interpretation, and multi-step mathematical consistency, revealing limitations of large language models (LLMs) when applied to technically interdependent tasks.

A central failure mechanism involved incorrect or incomplete application of equilibrium relationships governing compressive stresses in concrete, which compromised determination of the neutral axis and propagated inconsistencies into subsequent calculations. This behavior is consistent with prior literature showing that LLMs operate through statistical pattern recognition rather than semantic understanding or internally verifiable reasoning (Yan et al., 2025). As a result, the model may alternate between multiple formulations that are individually plausible but contextually incompatible, without detecting internal contradictions.

Because reinforced concrete design relies on cumulative verification, early inaccuracies were not corrected but amplified across stages, explaining the sharp performance decline observed in later criteria. These findings indicate that, in the absence of verification and conceptual anchoring, generative AI outputs may undermine rather than support learning in normatively constrained engineering tasks.

6.2. TRANSPARENCY, TRACEABILITY, AND EPISTEMIC RESPONSIBILITY IN AI-MEDIATED PROBLEM SOLVING

Beyond numerical inaccuracies, a significant portion of errors was associated with the absence of transparent and traceable reasoning. Many submissions omitted equations, intermediate variables, or justification of assumptions, even when final numerical results appeared coherent. This lack of traceability compromises both technical validation and pedagogical value, as results without documented reasoning do not support conceptual learning.

These findings reinforce concerns regarding epistemic opacity in AI-generated outputs (KISELEVA; KOTZINOS; DE HERT, 2022). Linguistically fluent responses may conceal missing assumptions or flawed procedures, increasing the likelihood that students accept outputs uncritically. In this context, transparency functions as a cognitive safeguard rather than a formal requirement, enabling verification, reflection, and error detection.

The absence of documented reasoning also indicates a transfer of epistemic responsibility from students to the AI system. Instead of monitoring and validating solution paths, students frequently accepted AI-generated steps as authoritative. This dynamic weakens metacognitive regulation and undermines higher-order cognitive processes fundamental to engineering education (SWART; DANETI, 2019).

6.3. GOVERNANCE, SUPERVISION, AND THE CENTRALITY OF THE INSTRUCTOR IN THE AGE OF GENERATIVE AI

The predominance of errors attributable to unverified acceptance of AI outputs (95.28%) indicates that the main vulnerability observed in this study is not prompt formulation, but the absence of cognitive governance during AI use. Under the GPT-Vanilla condition, students rarely questioned, validated, or contextualized AI-generated responses, revealing a pattern of epistemic passivity.

These findings challenge narratives suggesting that generative AI may reduce the instructional role in higher education. In contrast, the results demonstrate that AI integration increases the importance of instructional mediation in normatively demanding contexts. In AEC education, instructors play a critical role as regulators of epistemic quality, ensuring verification, coherence, and accountability in cumulative problem-solving processes.

Within this framework, cognitive governance emerges as a central competence. It refers to learners' ability to integrate domain knowledge, awareness of AI limitations, and metacognitive control over algorithmic support. Cognitive governance is not assumed to arise spontaneously from access to AI tools; rather, it requires explicit pedagogical guidance and sustained instructional involvement.

7. CONCLUSION

This study examined the pedagogical and epistemic risks associated with the unsupervised use of generative artificial intelligence in a high-responsibility engineering education context. Through a two-phase research design combining a diagnostic survey and an applied quasi-experimental study, the investigation moved from a broad characterization of AI adoption in engineering and architecture education to a focused analysis of its consequences in reinforced concrete beam design tasks. The findings demonstrate that, although generative AI tools are already deeply embedded in students' academic practices, their spontaneous and unmediated use introduces significant challenges to learning quality and epistemic responsibility.

The results show that unsupervised use of ChatGPT leads to systematic conceptual errors, procedural omissions, and cumulative error propagation across design stages. Rather than functioning as a neutral cognitive aid, generative AI under the GPT-Vanilla condition fostered epistemic passivity, characterized by reduced metacognitive monitoring and uncritical acceptance of algorithmically generated outputs. These effects were especially pronounced in criteria requiring higher-order reasoning, normative interpretation, and multi-step mathematical consistency.

Beyond documenting performance limitations, this study contributes to engineering education research in several ways. Empirically, it provides evidence of AI behavior in a structurally sensitive engineering task under realistic classroom conditions. Analytically, it offers a detailed classification of error types and demonstrates how early-stage inaccuracies propagate throughout interconnected design processes. Theoretically, it advances the concept of cognitive governance as a situated competence integrating disciplinary knowledge, metacognitive control, and critical awareness of AI limitations. Educationally, the findings indicate that conceptual fragility—rather than technological access—is the primary barrier to meaningful AI-supported learning.

The study also carries important pedagogical and institutional implications. Given the near-universal adoption of generative AI identified in Phase 1, prohibition or restrictive policies are neither realistic nor effective. Instead, the findings support a shift toward explicit regulation, curricular integration, and instructional mediation. In this context, the role of the instructor becomes more central, as effective learning with AI requires supervision, validation, and guidance that preserve epistemic accountability.

Several limitations should be acknowledged. Phase 2 was conducted within a single course and disciplinary context, and the sample size does not support statistical generalization. The study focused on unsupervised AI use prior to formal instruction and does not capture potential learning gains under guided conditions. Future research should examine how different levels of instructional mediation, prompt training, and conceptual scaffolding affect AI-supported learning across engineering subdisciplines.

In conclusion, the educational value of generative AI in engineering depends less on technological sophistication than on the cognitive, pedagogical, and institutional structures that govern its use. When embedded within transparent, supervised, and conceptually grounded learning environments, AI may support higher-order learning; without such governance, it risks undermining the competencies that engineering education seeks to cultivate.

8. STATEMENT ON ARTIFICIAL INTELLIGENCE

Generative AI tools were used only as part of the experimental condition (student use of ChatGPT) and as an object of analysis; no AI tools were used by the authors for data generation, coding, data analysis, or interpretation.

During the manuscript preparation process, a generative AI tool (ChatGPT) was used exclusively for language revision and improvement of textual clarity and cohesion in portions of the manuscript. All AI-assisted revisions were reviewed, edited, and validated by the authors, who take full responsibility for the accuracy, integrity, and originality of the content presented.

No AI tools were used in the assessment of student work, coding of errors, classification of results, or derivation of conclusions. All empirical analyses and interpretations were conducted by the authors.

9. ACKNOWLEDGMENTS

The authors thank the OMITTED FOR BLIND REVIEW for the financial support through project No. OMITTED FOR BLIND REVIEW, which funded undergraduate research scholarships, the acquisition of permanent equipment, and consumable materials.

REFERENCES

ASOK, Divya et al. Active learning environment for achieving higher-order thinking skills in engineering education. **IEEE 4th International Conference on MOOCs, Innovation and Technology in Education (MITE)**, 2016. Proceedings [...] [S.l.]: Institute of Electrical and Electronics Engineers, 2017. p. 47–52. Disponível em: https://doi.org/10.1109/MITE.2016.069.

BASKARA, FX. Risang. AI-Driven Dynamics: ChatGPT Transforming ELT Teacher-Student Interactions. **Lensa: Kajian Kebahasaan, Kesusastraan, dan Budaya**, *[S. l.]*, v. 13, n. 2, p. 261, 2023. DOI: 10.26714/lensa.13.2.2023.261-275.

CHEN, Yu; JENSEN, Scott; ALBERT, Leslie J.; GUPTA, Sambhav; LEE, Terri. Artificial Intelligence (AI) Student Assistants in the Classroom: Designing Chatbots to Support Student Success. **Information Systems Frontiers**, *[S. l.]*, v. 25, n. 1, p. 161–182, 2023. DOI: 10.1007/s10796-022-10291-4.

CLARIS, Lionel; RILEY, Donna. Situation critical: critical theory and critical thinking in engineering education. **IEEE Women in Engineering Magazine**, *[S. l.]*, v. 7, n. 1, p. 32–36, 2012. DOI: 10.1109/MWIE.2013.2249993.

DAKAKNI, Deema; SAFA, Nehme. Artificial intelligence in the L2 classroom: Implications and challenges on ethics and equity in higher education: A 21st century Pandora's box. **Computers and Education: Artificial Intelligence**, *[S. l.]*, v. 5, n. August, p. 100179, 2023. DOI: 10.1016/j.caeai.2023.100179. Disponível em: https://doi.org/10.1016/j.caeai.2023.100179.

HUSSAIN, Shafquat; AMERI SIANAKI, Omid; ABABNEH, Nedal. **A Survey on Conversational Agents/Chatbots Classification and Design Techniques**. [s.l.] : Springer International Publishing, 2019. v. 927 DOI: 10.1007/978-3-030-15035-8_93. Disponível em: http://dx.doi.org/10.1007/978-3-030-15035-8_93.

JOHRI, Aditya. Artificial intelligence and engineering education. **Journal of Engineering Education**, *[S. l.]*, v. 109, n. 3, p. 358–361, 2020. DOI: 10.1002/jee.20326.

KISELEVA, Anastasiya; KOTZINOS, Dimitris; DE HERT, Paul. **Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations. Frontiers in Artificial Intelligence**Frontiers Media S.A., , 2022. DOI: 10.3389/frai.2022.879603.

KOTSIS, Konstantinos T. ChatGPT as Teacher Assistant for Physics Teaching. **EIKI Journal of Effective Teaching Methods**, *[S. l.]*, v. 2, n. 4, p. 18–27, 2024. DOI: 10.59652/jetm.v2i4.283.

NGUYEN, Andy; NGO, Ha Ngan; HONG, Yvonne; DANG, Belle; NGUYEN, Bich Phuong Thi. Ethical principles for artificial intelligence in education. **Education and Information Technologies**, *[S. l.]*, v. 28, n. 4, p. 4221–4241, 2023. DOI: 10.1007/s10639-022-11316-w.

OECD LEGAL INSTRUMENTS. Oecd Legal. *[S. l.]*, 2019. Disponível em: http://legalinstruments.oecd.org.

ONU. **Objetivos de Desenvolvimento Sustentável.** 2015. Disponível em: https://sdgs.un.org/goals. Acesso em: 6 set. 2025.

PETERSON, Forest; HARTMANN, Timo; FRUCHTER, Renate; FISCHER, Martin. Teaching construction project management with BIM support: Experience and lessons learned. **Automation in Construction**, *[S. l.]*, v. 20, n. 2, p. 115–125, 2011. DOI: 10.1016/j.autcon.2010.09.009. Disponível em: http://dx.doi.org/10.1016/j.autcon.2010.09.009.

SHARD; VAIDYA, Pankaj; KOUL, Sapna; ARORA, Ankur; SONIA; JAIN, Pankaj. Exploring Student Perspectives on ChatGPT in Higher Education. **Proceedings of the 2024 3rd**

Edition of IEEE Delhi Section Flagship Conference, DELCON 2024, *[S. l.]*, p. 1–6, 2024. DOI: 10.1109/DELCON64804.2024.10866660.

SWART, Arthur James; DANETI, Marllene. **Analyzing Learning Outcomes for Electronic Fundamentals Using Bloom's Taxonomy.** *[S. l.]*, p. 39–44, 2019. DOI: 10.1109/EDUCON.2019.8725137.

UE. **Orientações éticas para uma IA de Confiança: Grupo de peritos de alto nível sobre a inteligência artificial.** *[S. l.]*, 2018. Disponível em: https://digital-strategy.ec.europa.eu/pt/library/ethics-guidelines-trustworthy-ai.

UNESCO. **Recomendação sobre a Ética da Inteligência Artificial - UNESCO Digital Library**. , 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por.

YAN, Biwei; LI, Kun; XU, Minghui; DONG, Yueyan; ZHANG, Yue; REN, Zhaochun; CHENG, Xiuzhen. On protecting the data privacy of Large Language Models (LLMs): A literature review. **Proceedings - 2024 International Conference on Meta Computing, ICMC 2024**, *[S. l.]*, v. 5, p. 1–12, 2025. DOI: 10.1109/ICMC60390.2024.00008.