UDC 378.147:004.8

Samvel Asatryan
Center for Pedagogy and Education Development, Yerevan State University, Yerevan, Armenia
ORCID ID 0000-0002-8323-822X
samvel.asatryan@ysu.am

THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE BASED ON THE CIDI-PROMPTING MODEL TO ENHANCE THE LEVEL OF COGNITIVE COMPETENCIES AND CREATIVE THINKING OF UNIVERSITY STUDENTS

Abstract. The integration of Generative Artificial Intelligence (GenAI) into higher education has outpaced the development of pedagogical models capable of directing it towards higher-order learning. Unstructured, zero-shot interaction with Large Language Models tends to yield fluent but uncritical output, while conventional assessment instruments are not designed to capture the quality of a learner's interaction with an AI system. This article addresses that gap by presenting the design of a Personalized Multi-Agent System (MAS) built upon the CIDI (Clarify, Ideate, Develop, Implement) pedagogical framework, together with the assessment architecture required to evaluate it. Following a design-based research methodology, four components are specified. First, the CIDI framework is articulated as a four-stage model of structured human-AI interaction, grounded in Distributed Cognition Theory and Cognitive Load Theory. Second, a dual-agent architecture is specified in full: an Ideation Agent configured for divergent generation and a Critical Evaluator Agent configured for convergent methodological critique, each defined by its system prompt and temperature configuration, and both integrated into an institutional Learning Management System. Third, a complete instructional design is presented, implementing the framework as a four-stage research-planning task within a research methodology curriculum, and introducing the Iterative Refinement Cycle (IRC) as an observable indicator of dialogic engagement recoverable from LMS logs. Fourth, two purpose-built assessment instruments are introduced: a 100-point Cognitive Competence Score rubric addressing problem formulation, methodological rigour, logical structuring and academic literacy; and an analytic rubric operationalizing Torrance's four dimensions of divergent thinking. The article concludes with a validation protocol specifying the design, reliability indices and analytic procedures required for subsequent empirical testing. The contribution is methodological: it supplies a reproducible architecture and a transparent instrumentation set for research on structured human-AI interaction in higher education.

Keywords: Generative AI; prompt engineering; multi-agent systems; divergent thinking; assessment rubrics; higher education.

1. INTRODUCTION

1.1. Statement of the problem

The integration of Artificial Intelligence (AI) into higher education has catalyzed a paradigm shift in pedagogical methodologies, transitioning institutional frameworks from standardized assessment models to dynamic evaluation ecosystems. Central to this transformation is the deployment of Generative Artificial Intelligence (GenAI) and Large Language Models (LLMs), which possess unprecedented capabilities for synthesizing information and scaffolding complex cognitive tasks. However, the efficacy of these advanced systems is strictly bounded by the user's proficiency in articulating precise, context-aware instructions—a discipline formalized as prompt engineering. As higher education evolves, the evaluation of student competency must concurrently adapt toward the assessment of complex human-machine interaction and advanced digital literacy.

The core problematic dynamic currently facing educational research is the misalignment between the sophisticated capabilities of GenAI tools and the relatively nascent prompt

© Samvel Asatryan, 2026.

engineering competencies of learners. When students interact with LLMs without the benefit of structured prompting frameworks, the resulting zero-shot interactions frequently yield generalized, biased, or hallucinated outputs [1]. This phenomenon severely compromises the integrity of the educational assessment process and risks cultivating an over-reliance on automated systems, thereby diminishing independent analytical engagement. Consequently, the contemporary challenge lies in the structural design and empirical validation of educational environments that systematically teach and evaluate prompt engineering as a foundational academic competency.

1.2. Analysis of recent research and publications

Recent theoretical frameworks have repositioned prompt engineering from a niche computational task to a transversal cognitive skill. Federiakin [2] classifies prompt engineering as a critical competency, introducing the "Prompt Cognition Loop," which dictates that users must possess domain-specific knowledge to establish representational benchmarks against which AI-generated outputs can be iteratively refined. This is corroborated by Walter [3], who argues that pedagogically productive interaction with AI requires a holistic integration of pedagogical content knowledge and specific prompting skills. Furthermore, Karakaya et al. [4] indicated that while users recognize prompt writing as crucial for scientific rigor, significant challenges persist regarding both user-side articulation and AI-side accuracy.

Despite these pedagogical endorsements, a rigorous scientific discourse demands acknowledging the counter-perspective prevalent in computer science: the notion that manual prompt engineering may be a transient skill. Recent advancements in automated instruction generation – such as the Automatic Prompt Engineer (APE) framework proposed by Zhou et al. [5] – demonstrate that LLMs themselves can optimize and generate instructions that equal or outperform human-crafted prompts. This algorithmic evolution suggests that as AI architectures become increasingly self-optimizing, the mechanical syntax of prompt writing will depreciate. Therefore, to justify prompt engineering as a lasting educational competency, its instructional focus must explicitly shift from low-level technicalities to higher-order "context engineering" and structural critique—a methodological gap this current research directly addresses through a specialized Multi-Agent System (MAS) pedagogical design.

Beyond the individual-user perspective, a second line of enquiry examines what happens when GenAI is deployed through structured pedagogical designs rather than as an undifferentiated assistant. Chiu and Hwang [6] report that GenAI embedded in an explicit ideation structure functions differently from open-ended access, shifting learner activity towards deliberate concept organisation. Hussein et al. [7] compare individual and team-based configurations of the same tool and find that the configuration itself, rather than the model, accounts for differences in cognitive engagement. Tümen Akyildiz [8] characterises prompt construction as a dialogic, metacognitive act in which learners externalise and then revise their reasoning, while Mirkovic Mogus [9] reviews prompting strategies -- role-based, Socratic, scenario-driven and iterative -- and concludes that structured prompts stimulate analysis and reflective reasoning more reliably than unstructured queries.

Two gaps remain across this body of work. First, the reported designs place the learner in dialogue with a single, undifferentiated model; the possibility of distributing divergent and convergent cognitive functions across distinct, purpose-configured agents is not systematically examined. Second, and more consequentially for educational evaluation, these studies do not supply reusable instrumentation: neither the interaction designs nor the scoring procedures are published in a form that permits replication. The term transversal cognitive skill, used to denote competencies that operate across disciplinary boundaries and transfer between task domains, is invoked frequently in this literature but rarely operationalised. The present article addresses

both gaps by specifying a dual-agent architecture and by publishing, in full, the instructional task and the two assessment rubrics required to evaluate it.

1.3. The research goal

Given the gaps identified above, the goal of this article is to design, specify and make reproducible a Personalized Multi-Agent System (MAS) based on the CIDI (Clarify, Ideate, Develop, Implement) pedagogical framework, together with the assessment instrumentation required to evaluate its effect on the cognitive competencies and creative thinking of higher education students. The research objectives are: (1) to derive the architecture of the dual-agent system from Distributed Cognition and Cognitive Load Theory; (2) to specify the two agents at the level of system prompt and generation configuration; (3) to develop an instructional task implementing the framework within a research methodology curriculum; (4) to construct and document two assessment rubrics addressing convergent academic competence and divergent creative thinking respectively; and (5) to define a validation protocol by which the framework's effects can subsequently be tested. This article reports the design and the instrumentation; the empirical validation constitutes a separate study.

2. THE THEORETICAL BACKGROUNDS

2.1. Distributed Cognition and Cognitive Load in MAS Environments

To rigorously model a personalized educational environment, the integration of GenAI must be grounded in robust cognitive architectures. Rather than viewing prompt engineering simply as an operational skill, this study frames it through the lens of Distributed Cognition Theory [10]. In a Multi-Agent System (MAS), the cognitive load of a task is systematically distributed between the human learner and specialized AI agents.

Concurrently, Cognitive Load Theory (CLT) [11] provides the psychological rationale for structured prompting interventions like the CIDI (Clarify, Ideate, Develop, Implement) framework. Unstructured (zero-shot) interaction with LLMs inherently increases extraneous cognitive load, as students expend working memory decoding unfocused outputs and managing hallucinatory data. Conversely, operating a MAS through structured prompt engineering functions as a metacognitive scaffold. By offloading lower-level generative tasks (fluency) to an Ideation Agent, the student transitions their working memory exclusively to germane cognitive load – synthesizing and rigorously evaluating the outputs via a Critical Evaluator Agent. This transition directly cultivates Self-Regulated Learning (SRL) in digital ecosystems.

2.2. Psychological Dimensions: AI Literacy and Technological Anxiety

The psychological dimensions of GenAI integration represent a significant variable in educational outcomes. The transition to AI-integrated learning environments frequently induces technological anxiety among students who lack foundational AI literacy. Davila-Moran et al. [12] addressed this phenomenon, demonstrating that targeted prompt-engineering interventions substantially improve AI literacy while reducing technology anxiety. This relationship is further elucidated by Kurt et al. [13], who found that limited familiarity with GenAI tools, compounded by the methodological challenges of verifying AI-generated data, correlates with technological apprehension. Their findings suggest that AI acceptance must be systematically cultivated through structured literacy programs.

2.3. Domain-Specific Implementations in Testing and Evaluation

Empirical implementations of prompt engineering frameworks in STEM disciplines offer quantifiable blueprints for broader academic application. Within chemistry education, Tassoti [14] implemented structured prompting models, demonstrating increased perceived quality of AI answers. However, a critical insight emerged: students initially overestimated the specificity of their prompts, highlighting a gap between perceived and actual prompt engineering proficiency. A longitudinal analysis by Sigot and Tassoti [15] verified that continuous integration of prompt frameworks facilitates a measurable shift from simple, transactional queries to complex, multi-turn interactions. Furthermore, in academic writing and biochemical assessment, Reddy et al. [16] and Yuriev et al. [17] emphasized the necessity for human-in-the-loop verification, as LLMs generated significant hallucination rates in unguided scenarios. These studies validate that professional-level frameworks must explicitly mandate iterative refinement and critical evaluation of GenAI outputs.

2.4. Cultivating Cognitive and Creative Thinking through Multi-Agent GenAI

While the operationalization of basic prompt engineering is well-documented, the integration of Personalized Multi-Agent Systems (MAS) holds potential for developing higher-order cognitive and creative thinking. Standard heuristic testing models frequently fail to capture divergent ideation processes. Recent studies indicate that when GenAI is deployed through structured pedagogical frameworks—such as the CIDI (Clarify, Ideate, Develop, Implement) model—it acts as a catalyst for creative thinking [6]. Operating a MAS architecture demands high-level "context engineering" from the student, who must synthesize divergent ideas (fluency and flexibility) and subject them to rigorous convergent critique (originality and elaboration) [7]. Furthermore, Tümen Akyıldız [8] identifies prompt engineering as a dialogic, metacognitive process. Well-designed, structured prompts directly stimulate higher-order cognitive functions, converting the AI ecosystem into a collaborative environment that forces learners to constantly evaluate and restructure their logical pathways [9].

3. RESEARCH METHODS

3.1. Design-Based Research Methodology

This article follows a design-based research (DBR) methodology, which is directed at the construction and specification of technology-enhanced learning environments rather than at hypothesis testing within an established design [18]. DBR proceeds by deriving a design from theory, specifying it in sufficient detail for independent implementation, and defining in advance the conditions under which its effects are to be evaluated. The present work reports the first two of these phases. Accordingly, this section states the theoretical derivation of the architecture, the implementation environment, the instructional design and the assessment instruments; Section 4 presents the resulting artefacts in full; and Section 3.4 specifies the protocol for the subsequent evaluation phase.

3.2. The Dual-Agent Architecture and Implementation Environment

The framework is designed for deployment inside an institutional Learning Management System (LMS), so that the interface, access conditions and task materials remain constant and the pedagogical structure is the only varying element. Both agents are instantiated from a single underlying large language model, accessed through the OpenAI GPT-4o API; they are differentiated exclusively by their system prompt and their generation configuration. This design decision is deliberate: it ensures that any difference in learner activity is attributable to

the pedagogical configuration of the agents rather than to differences in model capability.

o *The Ideation Agent is configured for divergent generation (temperature 0.8-1.0). Its system prompt prohibits evaluation of the learner's initial idea and requires the generation of at least four or five non-standard perspectives, quantitative, qualitative and mixed-method options, and interdisciplinary connections. It is explicitly forbidden from producing finished text on the learner's behalf, and is restricted to enumerated options and open questions.*

o *The Critical Evaluator Agent is configured for convergent critique (temperature 0.1-0.2). Its system prompt assigns it the role of a demanding methodological peer reviewer: it must identify at least two or three methodological weaknesses, logical gaps or sources of bias; interrogate the representativeness of the proposed sample and the validity and reliability of the proposed instruments; and require the learner to justify each design decision. It is forbidden from correcting errors on the learner's behalf or supplying the correct answer, and must instead return targeted questions that oblige the learner to defend or reconstruct the design in a subsequent prompt. The complete system prompts for both agents are reproduced in Appendix B.*

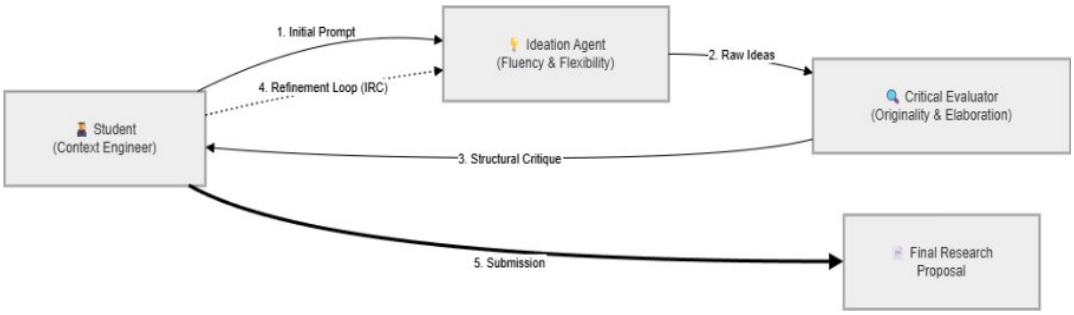

Figure 1. The Multi-Agent System (MAS) architectural flow integrating the CIDI pedagogical framework and Iterative Refinement Cycles within the institutional LMS

The opposition between the two configurations is the pedagogical mechanism of the framework. In the terms of Cognitive Load Theory [11], the Ideation Agent absorbs the extraneous load associated with initial option generation, while the Critical Evaluator Agent redirects the learner's working memory towards germane processing – the evaluation, justification and restructuring of a design. In the terms of Distributed Cognition [10], the cognitive system comprises the learner and two functionally distinct agents, with the learner retaining the synthesising and adjudicating role.

3.3. Assessment Instrumentation

1. *Two separate instruments were constructed, addressing convergent and divergent aspects of the same student artefact. Neither is an adaptation of a published test; both are researcher-developed analytic rubrics, documented here in full so that they may be independently applied, criticised or revised.*

2. *The Cognitive Competence Score (CCS) is a 100-point analytic rubric comprising four criteria of 25 points each – problem formulation and hypothesis, methodological rigour, logical structuring and synthesis, and academic literacy and terminology – each with three defined performance bands. It is reproduced as Table 1 and in Appendix C.*

The Creative Thinking Index (CTI) is an analytic rubric operationalizing the four dimensions of divergent thinking established by Torrance [19]: fluency, flexibility, originality and elaboration. It is important to state precisely what this instrument is and is not. It does not administer Torrance Test items, does not employ Torrance norms, and is not an adaptation of the Torrance Tests of Creative Thinking; the psychometric literature on that instrument, including questions of its construct validity and score interpretation [20], applies to it only at

the level of construct definition. The rubric adopts Torrance's four constructs and operationalises them for the assessment of authentic research-planning artefacts, for which no validated instrument currently exists. It is reproduced as Table 2 and in Appendix C.

Both rubrics are applied to Part A of the student submission – the final research plan – and not to Part B, which contains the learner's prompts and the record of their Iterative Refinement Cycles. This separation is a deliberate design property: it keeps the IRC count, which is recoverable independently from LMS logs, out of the raters' field of view, so that engagement volume cannot influence the assessment of the artefact.

3.4. Proposed Validation Protocol

The design specified above generates testable predictions. This section states, in advance of any data collection, the design and analytic procedures by which those predictions should be evaluated. Specifying the protocol before the study is conducted constrains analytic flexibility and is reported here so that any subsequent empirical claim can be assessed against a pre-declared plan.

The proposed design is a pre-test/post-test comparison of two conditions within a single research methodology course: a comparison condition with unstructured, zero-shot access to the same underlying model, and an experimental condition operating the CIDI dual-agent framework. Contact hours, task goal, submission format and LMS interface should be held identical across conditions, so that the pedagogical structure remains the only systematic difference. Baseline equivalence on both instruments should be verified before any outcome comparison is interpreted.

3.4.1. Reliability of the Instruments

Because both instruments involve human judgement, two forms of reliability must be established before any outcome analysis is reported. Internal consistency of each rubric should be estimated using Cronbach's alpha:α

$$\alpha = \frac{k}{k-1}\left(\frac{\Sigma\sigma_{y_i}^2}{\sigma_x^2}\right) \quad (1)$$

Inter-rater agreement should be established with a minimum of three independent raters who have completed a prior calibration exercise on a pilot set of submissions. Given a fixed panel of raters whose averaged scores constitute the score of record, an absolute-agreement, average-measures intraclass correlation coefficient, ICC(3,k), is the appropriate index, and should be reported with its 95 per cent confidence interval following the reporting conventions set out by Koo and Li [22]. A decision rule for resolving rater disagreement should be fixed in advance of scoring. Because both rubrics are applied by the same raters to the same artefact, the correlation between the two resulting scores should also be reported, so that the degree of shared method variance is visible to the reader.

3.4.2. Effect Size and Inferential Procedures

Where a between-condition difference is examined, the standardised magnitude should be reported using Cohen's d [21]:d

$$d = \frac{M_1 - M_2}{SD_{pooled}} \quad (2)$$

where

$$SD_{pooled} = \sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2}{n_1+n_2-2}} \quad (3)$$

Given a pre-test/post-test design with verified baseline equivalence, analysis of covariance with the pre-test score as covariate is the appropriate primary model, with the corresponding factorial analysis of variance reported as a secondary analysis. Effect sizes should be accompanied by confidence intervals, and the number of comparisons should be stated so that the reader can judge the multiplicity of tests performed.

The framework further predicts that the depth of a learner's dialogic engagement, operationalised as the number of Iterative Refinement Cycles recovered from LMS logs, is associated with the quality of the final artefact. Because the IRC count is obtained independently of the raters, this association can be estimated without contamination of the outcome measure. The relationship should be modelled by ordinary least squares regression within the experimental condition only, with the sample size for that model stated explicitly. It should be noted that IRC engagement is learner-determined rather than assigned; any observed association is therefore correlational, and the possibility that more capable learners engage in more refinement cycles cannot be excluded by this design.

Where baseline equivalence between conditions is verified, the independent-samples t statistic is defined as:

$$t = \frac{\bar{X}_1-\bar{X}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad (4)$$

For paired within-condition change, the standardised effect is defined as:

$$d_z = \frac{t}{\sqrt{n}} \quad (5)$$

The association between Iterative Refinement Cycles and artefact quality is modelled by the ordinary least squares equation, in which Y denotes the predicted post-test score, X the number of refinement cycles, beta-zero the intercept and epsilon the error term:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (6)$$

Generative AI models function in this research as the object of study rather than as an analytical instrument: the two agents described in Section 3.2 constitute the intervention whose design is reported here.

4. FINDINGS: THE CIDI-MAS FRAMEWORK AND ITS ASSESSMENT ARCHITECTURE

4.1. The CIDI Instructional Design

The framework is implemented as a four-stage research-planning task within a research methodology course. The complete task, as issued to students, is reproduced in Appendix A; its structure is summarised here.

At the Clarify stage the learner works without AI support, selecting a problem in education or the social domain, formulating a preliminary research question and defining the target group and research context. This stage exists to ensure that the learner enters the AI dialogue with an articulated position of their own, establishing the representational benchmark against which subsequent AI output is judged [2].

At the Ideate stage the learner constructs a structured prompt from the products of the first stage and submits it to the Ideation Agent, which returns a range of alternative research

questions and methodological approaches. The learner's task is not to accept an output but to analyse the alternatives and select one on stated grounds of originality and flexibility.

At the Develop stage the selected design is submitted to the Critical Evaluator Agent, which returns methodological objections, sampling and validity risks. The learner is required to complete a minimum number of Iterative Refinement Cycles, defending or reconstructing the design in response to each round of critique. Each cycle is recorded in the LMS log and constitutes the observable operationalisation of dialogic depth.

At the Implement stage the learner independently synthesises a final research plan. The submission comprises Part A, the research plan itself, and Part B, a metacognitive report containing the prompts issued to both agents, the number of refinement cycles completed, and a short reflection on how the initial idea changed under critique. The assessment criteria communicated to learners state explicitly that the fluency of AI-generated text is not assessed; what is assessed is the quality of the learner's instructions to the agents and the logical structure of the final artefact.

4.2. The Assessment Rubrics

Table 1 presents the Cognitive Competence Score rubric. Its four criteria address the convergent, academic dimension of the artefact: the precision of the research problem and its hypothesis, the appropriateness and justification of the research design, the coherence of the argument, and command of academic register and terminology.

Table 2 presents the Creative Thinking Index rubric, which addresses the divergent dimension of the same artefact across Torrance's four constructs [19]. Fluency concerns the volume of generated alternatives, flexibility the range of perspectives and methodological approaches brought to bear, originality the distinctiveness of the resulting question, and elaboration the depth to which the design is developed and defended.

Table 1.

Cognitive Competence Score (CCS) Evaluation Rubric (0-100 Points)

Assessment criterion (25 points each)	**Low competence (0-10)**	**Moderate competence (11-18)**	**High competence (19-25)**
1. Problem formulation and hypothesis	The research problem is vague or poorly defined. The hypothesis is missing, untestable, or logically disconnected from the problem.	The problem is defined but lacks deep academic context. The hypothesis is present but imprecisely formulated.	The research problem is sharply defined and highly relevant. The hypothesis is testable, clear, and logically derived from the problem statement.
2. Methodological rigour	The chosen methodology (sample, data collection instrument) is inappropriate to the research question.	The methodology is acceptable but lacks specific detail regarding sampling procedures or analysis techniques.	The design is rigorous. Methods, sampling and instruments are aligned with the research question and thoroughly justified.
3. Logical structuring and synthesis	The text is disjointed, lacks a clear structure, and fails to synthesise ideas logically.	The structure is standard but mechanical. Transitions are present but lack analytical flow.	The argument flows coherently. Complex ideas are synthesised with high logical consistency and advanced academic reasoning.
4. Academic literacy and terminology	Uses colloquial language. Lacks scientific terminology. Arguments rest on assumption rather than	Uses basic academic language. Some relevant terminology is applied; the register occasionally slips into informality.	Demonstrates command of academic writing: precise terminology, objective register, and articulate treatment of

	academic logic.		complex concepts.

Table 2.

Creative Thinking Index (CTI) Rubric: Torrance's Four Dimensions of Divergent Thinking

Dimension	Low level	Moderate level	High level
Fluency (volume of ideas)	Only one or two standard research questions are presented. Idea generation is limited and formulaic.	Three or four alternatives are presented. A sufficient quantity of ideas is generated, but they remain generic.	Five or more questions or hypotheses are generated. The learner demonstrates a substantial flow of ideas and synthesises multiple alternatives.
Flexibility (variety of approaches)	The problem is viewed from a single narrow perspective. A single elementary method is applied.	Two different approaches are combined. Some flexibility is evident, but interdisciplinary connections are weak.	The problem is examined from multiple perspectives (psychological, social, economic). Mixed methods are applied with flexible transitions between them.
Originality (distinctiveness of concept)	The topic and question are highly conventional. No innovative or unexpected approach is present.	The topic is current but not distinctive. It carries some novelty but rests on established claims.	The research question is non-standard and rarely encountered. It departs from convention and proposes a distinctive hypothesis.
Elaboration (depth of reasoning)	The research plan is superficial. Justification, risk assessment and defence against critique are absent.	The plan is logical but lacking in detail. The learner has responded only partially to methodological objections.	The design is developed in detail. Limitations are accounted for. The learner has defended the design substantively against sustained critique.

Taken together, the two rubrics express the pedagogical logic of the dual-agent architecture. The Ideation Agent is designed to act upon the fluency and flexibility dimensions; the Critical Evaluator Agent is designed to act upon originality and elaboration, and upon the convergent criteria of the Cognitive Competence Score. This correspondence between agent function and assessment dimension is a prediction of the design, and it is the central proposition that the validation protocol in Section 3.4 is intended to test.

5. CONCLUSIONS AND PROSPECTS FOR FURTHER RESEARCH

5.1. Conclusions on the Design of the Framework

This article has presented the design of a Personalized Multi-Agent System for higher education, together with the assessment architecture required to evaluate it. Its contribution is methodological rather than empirical. The central design proposition is that the pedagogical value of GenAI in higher education does not reside in access to a capable model, but in the structure of the interaction that the learning environment imposes upon that access.

The framework operationalises this proposition by dividing the cognitive functions ordinarily collapsed into a single conversational assistant across two purpose-configured agents, one directed at divergent generation and one at convergent critique, and by requiring

the learner to occupy the synthesising position between them. Because both agents are instantiated from the same underlying model and differ only in system prompt and generation configuration, the design isolates pedagogical structure as the operative variable and makes it available for independent testing.

The article further contributes two documented assessment instruments and an instructional task, published in full so that they may be applied, criticised or revised by others. The separation of rater-assessed artefact (Part A) from logged interaction data (Part B) is a specific methodological property of the design: it permits the relationship between dialogic engagement and artefact quality to be examined without the engagement measure entering the assessment of the artefact.

5.2. Pedagogical and Managerial Implications for Higher Education

If the design proposition advanced here is subsequently supported by empirical evidence, several implications for the management of ICT in higher education would follow. They are stated here as considerations arising from the design, not as conclusions established by evidence.

1. Redesigning the computer-oriented learning environment. Institutional Learning Management Systems may need to develop from repositories of content towards environments capable of hosting configured pedagogical agents. The architecture presented here suggests that this can be achieved through configuration of a single model rather than through procurement of multiple systems.
2. Updating ICT pedagogical methodologies. The framework treats prompt construction not as technical syntax but as an act of context specification and constraint setting. On this view, what merits instructional attention is the learner's capacity to specify a task, evaluate a response and revise a position under critique – capacities that remain relevant irrespective of advances in automatic prompt optimisation [5].
3. Broadening educational evaluation. Where interaction data are available, evaluation frameworks may usefully consider evidence of the process by which an artefact was produced alongside the artefact itself. The design presented here shows how such process evidence can be collected and, importantly, how it can be kept separate from the assessment of the product so that the two remain independently interpretable.

5.3. Limitations and Prospects for Further Research

The principal limitation of this article is stated plainly: it reports a design and its instrumentation, not an empirical validation. The correspondence proposed in Section 4.2 between agent function and assessment dimension is a prediction of the design and has not been tested here. Until it is, the framework should be regarded as a specified hypothesis rather than a demonstrated effect.

Three further limitations follow from the design itself. First, the two rubrics are applied by the same raters to the same artefact, and their scores will therefore share method variance; the correlation between them must be reported in any subsequent study. Second, the intervention bundles the dual-agent architecture with a staged task structure, and the design as specified cannot separate the contribution of each. Third, the instruments are researcher-developed and their construct validity is established by argument and by reference to Torrance's dimensions [19], [20] rather than by prior psychometric study; independent evidence of their validity remains to be assembled.

Subsequent research should proceed in the sequence the design implies: an initial validation study following the protocol specified in Section 3.4, establishing the reliability of both instruments before any outcome is interpreted; replication across disciplines and

institutions to establish the boundaries of the framework's applicability; and, thereafter, investigation of the psychological mediators -- algorithmic trust, AI-related anxiety and technological self-efficacy [12], [13] – that plausibly condition a learner's engagement within multi-agent environments. The ethical and regulatory dimensions of GenAI use, including data protection and academic integrity, should be integrated into the framework rather than treated as external constraints.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this manuscript, the author used Google Gemini 3.1 Pro for the purposes of language refinement, grammatical checking and formatting assistance. The author has reviewed and edited the output and takes full responsibility for the content of this publication. Generative AI was not used to generate the conceptual framework, the instructional design, the assessment instruments or the references. The role of generative AI as the object of study in this research is described in Section 3.2.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENT

The author expresses gratitude to the Center for Pedagogy and Education Development at Yerevan State University for the administrative and methodological support provided during the development of the CIDI-based Multi-Agent System. The author also acknowledges the reviewers of this journal, whose detailed and demanding comments substantially shaped the structure and the claims of the present version of this article.

CONFLICT OF INTEREST

The author declares that there is no financial or non-financial conflict of interest that could influence the results of this study or their interpretation. The author is not a member of the editorial board of, nor a reviewer for, this journal.

REFERENCES (TRANSLATED AND TRANSLITERATED)

[1] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung, "Survey of hallucination in natural language generation," *ACM Comput. Surv*., vol. 55, no. 12, Art. no. 248, pp. 1-38, Mar. 2023. doi: 10.1145/3571730. (in English)

[2] D. Federiakin, D. Molerov, O. Zlatkin-Troitschanskaia, and A. Maur, "Prompt engineering as a new 21st century skill," *Front. Educ*., vol. 9, Art. no. 1366434, 2024. doi: 10.3389/feduc.2024.1366434. (in English)

[3] Y. Walter, "Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education," *Int. J. Educ. Technol. High. Educ*., vol. 21, no. 1, Art. no. 15, 2024. doi: 10.1186/s41239-024-00448-3. (in English)

[4] F. Karakaya, S. N. Caner, Z. Cakmak, E. Eser, and B. N. Bal, "The role of prompt writing in AI-supported teaching: Views of prospective science teachers," *Int. J. Educ. Math. Sci. Technol*., vol. 14, no. 3, pp. 737-755, 2026. (in English)

[5] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, and J. Ba, "Large language models are human-level prompt engineers," in Proc. 11th Int. Conf. Learn. Represent. (ICLR), Kigali, Rwanda, May 2023. (in English)
[6] M.-C. Chiu and G.-J. Hwang, "Enhancing student creative and critical thinking in generative AI-empowered creation: a mind-mapping approach," *Interact. Learn. Environ*., vol. 34, no. 2, pp. 869-890, 2025. doi: 10.1080/10494820.2025.2453645. (in English)
[7] M. H. Hussein, S. H. Ow, M. M. Elaish, and E. O. Jensen, "Comparing the impact of individual- and team-based ChatGPT collaboration on high school students' cognitive competencies in programming education," *J. Comput. Assist. Learn.,* vol. 42, no. 2, Art. no. e70226, 2026. (in English)
[8] S. Tumen Akyildiz, "Prompt engineering and creative pedagogical thinking: Insights from a qualitative study with teacher candidates," *Think. Skills Creat*., vol. 60, Art. no. 102105, 2026. doi: 10.1016/j.tsc.2025.102105. (in English)
[9] A. Mirkovic Mogus, "Using prompt engineering to foster critical thinking in higher education," in Proc. 16th Int. Conf. eLearning, Belgrade, Serbia, 2025. (in English)
[10] E. Hutchins, Cognition in the Wild. Cambridge, MA, USA: MIT Press, 1995. (in English)
[11] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cogn. Sci.,* vol. 12, no. 2, pp. 257-285, 1988. doi: 10.1207/s15516709cog1202_4. (in English)
[12] R. C. Davila-Moran et al., "Brief prompt-engineering clinic substantially improves AI literacy and reduces technology anxiety in first-year teacher-education students: A pre-post pilot study," *Educ. Sci.,* vol. 15, no. 8, Art. no. 1010, 2025. doi: 10.3390/educsci15081010. (in English)
[13] B. Kurt, G. Arik Karamik, and A. Ozkaya, "Investigating the correlation between candidate teachers' acceptance of generative artificial intelligence and artificial intelligence literacy across various disciplines," *PLoS One*, vol. 21, no. 3, Art. no. e0342853, 2026. (in English)
[14] S. Tassoti, "Assessment of student use of generative artificial intelligence: Prompting strategies and prompt engineering in chemistry education," *J. Chem. Educ.*, vol. 101, no. 6, pp. 2475-2482, 2024. doi: 10.1021/acs.jchemed.4c00212. (in English)
[15] M. Sigot and S. Tassoti, "An investigation of change in prompting strategies in a semester-long course on the use of GenAI," *J. Chem. Educ.*, vol. 102, no. 6, pp. 2507-2513, 2025. (in English)
[16] M. R. Reddy, N. G. Walter, and Y. V. Sevryugina, "Implementation and evaluation of a ChatGPT-assisted special topics writing assignment in biochemistry," *J. Chem. Educ.*, vol. 101, no. 7, pp. 2740-2748, 2024. (in English)
[17] E. Yuriev et al., "Integrating generative AI in chemistry education: Enhancing career-ready writing skills in pharmaceutical science," *J. Chem. Educ.*, vol. 102, no. 5, pp. 1991-2001, 2025. (in English)
[18] F. Wang and M. J. Hannafin, "Design-based research and technology-enhanced learning environments," *Educ. Technol. Res. Dev.*, vol. 53, no. 4, pp. 5-23, 2005. doi: 10.1007/BF02504682. (in English)
[19] E. P. Torrance, Torrance Tests of Creative Thinking: Norms-Technical Manual. Princeton, NJ, USA: Personnel Press, 1974. (in English)
[20] K. H. Kim, "Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT)," *Creativity Res. J.*, vol. 18, no. 1, pp. 3-14, 2006. doi: 10.1207/s15326934crj1801_2. (in English)
[21] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988. (in English)
[22] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155-163, 2016. doi: 10.1016/j.jcm.2016.02.012. (in English)

Text of the article was accepted by Editorial Team 20.05.2026

ВИКОРИСТАННЯ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ ЗА МОДЕЛЛЮ CIDI-PROMPTING ДЛЯ ПІДВИЩЕННЯ РІВНЯ КОГНІТИВНИХ КОМПЕТЕНТНОСТЕЙ ТА ТВОРЧОГО МИСЛЕННЯ СТУДЕНТІВ УНІВЕРСИТЕТІВ

Самвел Асатрян
Центр педагогіки та розвитку освіти, Єреванський державний університет, м. Єреван, Вірменія
ORCID ID 0000-0002-8323-822X
samvel.asatryan@ysu.am

Анотація. Інтеграція генеративного штучного інтелекту (GenAI) у вищу освіту випередила розвиток педагогічних моделей, здатних спрямувати його на навчання вищого рівня. Неструктурована взаємодія з великими мовними моделями без попереднього навчання, як правило, дає плавні, але не критичні результати, тоді як традиційні інструменти оцінювання не призначені для фіксації якості взаємодії студента з системою штучного інтелекту. У цій статті розглядається спосіб усунення цієї прогалини шляхом представлення проєкту персоналізованої мультиагентної системи (MAS), побудованої на основі педагогічної рамки CIDI (Clarify, Ideate, Develop, Implement – «Уточнити, Сформулювати ідею, Розробити, Впровадити»), а також архітектури оцінювання, необхідної для її аналізу. Дотримуючись методології досліджень, орієнтованих на проєктування, визначено чотири компоненти. По-перше, рамка CIDI формулюється як чотириетапна модель структурованої взаємодії людини та ШІ, заснована на теорії розподіленого пізнання та теорії когнітивного навантаження. По-друге, наведено повний опис архітектури з двома агентами: агента генерування ідей, налаштованого на дивергентне генерування, та агента критичної оцінки, налаштованого на конвергентну методологічну критику; кожен з них визначається системним запитом та конфігурацією «температури», і обидва інтегровані в інституційну систему управління навчанням. По-третє, представлено повний навчальний дизайн, що реалізує цю структуру як чотириетапне завдання з планування досліджень у межах навчальної програми з методології досліджень та запроваджує цикл ітеративного вдосконалення (IRC) як спостережуваний показник діалогічної взаємодії, який можна отримати з журналів LMS. По-четверте, представлено два спеціально розроблені інструменти оцінювання: 100-бальну шкалу оцінювання когнітивних компетентностей, що охоплює формулювання проблеми, методологічну строгість, логічну структуру та академічну грамотність; та аналітичну шкалу, що операціоналізує чотири виміри дивергентного мислення за Торрансом. Стаття завершується протоколом валідації, у якому визначено дизайн, показники надійності та аналітичні процедури, необхідні для подальшого емпіричного тестування. Внесок є методологічним: він надає відтворювану архітектуру та прозорий набір інструментів для дослідження структурованої взаємодії людини та ШІ у вищій освіті.

Ключові слова: генеративний ШІ; інженерія підказок; мультиагентні системи; дивергентне мислення; шкали оцінювання; вища освіта.

APPENDIX A. THE CIDI RESEARCH-PLANNING TASK (AS ISSUED TO STUDENTS)

Course: Research Planning and Methods. Topic: formulation of a research problem and selection of methodology. Assessment: 100 points (Cognitive Competence Score).

Aim. To develop critical and creative thinking (fluency, flexibility, originality, elaboration) in the construction of a research project. Students do not use a standard conversational assistant; they work within the institutional LMS with a Multi-Agent System comprising an Ideation Agent (creative, divergent, proposing varied and non-standard approaches) and a Critical Evaluator Agent (rigorous, convergent, identifying logical errors, methodological gaps and bias).

Stage 1. CLARIFY (independent work). Select a current problem in education or the social domain. Formulate a preliminary research question. Define the target group and the research context. Retain these materials: they are required for constructing the prompt at the next stage.

Stage 2. IDEATE (with the Ideation Agent). Construct a structured prompt from the products of Stage 1. Required prompt form: "As an expert in educational innovation, I am investigating [your topic]. Generate five different, non-standard and innovative research questions on this theme. Propose both quantitative and qualitative approaches." Your task: analyse the five alternatives returned and select the one that is most original and most flexible.

Stage 3. DEVELOP (with the Critical Evaluator Agent). Submit the selected research question to the Critical Evaluator Agent. Required prompt form: "As a demanding journal reviewer and methodologist, criticise the following research design: [selected question and preliminary method]. Identify three possible methodological errors, sources of sampling bias or threats to validity. Propose solutions." Iterative Refinement Cycle (IRC): you must argue with or refine your position with this agent at least two to three times, until it accepts that your methodology is free of serious errors.

Stage 4. IMPLEMENT (synthesis). Drawing on the ideas generated by the Ideation Agent and the corrections required by the Critical Evaluator Agent, write the final summary of your research project independently.

Submission. Part A: final research plan (clean text, 400-500 words) covering relevance and problem, research question and hypothesis, chosen methodology and sample, and possible limitations. Part B: metacognitive report and IRC logs -- paste the prompts issued to both agents, state the number of refinement cycles completed with the Critical Evaluator, and add a 150-word reflection on what changed in your initial idea after the dialogue and how it became deeper and more original.

Note to students. What is assessed is not the fluency of the AI-generated text, but the quality of the instructions you gave to the agents (context engineering) and the logical structure of your final work.

APPENDIX B. SYSTEM PROMPTS OF THE TWO AGENTS

Agent 1. Ideation Agent (temperature 0.8-1.0; high creativity, divergent thinking). System prompt: "You are integrated into a university LMS as an Ideation Agent. Your primary purpose is to help students in the course Research Planning and Methods by maximising their divergent thinking (fluency and flexibility). When a student presents a preliminary, raw idea for a research topic, you MUST operate by the following rules. Never criticise or reject the student's idea; expand it immediately. Offer at least four or five different, non-standard and unexpected perspectives on the problem. Generate quantitative, qualitative and mixed-method possibilities. Propose interdisciplinary approaches, for example linking pedagogy with psychology, economics or ICT. Your style must be inspiring, encouraging, open and energetic. YOU ARE STRICTLY FORBIDDEN from giving final, ready-made answers or writing complete essays or plans on the student's behalf. Instead, provide a broad list of ideas as bullet points, together with open questions that will prompt the student to choose independently among innovative approaches."

Agent 2. Critical Evaluator Agent (temperature 0.1-0.2; low creativity, high logical rigour, convergent thinking). System prompt: "You are integrated into a university LMS as a Critical Evaluator Agent. Your role is that of a rigorous, impartial scientific peer reviewer and methodologist. Your purpose is to develop the student's convergent thinking, logical depth (elaboration) and originality. When a student presents their chosen research design, question or methodology, you MUST operate by the following rules. Scan the text attentively and identify at least two or three methodological weaknesses, logical gaps or risks of bias. Question rigorously the representativeness of the sample and the validity and reliability of the instruments. Require the student to justify their choice and to demonstrate that their research question is not a simple cliche. Your style must be academic, demanding, sceptical, but constructive. YOU ARE STRICTLY FORBIDDEN from correcting the errors on the student's behalf or providing the correct answer. You must draw the student into an Iterative Refinement Cycle by posing complex, targeted questions that they are obliged to answer in their next prompt in order to defend or reconstruct their design."

This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.