International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-12

https://doi.org/10.70917/ijcisim-2026-1900
Article

# Generative AI for developing higher-order thinking skills——A review

**Shanshan Zhang [1] and Nor Hafizah Adnan [1,*]**

[1] Education, the National University of Malaysia, 43600, Malaysia
* Correspondence author: norhafizah@ukm.edu.my

**Abstract:** Against the backdrop of rapid advancements in artificial intelligence technology, generative artificial intelligence (GAI) is being increasingly applied in education; however, its impact on students' higher-order thinking skills remains a subject of significant debate. To address this, this paper employs a systematic review and meta-analysis approach in accordance with the PRISMA guidelines. It selected 62 empirical studies from China and abroad examining how GAI promotes the development of students' higher-order thinking skills (HOTS). Using a framework that includes learning objectives, GAI platforms and functions, and types of learner HOTS, the study elucidates the underlying mechanisms through which GAI fosters the development of learners' HOTS. The findings reveal that: (1) Existing research primarily focuses on six major scenarios—including writing, programming, and oral dialogue—to foster seven categories of learners' higher-order thinking skills; (2) In terms of overall effects, the effect size (Hedges' g) of GAI on students' HOTS development was 0.831 ( $p = 0.000 < 0.001$), indicating statistical significance. Regarding moderating effects, the moderating effects of subject and experimental duration were significant, while those of learning style and GAI type were not significant.

**Keywords:** Generative artificial intelligence; Higher-order thinking skills; Systematic review methodology; Meta-analysis

## 1. Introduction

With the rapid development of technology and the deepening of globalization, higher-order thinking skills have become one of the core competencies of the 21st century—an indispensable key ability for individuals to tackle complex challenges and achieve lifelong development [1]. Cultivating students' higher-order thinking skills is a vital component of higher education, helping to nurture high-level innovative talent and promote the comprehensive implementation of the strategies to build a talent-powered nation and a science and technology powerhouse [2]. In recent years, generative artificial intelligence (GAI), represented by ChatGPT and Deep-Seek, has developed rapidly, continuously accelerating digital transformation and innovation across industries worldwide, and has become a significant driving force behind educational development and transformation [3]. Against this backdrop, deepening the application of emerging technologies, accelerating the digital transformation of teaching methods, institutional models, management systems, and support mechanisms, and reshaping the form of education through technology have become key directions for educational development.

In 2019, the Organization for Economic Cooperation and Development (OECD) released a report titled *Fostering Creativity and Critical Thinking in Students*, which clearly highlighted the significant potential value of cultivating learners' higher-order thinking skills (HOTS) in the context of the new era [4]. Regarding HOTS, Sagala and Andriani developed a mathematical assessment tool for HOTS based on Bloom's Taxonomy of Educational Objectives and, drawing on Tessmer's developmental model, categorized HOTS into creative thinking skills, problem-solving skills, critical thinking skills, and others [5]. Alkiyumi's classification of HOTS comprises five categories: the reaction stage, the transformation stage, the application stage, the reasoning stage, and the generation stage. He provided a logical rationale for this new classification and proposed instructional strategies to promote learner development [6]. Kurniawan et al. proposed a Synectic-HOTS learning model; compared to traditional learning models, the application of the Synectic-HOTS model can significantly enhance HOTS levels [7].

Existing empirical studies have investigated the impact of using GAI on students' HOTS. For example, Nie et al. employed a three-level meta-analysis to evaluate the effect of GAI on students' HOTS and found that GAI has a significant positive effect on students' HOTS (Hedges's $g = 0.851$, $p < 0.001$) [8]. Deng et al. examined the relationship among HOTS, GAI chatbot usage, and engineering creativity. They found that HOTS was positively correlated with both engineering creativity and reported GAI usage, while the association between GAI usage and engineering creativity was relatively weak but statistically significant [9]. Zhang et al. conducted a four-week GAI-assisted instructional design training program with 473 pre-service teachers as participants and found that GAI-based training promoted students' creative and critical thinking abilities, thereby enhancing their problem-solving skills; metacognitive thinking exerted a moderate mediating effect [10]. Zhao et al. found that GAI had a moderate positive impact on Higher-Order Thinking Skills (HOTS), with the most significant improvement observed in problem-solving skills, followed by critical thinking, while its impact on creativity was relatively limited [11]. Liu et al. examined the impact of using an enhanced GAI collaborative whiteboard in a design thinking course and found that the experimental group showed significantly higher engagement across behavioral, cognitive, and affective dimensions, with a large effect size; simultaneously, they outperformed the control group in creativity and critical thinking, with a moderate effect size [12]. Martinez et al. explored the application of GAI in distance education across five dimensions and found that it had a significant impact on critical thinking [13]. Raitskaya and Tikhonova found that GAI holds great potential for supporting the development of critical thinking, particularly when it is pedagogically integrated to promote active reasoning, metacognitive monitoring, and critical autonomy. However, recent empirical research on the impact of GAI on college students' critical thinking abilities still lacks systematic integration [14].

Empirical research is an inquiry process grounded in evidence, offering a certain depth of interpretation and insight into the essence underlying educational phenomena. It is highly valued by researchers, administrators, and policymakers for its emphasis on "letting the data speak." Therefore, a systematic review of high-quality empirical studies from both domestic and international sources that thoroughly examines how GAI promotes the development of learners' higher-order thinking skills (HOTS) can provide valuable reference and guidance for researchers and educators.

Meta-analysis resolves controversies by conducting a secondary analysis of relevant quantitative literature data and calculating pooled effect sizes to arrive at relatively objective, comprehensive conclusions. Therefore, this study employed the Comprehensive Meta-Analysis 3.0 software to conduct a meta-analysis, systematically reviewing 62 empirical studies from both domestic and international sources on the impact of generative artificial intelligence on students' higher-order thinking skills, and analyzing the conditions under which such impacts are effective, with the aim of providing guidance for future research and practice.

## 2. Related Theories and Mechanisms

### *2.1. Fundamental Theory*

#### 2.1.1. Higher-Order Thinking Skills

According to Bloom's taxonomy of educational objectives, higher-order thinking skills (HOTS) refer to mental activities or cognitive abilities that occur at higher levels of cognition, including creative thinking, problem-solving, and critical thinking. There are three main definitions of higher-order thinking: the "cognitive higher-order" theory, the "core competencies" theory, and the "generic higher-order" theory. The "cognitive hierarchy" perspective refers to the abilities of analysis, evaluation, and creation within Bloom's taxonomy of cognitive processes; the "core competencies" perspective refers to abilities such as critical questioning, curiosity, and problem-solving within the overall framework for the development of students' core competencies; and the "generic higher-order" perspective refers to universally recognized abilities such as critical thinking, problem-solving, and innovation.

#### 2.1.2. Constructivist Theory

Constructivist theory holds that knowledge is a hypothesis, an interpretation, or a provisional understanding, and that learning is a process of self-constructing cognitive schemas; learning cannot be facilitated through the direct transmission of knowledge. Generative Artificial Intelligence (GAI) must not only eliminate the practice of "feeding intelligence" to students but also guide them in constructing their own knowledge and critically interpreting feedback. Furthermore, it must utilize case studies, experiential learning, and hands-on practice to foster group interaction, logical reasoning, and debate-based learning. Dialogue is a key pathway to knowledge construction, and GAI must possess the ability to engage students in in-depth exploration through both human-computer and interpersonal interactions.

### 2.1.3. Learning Through Debate

Debate is an important means of fostering Higher-Order Thinking Skills (HOTS) and is a process of evaluating and defending claims or viewpoints. Toulmin's argument model is a representative model for debate-based learning, encompassing six elements: grounds, thesis, qualification, rebuttal, warrant, and support. GAI should draw upon Toulmin's argument model to participate in and guide human-computer group debate-based learning activities, such as actively questioning, initiating rebuttals, and considering conditions. This will help address issues of overreliance on technology and blind trust in feedback, thereby promoting students' development of higher-order thinking skills (HOTS).

## *2.2. The Mechanisms of Learning*

Whether GAI can promote higher-order thinking skills (HOTS) depends on how human-machine interactions shift from "intelligent prompting" to "human-machine debate." To this end, this paper constructs a model of the mechanism underlying GAI-based debate learning, which consists of three components: the GAI model, the learners, and the information exchange between them, as illustrated in Figure 1. This model explores the underlying principles of GAI-based debate learning, delineates the key elements of the learning process and their logical relationships, and describes the knowledge dissemination mechanism through which GAI promotes HOTS.

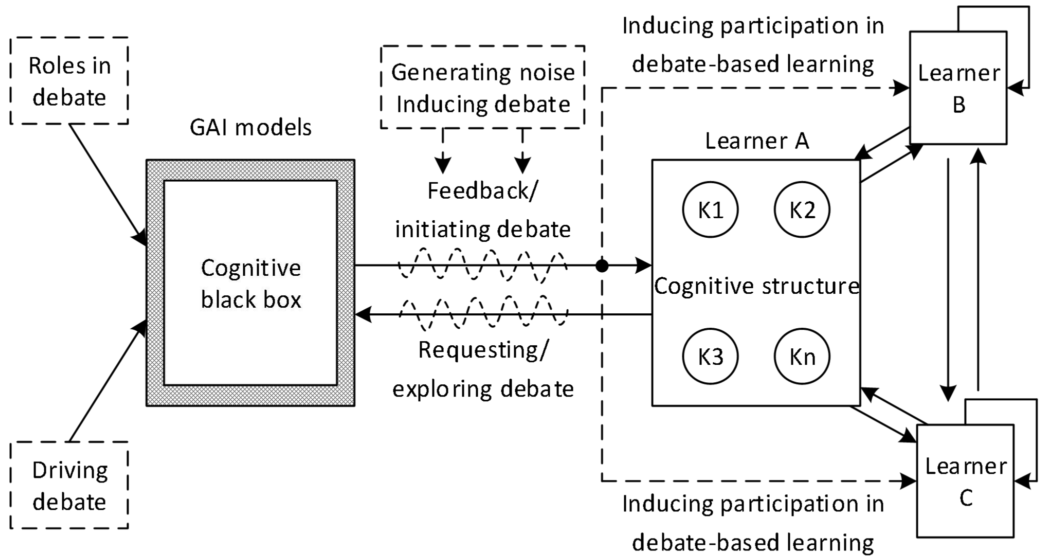


***Figure 1.*** Model e of the occurrence mechanism of GAI argument learning

(1) Mechanism for Generating "Debate Noise" to Induce Debate

The GAI model consists of a cognitive black box and a programming interface. It can generate human-like, systematic, and coherent feedback based on information prompts, simplifying the previously complex process of understanding, analyzing, and filtering search engine feedback. However, this not only leads many students to neglect the analysis, evaluation, and critical assessment of feedback but also fosters a learning attitude that discourages them from engaging in critical thinking, creativity, or problem-solving. The primary solution to this problem is to generate "debate noise" using large language models, learning analytics, and AI technologies during the human-computer question-and-answer process between requests and feedback. "Debate noise" disrupts the conventional human-computer dialogue by embedding questions or prompts that encourage mutual critical engagement. Based on Toulmin's argumentation model, "debate noise" is intelligently embedded into human-computer interactions to induce learners to engage in critical dialogue with the GAI model.

(2) Role-Based Mechanism to Drive Debate

GAI models have two shortcomings regarding debate-based learning. First, they lack the ability to actively guide students in debate-based learning. Simulating a classroom teacher's role in initiating debate-based learning is key to promoting students' higher-order thinking skills (HOTS). Learning systems can use learning analytics and AI technologies to monitor learning status in real time, thereby driving the GAI to actively initiate group debate-based learning. Second, they are not configured to play a role in debate-based learning. The GAI needs to function in the role of a "debate-oriented teacher." Assigning this debate role to the GAI is key to embedding feedback within "debate noise." "Debate noise" is likely to draw students' attention to the feedback content, prevent them from becoming mere "knowledge conveyors," and promote the development of their higher-order thinking skills (HOTS).

(3) Mechanisms for Inducing Group Participation in Debates

HOTS development requires the support of a learning community; however, the original GAI model only supports one-on-one question-and-answer activities. This necessitates the creation of a group debate learning space and mechanisms to induce participation. There are two main approaches to group debates: first, using role-playing to drive debates, thereby enabling multiple individuals to engage in debate-based learning with the GAI model; second, using "debate noise" to induce group participation in debate-based learning. "Debate noise" involves incorporating "argumentative" noise into feedback to prompt students to question, analyze, critique, or refine the feedback. At the same time, the GAI must guide students in resolving this noise. This requires using the results of conversational content analysis to guide students in evaluating, critiquing, or even refuting the content of the conversation.

## 3. Research Design

### *3.1. Research Methods*

Meta-analysis is a statistical method used to integrate and reanalyze multiple independent, related experimental or quasi-experimental studies toward a specific objective, in order to verify the common effects or identify the causes of differences among these independent studies. This study used the standardized mean difference (SMD) as the effect size to characterize the extent to which GAI-supported instructional interventions influence higher-order thinking skills (HOTS). Based on Cohen's theory, effect sizes of 0.2, 0.5, and 0.8 correspond to small, moderate, and large effects, respectively. The data were processed using the meta-analysis software Comprehensive Meta Analysis 3.0 to calculate the effect sizes for this study.

### *3.2. Literature Search and Screening*

#### 3.2.1. Literature Search

This study retrieved literature from databases such as Google Scholar, Web of Science, Springer, ScienceDirect, and Elsevier ScienceDirect. The types of literature selected for retrieval included journal articles and theses. Based on a preliminary review and analysis of the literature, it was found that research on higher-order thinking experienced a surge around 2015; to ensure the timeliness of the study, the search period was set from 2015 to 2025. Based on Bloom's Taxonomy of Educational Objectives and the concept of HOTS, the keywords for HOTS were selected as higher-order thinking, critical thinking, innovative thinking, creative thinking, logical thinking, metacognition, problem-solving skills, decision-making skills, and higher-order abilities. For GAI, the search keywords included "generative artificial intelligence," "AI-generated content," "AIGC," "large language models," "educational generative AI," and "chatbots." The search was supplemented by citation mining to ensure the comprehensiveness of the literature.

#### 3.2.2. Literature Screening

To avoid subjectivity, studies were included based on the following criteria: First, the research topic was higher-order thinking, and higher-order thinking was included among the dependent variables; second, the research methodology was empirical, experimental, or quasi-experimental, and included a control group and an experimental group, as well as data such as post-test means and standard deviations; third, the research participants were college students, including vocational college students, undergraduates, and graduate students; fourth, for studies with duplicate publications, only one publication was selected. Ultimately, 62 studies were included, comprising 58 high-quality journal articles and 4 theses. The flowchart of the literature screening process is shown in Figure 2.

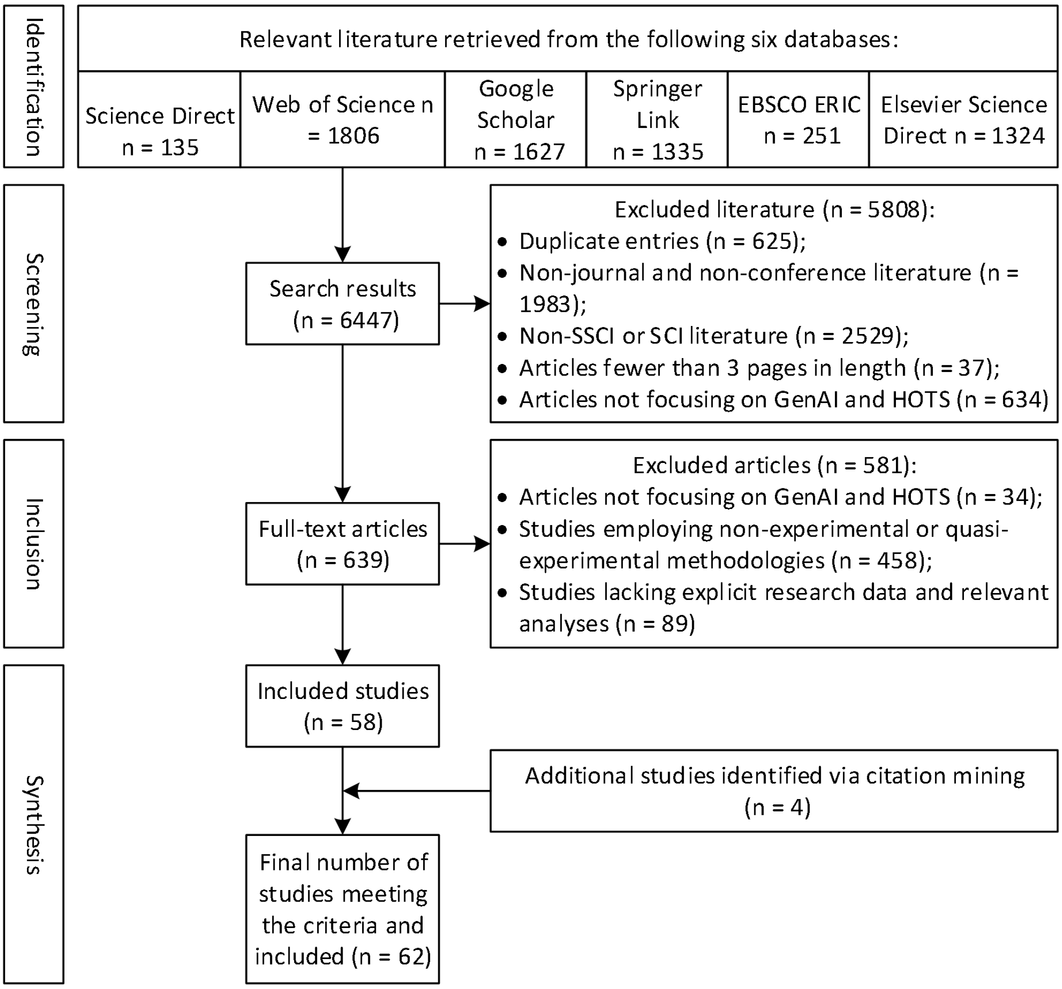


***Figure 2.*** PRISMA Flowchart

## 4. Research Findings

### *4.1. Types of GAI That Promote the Development of Learners' Higher-Order Thinking Skills (HOTS)*

Before addressing the question of which higher-order thinking skills (HOTS) GenAI can help learners develop, it is necessary to clarify the contexts in which existing empirical studies have focused on developing HOTS, which GenAI platforms were used, and what features they offer. Only then can we explore which HOTS GenAI promotes in learners across different contexts.

#### 4.1.1. Training Scenarios

An analysis of 62 studies revealed that there are six main types of scenarios in which GAI promotes the development of learners' higher-order thinking skills (HOTS), as shown in Figure 3. Among these, writing scenarios were the most common, with 19 studies (30.65%); professional course instruction scenarios accounted for 13 studies (20.97%). Next were programming scenarios, with 11 studies, accounting for 17.74%; oral conversation scenarios, with 8 studies, accounting for 12.90%; and reading scenarios, with 6 studies, accounting for 9.68%. Finally, artistic creation scenarios (including image creation, fashion design, etc.) accounted for 5 studies, representing 8.06%.

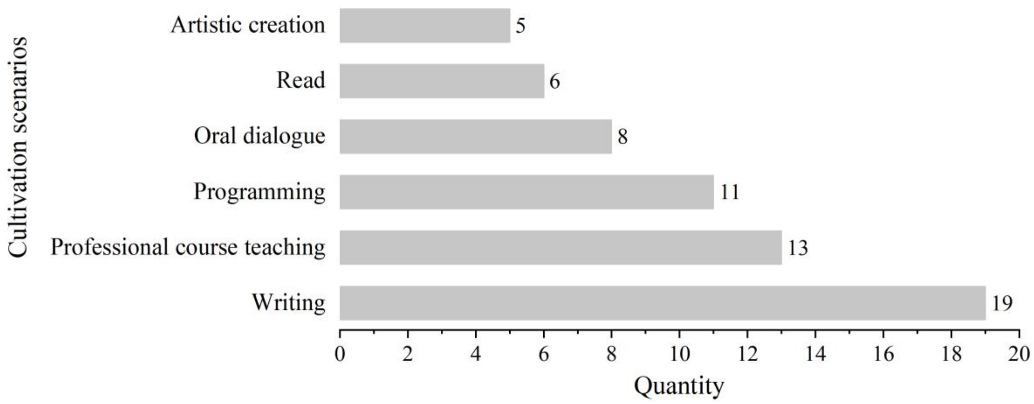


***Figure 3.*** Statistical chart of the cultivation scene

### 4.1.2. GAI Platform and Features

Statistical analysis revealed that the 68 empirical studies involved three categories of GenAI platforms. The first category consists of studies conducted using existing large-model platforms, such as ChatGPT, Wenxin Yiyan, and DeepSeek. These platforms primarily offer question-answering and guidance (n=22), essay generation (n=5), advice provision (n=3), content revision (n=2), and question generation (n=1) functions. The second category consists of independently developed large-model platforms. Compared to existing large-model platforms, these platforms have more focused functionality and do not place high demands on learners' ability to craft prompts. These tools or platforms primarily offer question-and-answer guidance (n=12), essay writing (n=11), evaluation (n=3), and resource recommendations (n=2). The third category consists of independently developed large-model-based educational agent platforms, such as the multi-agent teaching experiment platform developed using the AutoGen framework; these tools or platforms primarily provide functions such as question generation (n=1).

### 4.1.3. Learner HOTS Types

The results of the statistical analysis of learners' HOTS, as shown in Figure 4, indicate that existing empirical research focuses most on learners' innovative thinking skills (n=15, 24.19%), critical thinking skills (n=12, 19.35%), and problem-solving skills (n=9, 14.52%); some scholars refer to these three categories as the core components of HOTS. Second, there is a significant amount of research on self-regulation, human-computer collaboration, and metacognitive abilities, with seven studies each. Finally, there are five studies on computational thinking.

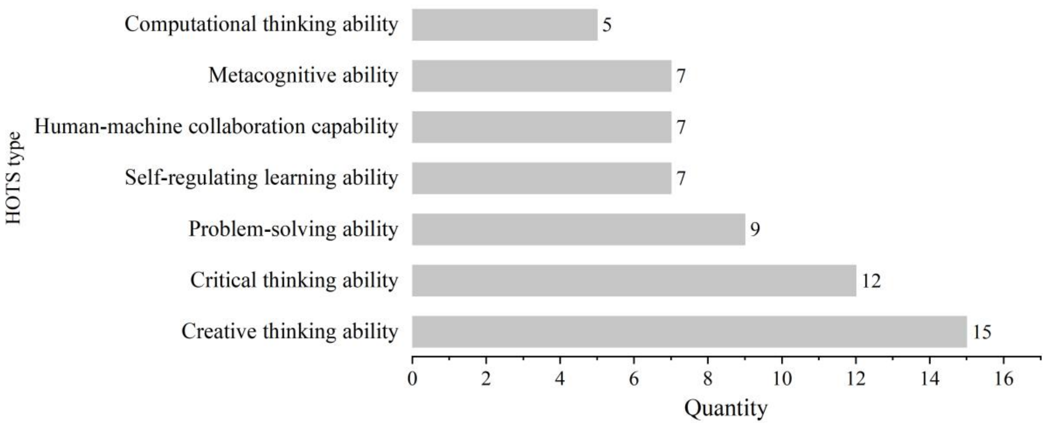


***Figure 4.*** Statistical Chart of higher-order thinking ability

Simply counting the frequency of learners' use of HOTS makes it difficult to observe the research focus in different learning contexts. Therefore, this study established a correspondence between development contexts and HOTS to analyze the contexts in which different HOTS are applicable and which HOTS are suitable for development in each context. The correspondence matrix between development contexts and HOTS is shown in Figure 5. Looking at the research context axis, the writing context involves the development of six categories of HOTS; in this context, researchers focus more on students' creative thinking abilities as well as their critical thinking abilities. Specialized course instruction also involves the development of six types of HOTS, with five of these studies focusing on the development of creative thinking skills. Programming involves the development of five types of HOTS, with a primary focus on learners' problem-solving and critical thinking skills. In the context of artistic creation, the types of HOTS developed are fewer—only three—with the primary focus on cultivating learners' critical thinking skills.

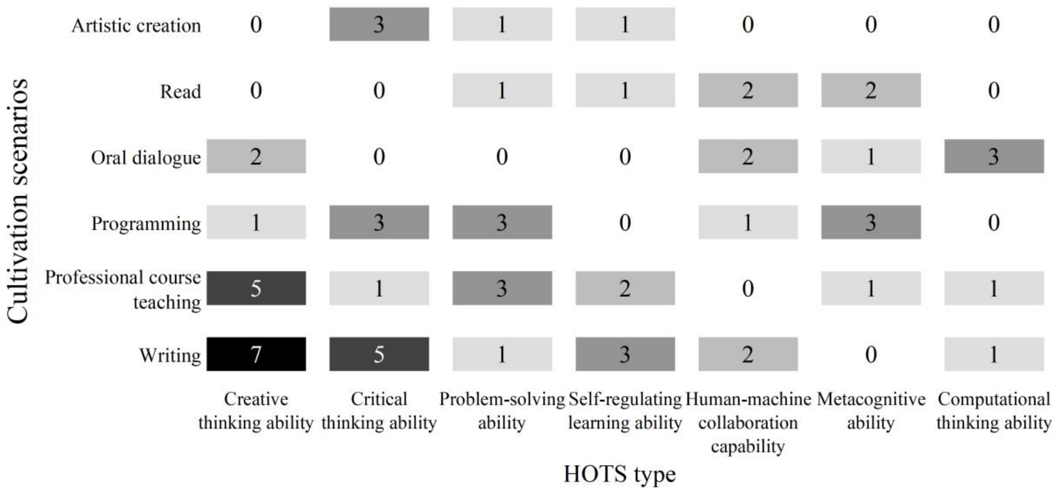


***Figure 5.*** The culture scene corresponds to the HOTS with bubbles

## *4.2. Meta-analysis Results*

### 4.2.1. Publication Bias Test

To avoid overestimating the effect size, a bias test must be conducted before performing the meta-analysis. First, a funnel plot was used to make a qualitative assessment of the presence of bias, as shown in Figure 6. Most effect sizes are symmetrically distributed on both sides of the effect size, suggesting that publication bias is unlikely. The Egger test yielded an intercept of 5.57 ( $p < 0.01$), indicating that publication bias may be present.

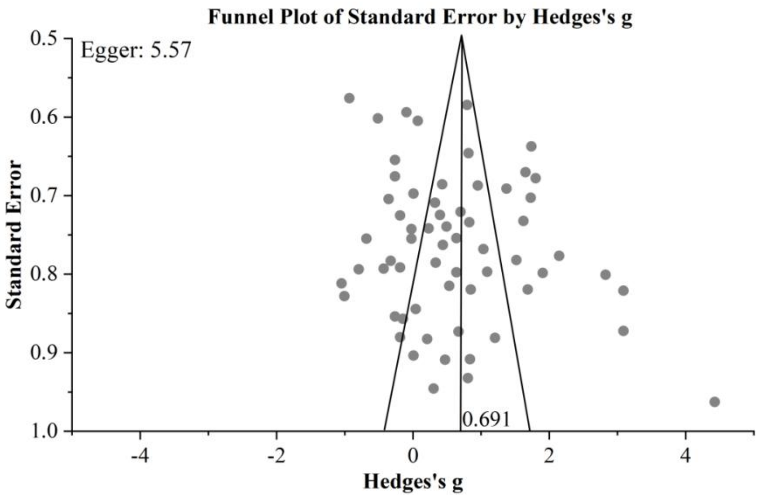


***Figure 6.*** Funnel plot of publication bias test

To maximize the robustness of the analysis results, the trimming method will be further applied in subsequent analyses to trim studies on both sides of the effect size, ensuring that the studies are distributed as symmetrically as possible on either side of the average effect size. It was found that the adjusted overall effect size remained significantly positive, with a 95% confidence interval of [0.587, 1.552]. The confidence interval for the effect size still does not include 0, indicating that publication bias had no significant impact on the results. Secondly, the loss-of-safety coefficient method was used to verify whether the effect was a spurious phenomenon; $N$ = 1,834, which is far greater than 5 $k$ +10 ( $k$ =62, where $k$ represents the total number of effect sizes included in this article), this indicates that unpublished effect sizes had a minimal impact on the results of this meta-analysis. Finally, a sensitivity analysis was conducted using the stepwise exclusion method, and the central trend remained stable at around 0.691, indicating that the meta-analysis results are robust. Based on the above analysis, there is no significant publication bias in the studies included in the meta-analysis, and the next phase of analysis can proceed.

### 4.2.2. Heterogeneity Test

Studies included in the same meta-analysis must undergo standardization tests to determine the level of heterogeneity; the $Q$ -test and $I^2$ statistic are typically used to assess the degree of heterogeneity among the included studies. The overall effect of GAI on HOTS is shown in Table 1. The test results indicate that $Q$ =705.782 ( $p$ =0.000 $< 0.001$) and $I^2$ = 93.827% > 75%, suggesting a high degree of heterogeneity among the studies. A random-effects model was selected for effect size pooling to fully account for the variability across different studies. Subsequent moderation analyses were conducted to explore variables that may influence the effect size, with the aim of providing a reasonable explanation for the sources of heterogeneity and enhancing the scientific rigor of the study's conclusions.

***Table 1.*** The Overall Impact of GAI on HOTS

| **Model** | | **Random effects model** |
|---|---|---|
| Sample size | | 64 |
| Effect size | | 0.831 |
| 95% confidence interval | Lower limit | 0.535 |
| | Upper limit | 1.156 |
| Double-tail test | Z value | 5.883 |
| | p value | 0.000 |
| Heterogeneity test | $I^2$ | 93.827 |
| | $Q$ | 705.782 |
| | df | 61 |
| | $p$ | 0.000 |

### 4.2.3. Analysis of Overall Effects

The "overall effect of GAI on HOTS development" was analyzed using a random-effects model. The results show that the pooled effect size, Hedges' g, is 0.831 ( $p$ =0.000 < 0.001), which is statistically significant. According to Cohen's effect size criteria, an effect size below 0.2 indicates a small effect; an effect size between 0.2 and 0.5 indicates a moderate effect; an effect size between 0.5 and 0.8 indicates an upper-moderate effect; and an effect size above 0.8 indicates a high effect. The current effect size of 0.831 indicates that the GAI has a very strong promotional effect on the development of students' HOTS.

Based on the analysis of overall effects, we further examined the influence of the GAI on different forms of HOTS, as shown in Table 2. Specifically, the effect sizes for metacognitive ability (Hedges' g = 1.252, $p < 0.05$) and critical thinking ability (Hedges' g = 0.934, $p < 0.005$) were greater than 0.8, indicating that the GAI has a highly positive impact on both students' metacognitive and critical thinking abilities. Although the effect on metacognition was greater than that on critical thinking, the effects on innovative thinking (Hedges' g = 0.352, $p > 0.05$), problem-solving ability (Hedges' g = 1.212, $p > 0.05$), computational thinking (Hedges' g = 0.838, $p > 0.05$), human-computer collaboration (Hedges' g = 0.981, $p > 0.05$), and self-regulation (Hedges' g = 0.633, $p > 0.05$). In terms of between-group effects, Q = 3.834, p = 0.621 > 0.05, indicating that there are no significant differences in the impact of the GAI on the various manifestations of students' higher-order thinking skills (HOTS).

**Table 2.** The Influence of GAI on Different Manifestations of HOTS

| Different Manifestations of HOTS | N | Effect size | 95% confidence interval | | Double-tail test | | Inter-group effect |
|---|---|---|---|---|---|---|---|
| | | | Lower limit | Upper limit | Z value | p value | |
| Creative thinking ability | 15 | 0.352 | -0.281 | 0.934 | 1.081 | 0.283 | *Q*=3.834 *p*=0.621 |
| Critical thinking ability | 12 | 0.934 | 0.345 | 1.451 | 3.052 | 0.004 | |
| Problem-solving ability | 9 | 1.212 | -0.435 | 2.829 | 1.520 | 0.145 | |
| Self-regulating learning ability | 7 | 0.633 | -0.382 | 1.533 | 1.321 | 0.238 | |
| Human-machine collaboration capability | 7 | 0.981 | -1.187 | 3.147 | 0.890 | 0.381 | |
| Metacognitive ability | 7 | 1.252 | 0.337 | 1.829 | 2.982 | 0.034 | |
| Computational thinking ability | 5 | 0.838 | 0.432 | 0.998 | 1.034 | 0.132 | |

#### 4.2.4. Analysis of Moderation Effects

This study further examined, through subgroup analysis, the effects of moderating variables—namely, discipline, learning style, experimental cycle, and GAI type—on the development of students' HOTS. The results of the subgroup analysis of moderating effects are shown in Table 3.

(1) The Impact of Discipline on the Development of Students' HOTS

By analyzing subgroup data from 25 humanities disciplines, 18 science and engineering disciplines, and 10 other disciplines, this study found that GAI had a significant positive impact on the development of students' HOTS in both humanities and science and engineering disciplines, while the impact on other disciplines was not significant. The results of the between-groups effect test ($Q = 13.84$; $p = 0.000$) indicate that the impact of different disciplines on students' HOTS development differs significantly.

(2) The Impact of Learning Modes on Students' HOTS Development

Analysis of data from 25 individual learning subgroups and 37 group learning subgroups revealed that GAI had a more significant promotional effect on HOTS development in individual learning (effect size = 0.951, $p < 0.001$) and a moderate promotional effect in group learning (effect size = 0.534, $p < 0.001$). However, a between-group effects test ($Q = 2.034$, $p = 0.105 > 0.05$) indicated that there was no statistically significant difference in the promotional effects of the two learning modes on HOTS development.

**Table 3.** Subgroup analysis results of moderating effects

| Adjusting variable | Category | N | Effect size | 95% confidence interval | | Double-tail test | | Inter-group effect |
|---|---|---|---|---|---|---|---|---|
| | | | | Lower limit | Upper limit | Z value | p value | |
| Subject | Humanities | 25 | 0.705 | 0.287 | 1.134 | 3.186 | 0.000 | *Q*= 13.84; *p*= 0.000 |
| | Science and engineering | 18 | 0.887 | 0.834 | 3.152 | 3.154 | 0.003 | |
| | Others | 10 | -0.081 | -0.380 | 0.337 | -0.038 | 0.835 | |
| Learning methods | Individual learning | 25 | 0.951 | 0.408 | 1.437 | 3.662 | 0.000 | *Q*= 2.034; *p*= 0.105 |
| | Group learning | 37 | 0.534 | 0.306 | 0.731 | 5.483 | 0.000 | |
| Cycle | Short-term | 22 | 1.135 | 0.253 | 2.305 | 2.005 | 0.018 | *Q*= 1.434; *p*= 0.485 |
| | Mid-term | 31 | 0.569 | 0.150 | 0.963 | 2.347 | 0.006 | |
| | Long-term | 9 | 0.752 | 0.289 | 1.189 | 3.062 | 0.003 | |
| GAI type | Domestic | 34 | 0.663 | 0.292 | 1.452 | 3.458 | 0.001 | *Q*= 0.434; *p*= 0.505 |
| | Abroad | 28 | 0.891 | 0.456 | 1.395 | 4.56 | 0.000 | |

(3) The Impact of Intervention Duration on the Development of Students' HOTS

The results of the between-group effect test ($Q = 1.434$, $p > 0.05$) indicate that there is no significant difference in the impact of different intervention durations on the development of HOTS. In terms of effect sizes for each intervention duration, the short-term intervention (effect size = 1.135, $p = 0.018 < 0.05$) > long-term intervention (effect size = 0.752, $p = 0.003 < 0.01$) > medium-term intervention (effect size = 0.569, $p = 0.006 < 0.01$). The short-term intervention had a more significant effect on promoting HOTS and achieved a large effect size, while the long-term and medium-term interventions achieved moderate effect sizes.

(4) The Impact of GAI Type on the Development of Students' HOTS

Analysis of data from 34 domestic and 28 international GAI subgroups revealed that international GAIs (effect size = 0.891, $p = 0.000 < 0.001$) had a greater promotional effect on students' HOTS than domestic GAIs (effect size = 0.663, $p = 0.001 < 0.005$). However, the between-group effect test ($Q = 0.434$, $p = 0.505 > 0.05$) indicated that there was no significant difference in the effect of promoting students' HOTS among GAIs of the same type.

## 5. Conclusion

Research Findings:

(1) GAI has a positive impact on the development of students' higher-order thinking skills (HOTS).

Overall, GAI plays a significant role in promoting the development of students' higher-order thinking skills. GAI effectively fosters students' metacognition and critical thinking. Regarding metacognition, the findings are consistent with previous research on the use of GAI to enhance metacognitive awareness. This may be because GAI continuously records the learning process and provides students with behavioral data feedback, thereby enabling them to better plan, monitor, and adjust their learning strategies—including goal setting, tracking cognitive processes, real-time self-testing, and self-assessment of learning outcomes. Regarding critical thinking, the findings align with previous research, all of which indicate that students' use of GAI enhances their critical thinking skills. This may be because GAI provides multi-perspective viewpoints and evidence, broadens the dimensions of students' thinking, and trains their argumentation skills through debate-style dialogue, thereby offering unique support for the development of students' critical thinking.

(2) The influence of moderating variables on GAI's cultivation of students' HOTS

At the disciplinary level, there are significant differences in GAI's impact on students' HOTS across different disciplines, indicating that disciplinary background plays a key moderating role in GAI-supported HOTS development. Specifically, the effect size is extremely high in science and engineering disciplines, while the impact is moderate in the humanities. This may be because GAI can directly assist students in science and engineering disciplines with problem decomposition, algorithm optimization, and experimental simulation, thereby fostering a "analysis–transfer–synthesis–innovation–monitoring" thinking model and promoting the spiral-like development of students' higher-order thinking. Although GAI can assist with text generation and argument structuring, deep critical thinking in the humanities still requires human leadership, with GAI serving more as a supplement than a core driving force. All disciplines generally face the challenge of weak critical thinking training: the humanities tend to blindly revere authority; STEM disciplines are prone to a "standard answer" mindset; and other disciplines are easily constrained by empirical judgments. During the use of GAI, students' rational scrutiny of the answers generated by the AI serves as the key lever for addressing these disciplinary challenges and activating critical thinking.

At the experimental cycle level, GAI did not show significant differences in its impact on students' higher-order thinking skills (HOTS) across different experimental cycles. Although no significant differences were found, the 0–1-month experimental cycle exhibited the largest combined effect size, indicating that students' HOTS improvements were most pronounced during short-term interventions supported by GAI. Experimental cycles of 3–6 months and 6 months or longer did not have a significant moderating effect on any higher-order thinking skills. This is likely because the convenience of GAI exacerbated students' cognitive laziness over the medium to long term, thereby offsetting the tool's potential gains in information processing efficiency. Medium- to long-term experimental cycles may produce a ceiling effect, where improvements in HOTS reach a limit, making it difficult for subsequent interventions to generate further significant impacts.

Research Implications:

(1) Technology-Led Approach: Following Cognitive Principles to Empower Collaborative Learning

In the future, we should enhance the educational applicability of GAI to promote students' development of Higher-Order Thinking Skills (HOTS). First, we must precisely align with the cognitive development characteristics of students at different educational stages and develop stage-specific educational large language models. Individual cognitive development comprises three stages: concrete-image thinking, image-abstract thinking, and abstract-logical thinking. Large models for higher education should establish an AI project mentor system to strengthen the analysis and exploration of complex, abstract problems, serving as advanced cognitive partners for students. Second, we should build a collaborative learning platform centered on GAI. GAI can intelligently assign learning tasks based on students' past performance, construct learning frameworks, summarize and distill key points, and generate personalized evaluation feedback based on student participation, ultimately forming a teaching feedback loop characterized by "technology empowerment—full participation—competency enhancement.”

(2) Faculty Leadership: Daring to Innovate and Adapting Instruction to the Times and Students’ Needs

The unique role of faculty in education is irreplaceable. Faculty should proactively adapt to technological changes in the AI era and guide students in exploring the application of GAI in teaching and learning. Teachers should prioritize the development of students’ higher-order thinking skills as their core objective, construct a “teacher-student-computer” tripartite interactive teaching model, and drive the transformation of the assessment system from a knowledge-based to a competency-based approach. They should establish an assessment framework focused on the development of comprehensive competencies such as critical thinking, reflective thinking, and problem-solving skills. Furthermore, teachers need to scientifically design GAI teaching strategies tailored to students’ educational stages.

(3) Student-Centered Approach: Beware of the Cult of Digital Technology; Make Effective Use of Tools to Empower Students

As the primary agents of their own learning, students’ development of higher-order thinking skills (HOTS) requires not only technological support and teacher guidance but also their own active participation. First, students need to cultivate a sound perspective on the relationship between humans and machines. They must not only be adept at leveraging “machine intelligence” to enhance “human intelligence”—using GAI to expand the breadth and depth of their cognition—but also uphold their own value-based rationality to prevent “machine intelligence” from undermining “human intelligence” due to overreliance or blind worship. Second, students must engage in metacognitive monitoring during human-machine interactions, maintain autonomy in their thinking, and carefully evaluate and critically adopt generated content. Third, beyond classroom learning tasks, students can explore the application of GAI in self-directed learning scenarios to enhance their own HOTS through practice.

**References**


1. Widana, I. W. (2017). Higher order thinking skills assessment (HOTS). JISAE: Journal of Indonesian Student Assessment and Evaluation, 3(1), 32-44.
2. Setiawan, A., Malik, A., Suhandi, A., & Permanasari, A. (2018, February). Effect of higher order thinking laboratory on the improvement of critical and creative thinking skills. In IOP Conference Series: Materials Science and Engineering (Vol. 306, No. 1, p. 012008). IOP Publishing.
3. Lang, Q., Wang, M., Yin, M., Liang, S., & Song, W. (2025). Transforming education with generative AI (GAI): Key insights and future prospects. IEEE Transactions on Learning Technologies, 18, 230-242.
4. Jihannita, J., Prasetyo, Z. K., & Wilujeng, I. (2023). How to prepare HOTS to face the 21st century?. Jurnal Penelitian Pendidikan IPA, 9(8), 486-492.
5. Sagala, P. N., & Andriani, A. (2019, March). Development of higher-order thinking skills (HOTS) questions of probability theory subject based on Bloom’s taxonomy. In Journal of Physics: Conference Series (Vol. 1188, No. 1, p. 012025). IOP Publishing.
6. Alkiyumi, M. (2024). New classification of higher-order thinking skills (HOTS). Onomázein, 62(December 2023), 1053-1064.
7. Kurniawan, E. S., Mundilarto, E. I., & Istiyono, E. (2024). Improving student higher order thinking skills using Synectic-HOTS-oriented learning model. Int J Eval & Res Educ ISSN, 2252(8822), 1133.
8. Nie, X., Tian, Y., Liu, M., Wu, D., & Guo, Y. (2025). The impact of generative artificial intelligence on students’ higher order thinking: Evidence from a three-level meta-analysis. Education and Information Technologies, 30(17), 25359-25390.
9. Deng, Y., Guo, H., & Dai, Y. (2026). Examining the impact of higher-order thinking and GAI chatbots on engineering creativity in higher education. Scientific Reports.
10. Zhang, Y., Tian, H., & Lu, J. (2025). The relationship between higher-order thinking and problem-solving skills development among pre-service teachers using generative AI: an analysis based on moderated mediation. BMC psychology, 13(1), 1094.

11. Zhao, Y., Yue, Y., Sun, Z., Jiang, Q., & Li, G. (2025). Does Generative Artificial Intelligence Improve Students' Higher-Order Thinking? A Meta-Analysis Based on 29 Experiments and Quasi-Experiments. Journal of Intelligence, 13(12), 160.
12. Liu, Y. L. E., Lee, T. P., & Huang, Y. M. (2026). Enhancing student engagement and higher-order thinking in human-centred design projects: The impact of generative AI-enhanced collaborative whiteboards. Interactive Learning Environments, 34(1), 258-275.
13. Martinez, C. M., Roger-Monzo, V., & Sirvent, F. C. (2025). Generative AI and critical thinking in online higher education: challenges and opportunities. RIED-Revista Iberoamericana de Educación a Distancia, 28(2), 233-273.
14. Raitskaya, L., & Tikhonova, E. (2025). Enhancing critical thinking skills in ChatGPT-human interaction: A scoping review. Journal of Language and Education, 11(2 (42)), 5-19.