ISSN 2669-2023 (print) / ISSN 2538-8959 (on-line) **/ https://doi.org/10.47459/cndcgs.2026.121**

Artificial Intelligence in Higher Education: Safety, Reliability, and Risk Management in a Socio-Technical System

Ibraim Didmanidze[1], Ia Khasaia[2*], Nato Sherozia[2], Nino Jintcharadze[3], Nazi Jintcharadze[3], Aleksadre Bakuridze[4]

[1]Center of languages and information technologies, Faculty of Exact Sciences and Education, Batumi Shota Rustaveli State University, 32/35 Rustaveli/Ninoshvili st. Batumi, 6010, Georgia
[2]Department of Pedagogical Sciences, Faculty of Exact Sciences and Education, Batumi Shota Rustaveli State University, 32/35 Rustaveli/Ninoshvili st. Batumi, 6010, Georgia
[3]Department of European Studies, Faculty of Humanities, Batumi Shota Rustaveli State University, 32/35 Rustaveli/Ninoshvili st. Batumi, 6010, Georgia
[4]Department of Mathematics, Batumi Shota Rustaveli State University, 32/35 Rustaveli/Ninoshvili st. Batumi, 6010, Georgia

*Correspondence: *ia.khasaia@bsu.edu.ge*

Abstract

The study conceptualizes artificial intelligence in higher education as a socio-technical system in which human cognition, institutional norms, and algorithmic processes interact. It aims to develop a model of functional differentiation that preserves students' cognitive responsibility while enabling the safe and effective use of artificial intelligence. The study employs a mixed-methods design combining conceptual analysis with empirical data from student surveys conducted in higher education institutions in Georgia. The findings indicate that artificial intelligence enhances operational efficiency and supports academic tasks; however, its use for content generation is associated with reduced cognitive engagement and risks to knowledge reliability and academic integrity. The study concludes that clear functional boundaries, process-oriented assessment, AI literacy, and institutional governance are essential for responsible integration.

KEY WORDS: *artificial intelligence in higher education, socio-technical systems, functional differentiation, system reliability, risk management.*

Citation: Didmanidze, I.; Khasaia, I.; Sherozia, N.; Jintcharadze, N.; Jintcharadze, N.; Bakuridze, A. Artificial Intelligence in Higher Education: Safety, Reliability, and Risk Management in a Socio-Technical System. In Proceedings of the Challenges to National Defence in Contemporary Geopolitical Situation, Brno, Czech Republic, 7-10 September 2026. ISSN 2538-8959, https://doi.org/10.47459/cndcgs.2026.121

1. Introduction

In the contemporary era, artificial intelligence has emerged as a leading technological force fundamentally transforming the landscape of higher education [1-3]. However, higher education cannot be reduced to a mere system of knowledge transmission; rather, it functions as a social institution in which cognitive, ethical, and professional competencies are developed [4]. These competencies form the basis of institutional trust, democratic resilience, and the quality of decision-making in high-responsibility domains.

In this context, the integration of artificial intelligence into higher education should be understood not only as a pedagogical innovation, but also as a matter of socio-technical governance, closely linked to the quality of knowledge production, the distribution of responsibility, and the sustainability of institutional systems [5,6]. AI is no longer limited to facilitating access to information; it increasingly participates in the generation, structuring, and refinement of academic content. Students use artificial intelligence for a wide range of purposes, including information retrieval, text editing, translation, idea generation, and support in completing academic tasks, thereby embedding it within everyday academic activity [7-10].

As artificial intelligence becomes embedded in academic practice, it increasingly influences not only efficiency, but also authorship, cognitive engagement, and the distribution of intellectual responsibility [11-13]. This transformation has

Peer-review under responsibility of General Jonas Žemaitis Miltitary Academy of Lithuania and University of Defence, Czech Republic

generated a broad academic debate. On the one hand, AI is associated with risks such as cognitive offloading, reduced critical thinking, and academic misconduct [14-19]. On the other hand, it is viewed as a means of enhancing efficiency, accessibility, and personalization in learning [2,20].

This duality indicates that the impact of artificial intelligence cannot be explained solely by its technical capabilities. Rather, it depends on patterns of use, institutional regulation, and the educational contexts in which it operates [3,5,6]. The issue becomes particularly acute in relation to generative AI, whose capacity to produce academically acceptable content may reduce the need for visible cognitive effort and thereby challenge traditional understandings of authorship, assessment integrity, and students' intellectual responsibility [7,11,14,15]. Under these conditions, the key question is not whether AI should be used in higher education, but how the boundary should be defined between technological support and the substitution of cognitive processes.

Despite the growing body of research, existing studies tend to examine either the benefits or the risks of artificial intelligence in isolation. While risks related to academic misconduct and cognitive decline have been widely discussed [18], [19], comparatively limited attention has been devoted to integrative analytical frameworks that jointly address students' cognitive responsibility, the functional role of artificial intelligence, and the institutional conditions required for its safe and responsible use [3,5]. As a result, a conceptual gap persists, extending beyond pedagogy to issues of governance, reliability, and trust.

To address this gap, the present study proposes a conceptual framework based on functional differentiation, which defines a structured distribution of roles between students and artificial intelligence within academic activity. Within this framework, thinking, analysis, argumentation, interpretation, and the formulation of conclusions are conceptualized as components of students' cognitive responsibility, while artificial intelligence is positioned as a tool supporting executive and operational functions, such as structuring, editing, and information organization [21,22]. This differentiation provides a basis for balancing efficiency with the preservation of cognitive autonomy, while also supporting the transparency of responsibility and the quality of knowledge production.

The aim of the study is to develop a conceptual and empirical understanding of the role of artificial intelligence in higher education as a socio-technical system and to formulate a model of functional differentiation that supports responsible integration. The study adopts a mixed-methods approach, combining conceptual analysis with empirical data derived from a student survey, thereby providing exploratory evidence on patterns of use, perceived risks, and emerging governance challenges.

Research Objectives. To achieve the aim of the study, the following objectives are defined:

- To analyze patterns of students' use of artificial intelligence in academic activities, with particular attention to its impact on cognitive engagement and the reliability of knowledge production;
- To identify the cognitive processes and competencies that should remain with-in the domain of student responsibility as a basis for cognitive autonomy and academically reliable outcomes;
- To delineate and systematize AI-supported executive functions in a way that enhances efficiency without substituting cognitive processes;
- To develop a conceptual model of functional differentiation between students and artificial intelligence, ensuring a clear distribution of responsibility, the reliability of knowledge, and the safe integration of AI;
- To assess the risks associated with the misuse of artificial intelligence (cognitive, ethical, epistemic, and governance-related) and to define corresponding pedagogical and institutional implications, including mechanisms of regulation, oversight, and accountability.

Research Questions. The study seeks to address the following research questions:

- Which cognitive processes and learning competencies should remain within the domain of student responsibility in AI-supported education in order to preserve cognitive autonomy and ensure reliable knowledge?
- What functional role can artificial intelligence perform in academic activity without undermining critical thinking, argumentation, and independent decision-making?
- What types of risks (cognitive, ethical, epistemic, and governance-related) arise when artificial intelligence partially or fully substitutes student cognitive engagement?
- On what principles should teaching and assessment practices be adapted to ensure the ethical, reliable, and safe integration of artificial intelligence?
- How does the clarity or ambiguity in the distribution of functions between students and artificial intelligence relate to the reliability of knowledge, institutional trust, and the resilience of the educational system as a socio-technical system?

2. Theoretical Framework and Literature Review

The study is grounded in a structured analysis of contemporary research on artificial intelligence in higher education, with particular attention to its implications for cognitive processes, academic integrity, and institutional governance. Existing scholarship can be broadly grouped into three interrelated analytical strands, each offering a partial perspective on the role of AI in educational contexts.

The first strand emphasizes the functional benefits of artificial intelligence in education. Within this perspective, AI is primarily conceptualized as a tool that enhances the operational effectiveness of learning processes. Studies highlight its capacity to improve efficiency, expand access to information, and support personalized learning, as well as to assist students in tasks such as information retrieval, text production, and the organization of academic work [1, 2, 20]. While this line of research demonstrates the practical advantages of AI, it tends to focus predominantly on performance outcomes, often overlooking implications for cognitive engagement and the distribution of intellectual responsibility.

The second strand focuses on the risks associated with AI use, particularly in relation to cognitive processes and academic integrity. A growing body of literature suggests that the use of generative AI may contribute to cognitive offloading, reduced critical thinking, and an increased likelihood of academic misconduct [19, 15, 16]. The studies emphasize that the delegation of cognitive tasks to AI can weaken independent reasoning, diminish the quality of argumentation, and challenge established notions of authorship and intellectual ownership. However, this perspective, while analytically valuable, often adopts a predominantly risk-oriented stance without fully accounting for the constructive and supportive functions that AI may perform within structured learning environments.

A third and increasingly significant strand situates artificial intelligence within broader socio-technical and governance frameworks. From this perspective, AI is understood not merely as a pedagogical tool, but as a systemic factor embedded within institutional environments, interacting with norms, regulatory structures, and organizational practices [5, 6]. The approach draws attention to issues of accountability, oversight, and the distribution of responsibility in AI-mediated contexts. Nevertheless, while it expands the analytical scope beyond pedagogy, it does not provide a sufficiently operational and pedagogically actionable model for distinguishing between human and technological functions within academic activity.

Taken together, these strands reveal a fragmented research landscape. Although the benefits and risks of artificial intelligence have been extensively examined, comparatively limited attention has been devoted to integrative frameworks that systematically connect cognitive responsibility, the functional role of AI, and mechanisms of institutional governance. As a result, a conceptual gap persists in understanding how these dimensions interact within a unified socio-technical system.

This fragmentation is not merely thematic; it reflects competing assumptions about what constitutes educational value in AI-mediated contexts. Efficiency-oriented studies tend to prioritize productivity, access, and task performance, whereas risk-oriented approaches emphasize academic integrity, cognitive autonomy, and the preservation of authorship. Governance - oriented perspectives, in turn, focus on accountability, institutional control, and the conditions required for trustworthy system operation. The absence of a framework capable of integrating these dimensions limits the analytical coherence of the field.

To address this gap, the present study adopts a conceptual framework based on functional differentiation, which conceptualizes academic activity as a structured distribution of roles between students and artificial intelligence. Within this framework, cognitive processes - including thinking, analysis, argumentation, interpretation, and the formulation of conclusions - are defined as the responsibility of the student, whereas artificial intelligence is positioned as a tool supporting executive and operational functions, such as text structuring, linguistic refinement, and information organization [21, 22].

The proposed framework serves both as a theoretical foundation and as an analytical instrument for examining the relationship between cognitive engagement, the quality of knowledge production, and the responsible integration of artificial intelligence in higher education. By integrating pedagogical, cognitive, and governance perspectives, it enables a more coherent understanding of AI as a component of a broader socio-technical system.

On this basis, the methodological design of the study is structured to empirically explore how the differentiation between cognitive responsibility and AI-supported functions is reflected in students' academic practices.

3. Methodology

The study adopts a mixed-methods design, combining conceptual analysis with empirical research to operationalize the framework of functional differentiation and examine students' use of artificial intelligence in academic contexts. The approach enables the integration of theoretical and empirical perspectives, allowing for both the identification of usage patterns and the interpretation of associated cognitive, ethical, and governance-related risks.

The research design incorporates three interrelated components:

(1) a conceptual-analytical synthesis of the literature;

(2) a quantitative phase based on an online survey;

(3) a qualitative phase involving the analysis of open-ended responses.

This integrated design supports a comprehensive examination of how artificial intelligence is used in academic practice and how its role is perceived in relation to students' cognitive responsibility.

3.1 Quantitative Phase – Survey

Quantitative data were collected through an online survey instrument designed to examine students' use of artificial intelligence and their attitudes toward its application in academic contexts. The survey was distributed online using a convenience sampling strategy among students enrolled in higher education institutions in Georgia. Although the sampling

approach limits generalizability, it is appropriate for exploratory research aimed at identifying emerging patterns of behavior and perception.

A total of 230 students participated in the study, including both undergraduate and graduate students from different academic disciplines. Participants were aged between 18 and 25 years. The sample provides a broad descriptive overview of AI use across study levels, although it is not statistically representative.

3.1.1 Instrument Design

The survey instrument was developed based on a review of relevant literature on artificial intelligence in education, academic integrity, and cognitive engagement. It com-bined adapted elements from prior studies with items specifically designed to reflect the conceptual framework of functional differentiation.

The questionnaire consisted of four main components:

- demographic information;
- frequency and purposes of AI use;
- Likert-scale items measuring key constructs;
- open-ended questions.

The Likert-scale section included multiple items related to:

- cognitive engagement,
- academic efficiency,
- ethical considerations,
- institutional aspects of AI use.

Responses were measured using a five-point scale ranging from strong disagreement to strong agreement. The open-ended section included several questions designed to explore students' perceptions of the roles of AI and students, the boundaries of cognitive responsibility, ethical considerations, and the need for institutional regulation. Responses were collected in Georgian.

Prior to data collection, the instrument was pilot-tested with a small group of students to improve item clarity and interpretability. Minor adjustments were made accordingly, supporting the face validity of the instrument.

3.1.2 Analytical Focus

The quantitative phase was designed to:

- identify patterns of AI use;
- examine the functional role of AI (cognitive vs. executive);
- analyze students' attitudes toward academic responsibility and ethics;
- assess perceptions of the need for institutional regulation.

Quantitative data were analyzed using descriptive statistics, including frequencies, percentages, and mean values. The approach is consistent with the exploratory nature of the study, which aims to identify tendencies rather than establish causal relationships. In the case of multiple-response questions, percentages do not sum to 100% due to the possibility of selecting more than one option.

Given that the data are based on self-reported responses, the results should be interpreted as indicative of perceived practices and attitudes rather than direct observations of behavior.

3.2 Qualitative Phase – Analysis of Open-Ended Responses

The qualitative component of the study was based on responses to open-ended survey questions, aimed at providing a deeper interpretative understanding of students' experiences and perceptions. The responses addressed:

- practical uses of artificial intelligence;
- perceptions of the roles of students and AI;
- boundaries of cognitive responsibility;
- evaluations of ethical use;
- views on institutional regulation.

The textual data were analyzed using inductive thematic analysis. The coding process involved systematic reading, initial categorization, and iterative refinement of themes based on recurring patterns in the data. This approach enabled the identification of key thematic categories that complement and deepen the quantitative findings.

3.3 Data Analysis

Quantitative data were analyzed using descriptive statistics, including frequencies, percentages, and mean values. Qualitative responses were examined through inductive thematic analysis involving iterative coding and theme refinement. The findings were then integrated through methodological triangulation in order to identify recurring patterns and interpret their cognitive, ethical, and governance-related implications. This integration enhanced the robustness of the analysis and enabled the identification of both observable patterns of AI use and underlying dynamics related to cognitive engagement, ethical considerations, and institutional conditions. The analysis was guided by the conceptual framework of functional differentiation, which provided the basis for distinguishing between cognitive responsibility (assigned to the student) and executive functions (supported by artificial intelligence). This framework functioned both as a theoretical reference and as an analytical lens for interpreting the empirical findings. Given the reliance on self-reported data and descriptive analysis, the findings should be understood as exploratory, providing indicative insights into emerging trends rather than definitive conclusions.

3.4 Ethical Considerations

The study was conducted in accordance with established principles of academic ethics. Participation was voluntary, and the survey was anonymous. Participants were informed about the purpose of the study and the intended use of the data. All responses were processed in aggregated form to ensure confidentiality. Participants had the right to withdraw from the study at any stage without any negative consequences.

4. Results

The quantitative and qualitative findings provide insight into patterns of students' use of artificial intelligence and their perceptions within academic activity. The results describe reported usage practices and indicate emerging tendencies related to cognitive engagement, authorship, and institutional conditions. Within the limits of self-reported data, the findings offer exploratory evidence of how AI is integrated into academic work and where associated risks may arise.

4.1 Frequency and Purposes of AI Use

The findings indicate that the use of artificial intelligence is widespread among students and increasingly embedded in routine academic practice. Respondents report using AI regularly across a range of tasks, suggesting that it functions as a normalized component of academic activity rather than an occasional support tool.

The primary purposes of AI use include:

- information retrieval;
- explanation of complex concepts;
- text translation;
- grammatical and stylistic editing;
- text structuring;
- idea generation.

Less frequently, though still to a notable extent, students report using AI for generating full texts and preparing presentations.

Table 1. Main Purposes of AI Use (%)

Purpose	Value
Information retrieval	84
Explanation of complex concepts	77
Translation	71
Text editing	75
Idea generation	68
Full-text generation	42

The distribution suggests a functional predominance of operational and support-oriented uses. However, the relatively high rate of full-text generation (42%) indicates that, in a substantial proportion of cases, AI use extends beyond executive support into the domain of content production. This pattern may suggest movement from assistance toward partial cognitive substitution, raising concerns regarding authorship and the degree of independent intellectual engagement.

4.2 Cognitive Engagement

The analysis of Likert-scale responses indicates that students report relatively high levels of cognitive engagement when using artificial intelligence. The majority of respondents state that they:

- formulate core arguments independently;

- critically evaluate AI-generated information;
- retain control over the final version of their texts.

Table 2. Assessment of Cognitive Engagement (Mean Values)

Statement	**Mean**
Independent argumentation	4.1
Critical evaluation	3.9
Direct use of AI-generated answers	3.2

Despite these self-reported indicators of active engagement, the mean score for the direct use of AI-generated responses (3.2) suggests the presence of partial cognitive delegation. This may indicate a discrepancy between students' self-perception of intellectual autonomy and the practices reflected in their responses. While students emphasize their independence, the data also appear consistent with a degree of cognitive offloading, which may affect the depth of reasoning and the quality of argumentation.

4.3 Academic Efficiency

The findings indicate that students perceive artificial intelligence primarily as a tool for enhancing academic efficiency. Respondents report that AI:

- reduces the time required to complete tasks;
- improves the structure of written texts;
- enhances the technical quality of writing.

However, these efficiency gains are predominantly associated with time-saving and technical refinement rather than with the development of cognitive processes. This pattern appears consistent with a structural tension between operational efficiency and deeper learning. Over time, such reliance on efficiency-oriented use could contribute to reduced cognitive engagement and may limit the development of higher-order thinking skills.

4.4 Academic Integrity

The results indicate that students demonstrate a clear normative understanding of academic integrity in the context of AI use. Most respondents agree that:

- artificial intelligence should not replace their own thinking;
- authorship must be preserved;
- clear boundaries for ethical use are necessary.

At the same time, uncertainty remains regarding how AI use should be appropriately disclosed in academic work. This may suggest a gap between ethical awareness and practical implementation. The absence of clear standards for attribution and disclosure could increase the likelihood of unintentional violations of academic integrity.

4.5 Institutional Aspects

The findings indicate that students perceive existing institutional frameworks as insufficiently defined. Respondents highlight:

- the absence of clear guidelines for AI use;
- ambiguity in assessment criteria;
- limited guidance from instructors.

At the same time, students express support for:

- process-oriented assessment;
- transparent disclosure of AI use;
- the integration of AI literacy into academic curricula.

These responses suggest that the regulation of AI use is not yet fully institutionalized. The lack of consistent rules and assessment practices may create an environment in which AI use remains partially unstructured, increasing both ethical and cognitive risks.

4.6 Qualitative Findings

The analysis of open-ended responses identified several recurring thematic categories.

Perceived benefits:

- time efficiency;
- rapid access to information;

- simplification of complex topics.

Perceived risks:

- overreliance on technology;
- weakening of critical thinking;
- challenges to academic integrity

The themes are consistent with the quantitative findings and provide additional interpretative depth. Notably, students themselves recognize the dual nature of AI use, acknowledging both its practical advantages and its potential impact on cognitive processes. This alignment between quantitative and qualitative data supports the internal consistency of the findings.

4.7 Integrated Risk Analysis

Overall, the findings suggest several interrelated categories of risk associated with the use of artificial intelligence in academic contexts:

Cognitive risk — the potential weakening of independent thinking and analysis due to partial delegation of cognitive processes;

Ethical risk — ambiguity in authorship, attribution, and the appropriate use of AI-generated content;

Reliability risk — the use of unverified or AI-generated information, which may affect the quality and validity of knowledge;

Governance risk — insufficiently defined institutional rules, assessment criteria, and regulatory mechanisms.

The risks are interconnected and should not be considered in isolation. Their interaction reflects the broader transformation of higher education into an AI-mediated socio-technical environment. Within this context, the findings suggest that the impact of artificial intelligence extends beyond individual learning practices to include systemic challenges related to responsibility, regulation, and the quality of academic outcomes.

5. Discussion

The findings indicate that artificial intelligence occupies a dual role in students' academic practice. On the one hand, it is used primarily for executive and support-oriented purposes, such as information retrieval, translation, text editing, structuring, and idea generation. On the other hand, the reported use of AI for content generation and the partial direct use of AI-produced outputs suggests that its function may extend beyond operational assistance into cognitively more consequential domains.

The distinction is analytically important. The results indicate that the central issue is not AI use as such, but the point at which executive support begins to overlap with argumentation, evaluation, and other core cognitive processes that should remain the student's responsibility. In this respect, the findings provide empirical support for the framework of functional differentiation proposed in the study.

The coexistence of high levels of self-reported cognitive engagement with indicators of partial cognitive delegation further highlights this tension. This pattern suggests a potential misalignment between students' perceived intellectual autonomy and their reported practices. Such a discrepancy has implications not only for individual learning outcomes, but also for the epistemic quality of academic work, including the depth of reasoning and the coherence of argumentation [15, 16].

The findings also indicate that students' practices are shaped not only by individual choices, but by institutional conditions. Reported uncertainty regarding disclosure, assessment standards, and instructor guidance suggests that AI use is developing in a partially under-regulated environment. As a result, the responsible integration of artificial intelligence cannot be treated solely as a matter of personal ethics; it must also be understood as a matter of institutional governance.

At the same time, the data indicate that students demonstrate an intuitive distinction between cognitive and executive functions. Activities such as thinking, argumentation, and critical evaluation are consistently associated with human responsibility, whereas tasks related to structuring, linguistic refinement, and information organization are perceived as appropriate for AI-supported assistance. This alignment provides additional empirical grounding for the conceptual model proposed in the study.

Overall, the findings suggest that effective integration of artificial intelligence depends not on restricting its use, but on clearly defining functional boundaries between human cognition and technological support. Such differentiation enables the preservation of cognitive autonomy while allowing AI to contribute to operational efficiency in a controlled and transparent manner.

5.1 AI in Higher Education as a Socio-Technical System

From a systems perspective, the significance of artificial intelligence in higher education lies not merely in its instructional utility, but in its capacity to reshape the relationships among cognition, institutional norms, assessment practices, and accountability structures [12, 13, 6]. In this sense, higher education is increasingly functioning as an AI-mediated environment in which artificial intelligence operates not simply as a tool, but as a systemic factor influencing knowledge construction, assessment process-es, and the distribution of responsibility. This transformation reflects a broader recon-

figuration of academic activity, extending beyond individual learning practices to the structural organization of educational systems.

Within this context, the principle of functional differentiation acquires systemic importance. It establishes a structured distinction between the domain of student cognitive responsibility - such as thinking, analysis, argumentation, and interpretation—and the executive functions of artificial intelligence, including text structuring, editing, and information organization [21, 22]. The absence of clearly defined functional boundaries may result in ambiguity in authorship, reduced accountability, and challenges to the reliability of academic outcomes. From an assessment perspective, this distinction is critical, as it determines both the locus of intellectual contribution and the attribution of responsibility.

More broadly, higher education can be understood as a socio-technical subsystem in which patterns of cognition, responsibility, and technology use are formed. These patterns may extend beyond the educational context and influence professional practices in domains where reliability, accountability, and decision quality are essential. In this regard, the integration of artificial intelligence in education has implications not only for learning processes, but also for the functioning of wider socio-technical systems.

5.2 Implications for Safety, Reliability, and Risk Management

The findings suggest that the integration of artificial intelligence in higher education should be understood as a trust-dependent process requiring clearly defined governance and regulatory mechanisms. From a risk management perspective, AI use in academic contexts should be guided by principles such as transparency, accountability, traceability, and governability [6].

A central issue lies in the distinction between operational efficiency and educational reliability. The results indicate that AI enhances efficiency when used for executive tasks, while its use for content generation is associated with reduced cognitive engagement and weaker independent reasoning [15] – [17]. This divergence suggests that improvements in efficiency do not necessarily correspond to improvements in learning quality or the reliability of knowledge.

These patterns indicate that unregulated AI use may gradually affect cognitive process-es and, consequently, the quality of academic outcomes. From a systemic perspective, this constitutes a governance challenge rather than merely a pedagogical concern.

A risk-sensitive approach to AI integration should therefore include:

- clear differentiation between cognitive responsibilities and AI-supported functions;
- transparent disclosure of AI use;
- process-oriented assessment models;
- the development of AI literacy;
- coherent institutional policy frameworks aligned with academic integrity and accountability.

The relevance of these principles can be further clarified through comparison with other high-responsibility socio-technical systems, particularly in the defence sector. The comparison with defence systems is intended as an analytical analogy at the level of governance principles rather than functional equivalence between the two domains.

In defence contexts, the integration of artificial intelligence is governed by strict requirements related to validation, reliability assurance, controlled data inputs, and multilayered monitoring. Artificial intelligence does not substitute command authority; rather, it restructures decision-support architectures by providing analytically filtered representations of operational information. Under such conditions, artificial intelligence functions as a decision-support system, while final responsibility, judgment, and accountability remain with human actors. The safe operation of these systems depends on clearly defined functional roles and robust mechanisms of oversight.

A comparable governance principle applies to higher education. While the consequences differ in scale and nature, the uncontrolled delegation of cognitively decisive processes to AI may undermine the quality, reliability, and integrity of academic out-comes. In both contexts, artificial intelligence should be conceptualized as a support system operating within clearly defined boundaries and under human supervision [5, 3, 6]. From this perspective, the framework of functional differentiation can be understood as a risk-governance model. It provides a structured basis for preserving cognitive responsibility while enabling the effective and controlled use of artificial intelligence. By aligning pedagogical practices with principles of accountability, transparency, and role clarity, the framework supports the safe, reliable, and responsible integration of AI within higher education as a socio-technical system.

6. Practical Implications and Limitations

The findings provide several practical implications for higher education institutions, particularly in relation to the safe, reliable, and responsible integration of artificial intelligence.

First, there is a clear need for the development of coherent institutional frameworks governing AI use. These should include explicit guidelines, standardized disclosure practices, and clearly defined boundaries of academic responsibility. A shared understanding among students and instructors is essential to ensure consistent and transparent application across different types of academic tasks [5, 3].

Second, assessment models require adaptation. Traditional approaches focused solely on final outputs are insufficient in contexts where AI is widely used. Process-oriented assessment should therefore be implemented, incorporating intermediate stages of work, transparency of workflow, the development of argumentation, and evidence of cognitive engagement [5, 3].

Third, the development of AI literacy should be treated as a core educational objective. This includes not only technical competence, but also skills related to critical evaluation, verification of information, and responsible use of AI-generated content. Integrating AI literacy into curricula can support more reflective and informed engagement with technology [22, 3].

Fourth, institutions should embed the principle of functional differentiation into both policy and oversight mechanisms. Educational practices should clearly distinguish between tasks that require independent cognitive engagement and those where AI may be appropriately used as an executive support tool. This may include documenting stages of academic work, reviewing draft processes, and incorporating oral or analytical verification where necessary in order to distinguish cognitive responsibility from permissible AI-supported assistance [6, 5].

Overall, the findings suggest that the integration of artificial intelligence in higher education requires not only technological adaptation, but also structured, risk-sensitive governance. Such an approach enables a balance between efficiency and cognitive engagement, while supporting the reliability of academic outcomes and the responsible use of AI within broader socio-technical systems.

Limitations. Several methodological limitations should be considered when interpreting the findings of this study.

First, the use of a convenience sampling strategy limits the generalizability of the results. Although the sample includes students from different higher education institutions and academic disciplines, it does not constitute a statistically representative population.

Second, the study relies on self-reported data, which may be influenced by social desirability bias and subjective interpretation. In particular, reported levels of cognitive engagement and patterns of AI use may not fully correspond to actual behavior in academic practice.

Third, the use of descriptive statistical analysis enables the identification of general trends but does not allow for the establishment of causal relationships. The findings should therefore be interpreted as indicative and exploratory rather than explanatory.

Finally, the study is situated within the context of higher education in Georgia, reflecting specific institutional and cultural conditions. While the identified patterns may be relevant to other contexts, their direct transferability should be considered with caution.

7. Conclusion

The study has argued that the integration of artificial intelligence in higher education should be understood not only as a pedagogical development, but as a socio-technical transformation affecting knowledge construction, cognitive responsibility, academic authorship, and institutional trust. The findings indicate that students use AI primarily for executive and support-oriented purposes, but also in ways that may extend into content generation and partial cognitive delegation.

In response to this dual dynamic, the study proposed a framework of functional differentiation, according to which core cognitive processes remain the responsibility of the student, while artificial intelligence supports executive and operational functions. This framework provides a basis for balancing efficiency with cognitive autonomy and may also be interpreted as a risk-governance model for AI integration in higher education.

The findings further suggest that responsible use of AI requires clear institutional rules, transparent disclosure practices, process-oriented assessment, and the integration of AI literacy into curricula. Without such structures, the benefits of AI may be accompanied by risks to cognitive engagement and academic integrity.

More broadly, higher education functions as a formative socio-technical environment in which patterns of cognition, responsibility, and technology use are established. These patterns may extend beyond the educational context and influence professional practices in domains where reliability, accountability, and decision quality are critical.

Ultimately, effective AI integration depends not on technological expansion alone, but on preserving human cognitive centrality, maintaining clear functional boundaries, and implementing risk-sensitive governance mechanisms that support safe, reliable, and responsible use.

References

1. **Holmes W., Bialik M., Fadel C.** *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Boston: Center for Curriculum Redesign, 2019. ISBN 978-0-578-50200-2.
2. **Kasneci E., Sessler S., Kühl S., et al.** ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 2023, 103, 102274. Available from: https://doi.org/10.1016/j.lindif.2023.102274

3. **UNESCO**. *Guidance for Generative AI in Education and Research* [online]. Paris: UNESCO, 2023 [viewed 2026-04-19]. Available from: https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research
4. **Barnett R.** *Knowing and Becoming in the Higher Education Curriculum*. London: Routledge, 2009. ISBN 978-0415473613.
5. **High-Level Expert Group on AI.** *Ethics Guidelines for Trustworthy AI* [online]. Brussels: European Commission, 2019 [viewed 2026-04-19]. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
6. **National Institute of Standards and Technology (NIST).** *AI Risk Management Framework (AI RMF 1.0)* [online]. Gaithersburg: NIST, 2023 [viewed 2026-04-19]. Available from: https://doi.org/10.6028/NIST.AI.100-1
7. **Zhao W., Cox A.** Generative AI and academic writing: Opportunities and risks. *Computers and Composition*, 2024, 72, 102825. Available from: https://doi.org/10.1016/j.compcom.2024.102825
8. **Basch C., Hillyer G., Gold B., Yousaf H.** Artificial intelligence in higher education: Student knowledge, attitudes, and ethical perceptions in the United States. *SDGs Studies Review*, 2025, 6, e034. Available from: https://doi.org/10.37497/sdgs.v6istudies.342-12
9. **Mogavi R.H., Deng C., Kim J.J., et al**. ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans*, 2024, 2(1), 100027. Available from: https://doi.org/10.1016/j.chbah.2023.100027
10. **Sousa M.J., Cardoso P.** Use of generative AI by higher education students. *Electronics*, 2025, 14(7), 1258. Available from: https://doi.org/10.3390/electronics14071258
11. **Kotsis K.T.** The role of ChatGPT in academic writing: Pedagogical and ethical dimensions. *EthAIca: Journal of Ethics, AI and Critical Analysis*, 2025, 4, 420. Available from: https://doi.org/10.56294/ai2025420
12. **Leveson N.** *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge: MIT Press, 2011. ISBN 978-0262016629.
13. **Hollnagel E. Safety-I and Safety-II.** *The Past and Future of Safety Management*. Farnham: Ashgate, 2014. ISBN 978-1472423085.
14. **Georgiou G.P.** ChatGPT produces more "lazy" thinkers: Evidence of cognitive engagement decline [online]. *arXiv preprint arXiv:2507.00181*, 2025 [viewed 2026-04-19]. Available from: https://doi.org/10.48550/arXiv.2507.00181
15. **Gerlich M.** AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 2025, 15(1), 6. Available from: https://doi.org/10.3390/soc15010006
16. **Lee H.-P.H., Sarkar A., Tankelevitch L., et al.** The impact of generative AI on critical thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. New York: ACM, 2025. Available from: https://doi.org/10.1145/3706598.3713778
17. **Salido A., Syarif I., Sitepu M.S., et al.** Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies. *Social Sciences & Humanities Open*, 2025, 12, 101924. Available from: https://doi.org/10.1016/j.ssaho.2025.101924
18. **Bittle K., El-Gayar O.** Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 2025, 16(4), 296. Available from: https://doi.org/10.3390/info16040296
19. **Cotton D.R.E., Cotton P.A., Shipway J.R.** Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 2024, 61(2), p. 228–239. Available from: https://doi.org/10.1080/14703297.2023.2190148
20. **Lo C.K.** What is the impact of ChatGPT on education? *Education Sciences*, 2023, 13(4), 410. Available from: https://doi.org/10.3390/educsci13040410
21. **Deane P., Song Y.** The key practice: Discuss and debate ideas. *ETS Research Report Series*, 2015, 2015(1), p. 1–21. Available from: https://doi.org/10.1002/ets2.12079
22. **OECD**. *OECD Learning Compass 2030: A Series of Concept Notes* [online]. Paris: OECD, 2019 [viewed 2026-04-19]. Available from: https://www.oecd.org/education/2030-project/