

Critical AI Literacy in Foreign Language Writing Instruction in the Age of Generative AI: Theoretical Framework and Pedagogical Pathways

Mu He*

School of Foreign Languages and Literatures, Chongqing Normal University, Chongqing, China

**Corresponding Author*

Abstract: This study conceptualizes critical AI literacy to address the subjective risk of the erosion of authorial voice, cultural stance and judgmental accountability in AI-assisted writing instruction. Drawing on critical language awareness and critical information literacy, a theoretical framework grounded in the Awareness-Positionality-Strategies-Evaluation (APSE) model is established. Unlike general or generative AI literacy, critical AI literacy shifts the focus from effective tool use to the maintenance of authorial judgment in human-AI collaboration. The framework analyzes the subjective risk, embeds critical literacy into the three-stage instructional design, and establishes a three-tiered evaluation structure. It repositions APSE as a pedagogical instrument and enables the paradigm shift from technological competence to authorial agency. This study provides both theoretical basis and practical pathways for AI-mediated writing instruction.

Keywords: Critical AI Literacy; Foreign Language Writing Instruction; APSE Framework; Subjectivity; Human-AI Collaboration; Generative AI

1. Introduction

Generative artificial intelligence (AI) is rapidly penetrating into the field of foreign language teaching, especially writing instruction. It fundamentally rewrites the practical form of traditional writing support through instant formative feedback, the extension of discourse structure, the optimization mechanism of language expression and sustainable dialogue support. Compared with dictionaries, corpora or automatic scoring systems that only serve specific aspects of language processing, generative AI shows significant heterogeneity.

It is no longer a passive query object but a generation subject, and directly embeds and reshapes the writing ecology from concept to evaluation [1,2].

In the collaborative situation with AI as the main generator, how can learners guard their cognitive sovereignty in the writing process and avoid the weakening of critical thinking due to technology dependence? In addition, as language expression, cultural stances and value judgment are increasingly mediated by algorithms, foreign language education must face these basic problems related to the true nature of education.

Faced with the educational embeddedness of AI, the academic community uses "AI literacy" as a conceptual tool to outline the technological cognition, application ability and ethical awareness that learners should have [3,4]. This framework effectively promotes the acceptance and internalization of technology in foreign language education. However, when AI is transformed from a passive auxiliary tool to an active cognitive partner, the boundary limitations of existing AI literacy are increasingly prominent. It emphasizes more on "how to coexist with technology", rather than "how to guard the irreplaceable nature of people in technology". Research shows that under the surface of generative AI improving writing fluency, there are risks such as the erosion of authorial voice, the loss of intellectual depth, the outsourcing of evaluative responsibility and the intensification of technological dependency [5,6]. This shows that the core focus of foreign language education has changed from "how to make good use of tools" to "how to establish people's cognitive sovereignty in man-machine cooperation". Xu's action research further confirmed that the generation of thought depth cannot be realized automatically through technical proficiency, but depends on the

reflective turn of teaching intervention [2]. As pointed out by Sperling et al, how can foreign language education realize the transformation from technical rationality to humanistic rationality in the context of accelerating technological iteration [3]?

From the perspective of writing instruction, the limitations of the explanatory power of general AI literacy become increasingly apparent. The "technical competence" it cultivates may cover up the profound subjective risk brought by AI participating in writing. The first risk is the outsourcing of cognitive process. AI quickly generates the text prototype, which makes it easy for learners to transfer the internal thinking process of idea and modification to the machine, and the voice of "human" in writing practice gradually fades. The second risk is the distortion of self-cognition. The immediate acquisition of normative texts may lead to learners' misjudgment of writing ability, forming a vicious circle of technology dependence and false self-efficacy. The third is the concealment of the value position. AI output, with its "natural" and "reasonable" appearance, obscures the hidden English centered discourse, cultural flattening mechanism and ideological prejudice, making learners unconsciously accept the cultural discipline of the algorithm [7,8]. It can be seen that a literacy framework that cannot guide learners to ask "who is speaking", "who is deciding the meaning" and "who is taking responsibility" is still not enough, even if it is more complete at the technical application level.

Therefore, the limitations of general AI literacy in writing instruction are increasingly evident. At present, foreign language education faces a fundamental choice that whether to continue to stay at the technical optimization level of "how to use AI more effectively", or to move towards a "critical AI literacy" that can protect learners' subjectivity. The "criticality" emphasized by the latter is not a simple negation of technology, but refers to learners' ability of reflective intervention, dialogic negotiation and independent evaluation in the process of collaborating with AI [9]. The necessity of this literacy turn comes not only from the inevitable logic of theoretical deduction, but also from the urgent call of practical difficulties. The quantitative survey conducted by Chen et al shows that learners'

cognitive dimension of feedback is generally weak. The qualitative observation of front-line teachers is the evidence of reverse replication. Students often stop content replication in ChatGPT-assisted writing, lacking critical inspection of the generated text. The dual pressure of theory and practice requires the foreign language education community to provide a systematic plan to transform the critical AI literacy from concept to teaching practice.

This study aims to provide a systematic theoretical framework and teaching plan for cultivating critical AI literacy in foreign language writing instruction. It is positioned at the intersection of theoretical construction and instructional design, with "critical AI literacy" as the core concept, drawing on and transforming the APSE model to construct an analytical framework suitable for foreign language writing instruction [10]. On this basis, it provides a teaching implementation path. At the methodological level, this study employs a combination of literature review and concept analysis methods, following the research procedure of "concept definition-framework construction-risk analysis-path design".

Compared with existing body of research, the innovation of this study lies in the following three aspects. The first aspect is the integration and innovation of perspectives, introducing critical language awareness and critical information literacy into the research of AI literacy, endowing the framework with a unique critical dimension of foreign language education language, and enabling AI literacy to move from general abilities to disciplinary construction. The second is to expand the effectiveness of the tool, extending the APSE model from a static capability analysis framework to a dynamic instructional design tool, enabling it to systematically integrate practical aspects such as risk identification, teaching intervention, and effectiveness evaluation, achieving a functional transition from "description" to "intervention". The third is the return of ontological attention, with the research core proposition shifting from "how to use AI more effectively" to "how to maintain author judgment in human-computer cooperation", elevating the research focus from "technical effectiveness list" to "subjective judgment mechanism", completing the paradigm reconstruction at the methodological

level. The above innovations all point towards a common goal, which is to transform critical AI literacy from an abstract concept into an operational and evaluative foreign language writing teaching practice plan.

2. The Connotations and Theoretical Framework of Critical AI Literacy

Critical AI literacy is not born in a vacuum. On the contrary, it represents a fundamental shift in the established theoretical tradition of foreign language education in the era of AI. Its theoretical resources mainly come from two intertwined academic contexts. The first is critical language awareness, which reveals the intrinsic connections between language and power, identity and ideology, providing a critical perspective for studying the language bias output by AI [11]. The second is critical information literacy, which elevates information literacy from the skill level to the value reflection level, emphasizing that information acquisition, understanding, and use must be embedded in the process of problem localization and knowledge construction, providing methodological resources for evaluating the credibility and legitimacy of AI generated content [12]. In the context of the deep involvement of AI in writing instruction, the technologization of language and the verbalization of technology have formed a dual cover. The "technical ability" cultivated by general AI literacy may only be overlooked by learners, and the fusion of the above two clues provides a theoretical weapon to penetrate this barrier. Critical language awareness helps learners identify 'who is speaking' and 'whose voice is amplified or silenced', while critical information literacy guides learners to ask 'how knowledge is generated' and 'where is naturalized'. Both of them point to the core focus of "authorship judgment", laying a theoretical foundation for this study to draw on and reform the post chamber model.

In order to clarify the theoretical positioning of critical AI literacy in foreign language writing instruction, it is necessary to systematically analyze the concepts of "AI literacy", "generative AI literacy", and "critical AI literacy". These three constitute an evolutionary lineage from "instrumental rationality" to "subjective consciousness". AI literacy represents the cognitive mastery of technology, with its core being the

understanding of the basic principles, application methods, and ethical boundaries of AI [3,4]. Generative AI literacy is a step forward, focusing on unique interactive skills for generating scenarios and the key role of prompt design as a human-machine interface in library environments. Agirdag points out from a critical perspective that prompt design not only involves technical operations, but also embeds complex relationships between language mechanisms and social inequality [13]. Critical AI literacy goes beyond this spectrum, as it is no longer satisfied with technical optimization of "how to use AI more effectively", but places learners at the core of human-machine collaboration and asks "can we maintain authorial judgment when collaborating with AI". This questioning fundamentally distinguishes between critical AI literacy and AIEL3 framework [4]. The latter involves adding ethical considerations within the framework of 'technological application capabilities', while the former is the reconstruction of problem awareness itself. This transformation is a specific cultural response to the previous article's proposition of the transition from technological rationality to humanistic rationality.

After clarifying the lineage relationship between critical artificial intelligence literacy and existing concepts, this study defines it as a comprehensive ability centered around "subjective judgment mechanisms". Specifically, critical AI literacy refers to the ability of learners to identify technological logic and its social and cultural consequences during the deep involvement of AI in language learning, maintain a cautious rather than dependent stance on AI output, negotiate with AI through conscious interaction strategies, and make independent judgments and value evaluations of the resulting outcomes. This definition can be decomposed into four interdependent core dimensions, forming a logical isomorphism with the APSE model. The first dimension is consciousness, which refers to the cognitive awareness of the principles, operational logic, and potential biases of artificial intelligence technology. The second dimension is stance, which refers to establishing ethical responsibility and independent value judgments in human-computer interaction, and maintaining learners' "cognitive sovereignty". The third

dimension is strategy, which refers to maintaining discourse dominance in the writing process through strategic interventions such as rapid design, questioning and iteration, and comparing multiple options. The fourth dimension is evaluation, which refers to checking the authenticity, logical review, cultural suitability judgment, and "author identity" of the content generated by artificial intelligence.

Compared to general artificial intelligence literacy, this definition achieves three paradigm shifts. Firstly, there is a difference in goals, with "preserving subjectivity" rather than "proficiency in operation" as the basic direction for cultivating literacy. The second is the difference in objects, reconstructing the output of artificial intelligence from "natural and trustworthy sources of knowledge" to "dialogue objects that require critical questioning". The third is the difference in process, which transforms the integration of technology into teaching from "passive acceptance of output" to "continuous negotiation with algorithms". This definition elevates critical artificial intelligence literacy from a "list of abilities" to a "judgment mechanism," directly responding to the core focus of "author judgment" raised in the introduction.

Based on the previous definition of critical AI literacy, this study further draws on APSE model to construct an analytical framework suitable for the context of foreign language writing instruction [10]. The four dimensions of this framework are the specific implementation of the "triple paradigm shift" at the operational level. Critical awareness corresponds to "object differences", reconstructing the output of AI from "natural and reliable sources of knowledge" to "dialogue objects that require questioning traceability", requiring learners to have insight into the implicit operations, algorithmic mechanisms, language biases, and cultural presuppositions of training data. The critical stance carries the "goal difference" - taking "subjectivity preservation" rather than "operational proficiency" as the basic direction of literacy cultivation, enabling learners to position themselves as "interlocutors" rather than "receivers" in human-computer interaction, and intervene in writing as guardians of "cognitive sovereignty". The critical

interaction strategy reflects the "process difference", and the integration of technology into teaching has shifted from "passive acceptance of output" to "continuous negotiation with algorithms". Through strategic interventions such as real-time design, questioning and iteration, constraint setting, multi scheme comparison, and meaning reconstruction, the author maintained their dominant position in the interaction. Critical evaluation integrates the ultimate demands of the triple paradigm shift, systematically assessing the authenticity verification, logical review, cultural appropriateness analysis, and questioning of "author identity" in the resulting and forming processes, and implementing the "subjective judgment mechanism" as an actionable ability behavior. These four dimensions are interdependent and progressive. Consciousness provides a cognitive prerequisite for learners to recognize the existence of default values in algorithms, which ensures the subject's posture and prevents learners from becoming passive recipients of technology. Strategies provide operational means to transform subjective intentions into effective interactive practices. Evaluate and implement the final judgment, and review the value of the resulting outcomes and processes. Around the core proposition of "maintaining writing subjectivity", the four have formed a complete ability loop, from "seeing" to "standing", and then to "intervening" in the final "decision-making", thus truly realizing the teaching practice framework of critical AI literacy from abstract concepts to analytical, trainable, and assessable.

It should be pointed out that critical AI literacy is not a natural development of general AI literacy in the technical capability spectrum, but a methodological reconstruction of the concept of literacy. This is a concentrated manifestation of the "triple paradigm shift" in the essence of literacy, and its core proposition has shifted from "how to use AI more effectively" to "how to maintain authorial judgment in human-machine collaboration". The key to critical AI literacy is not whether learners can generate normative texts faster, but whether they can answer four subjective questions in writing practice with deep technological intervention. Firstly, can they recognize the implicit operations of the

algorithm's default values (corresponding to "object differences")? Secondly, can they protect the unique voice of the author in technical assistance (corresponding to "target differences")? Thirdly and fourthly, can they make conscious cultural judgments on the generated content and ultimately take responsibility for the expression behavior (corresponding to the ultimate demand of "process differences"). For foreign language education, this reconstruction has disciplinary ontological significance. Language learning is essentially a continuous practice of expression selection, cultural status, and meaning negotiation, and generative AI has invested the most profound technological variables in these "core areas" that constitute the foundation of language education. It is precisely the mission of this discipline that makes foreign language education a key area for testing the proposition of the transition from technological rationality to humanistic rationality.

3. AI-Assisted Writing: Risk Analysis and APSE Application

Under the surface of technological empowerment, the reshaping of the writing process by generative AI hides a profound crisis of subjectivity. The APSE framework constructed in the previous section is an analytical tool that can attribute these crises to three interrelated risk dimensions, namely the negative manifestations of the old paradigm in the "triple paradigm shift" discussed in Section 2.

The first risk is a weakening of the author's voice and the absence of critical stance. In foreign language writing instruction, the author's voice is a concentrated reflection of the author's position on the topic, the active organization of the argumentation path, and the sense of responsibility for expressing the consequences [2]. However, the output of generative AI is based on statistical fitting of high-frequency corpora, which naturally tends to produce "surface coherent" paradigm answers. When learners mistake the "standard text" generated by AI for "their own expression," subjectivity will be unconsciously diluted [6,10]. This is a typical symptom of the absence of a 'critical stance', where learners fail to communicate as 'interlocutors' rather than 'receivers' in their writing, resulting in their raw and vivid personal voices being drowned

out by the general expression of algorithms. Taking the argumentative essay "The Influence of Social Media" as an example, if learners only input a summary instruction of "discussing the pros and cons of social media", although the opening language generated by AI is standardized, it may completely cover a unique perspective based on their own network experience.

The second risk is the blind spot of critical consciousness caused by the drift of cultural stance. Many seemingly 'objective' and 'neutral' AI outputs are actually embedded with specific cultural centrism and value assumptions. When students accept the argumentative logic provided by AI without careful consideration, their text may become standardized at the linguistic level, but tend to converge to the default cultural patterns of the algorithm at the positional level, due to the unconscious cultural implantation caused by a lack of "critical awareness". Taking the theme of "environmental protection" as an example, texts generated by AI tend to cite policy cases and discourse frameworks from Western countries, but pay little attention to local environmental protection practices and their cultural logic. If learners do not pay attention to this, their writing will unintentionally complete a shift in cultural status and transform from "meaning constructors" to "passive recipients of cultural paradigms" [2,14].

The third risk is the lack of critical assessment caused by outsourcing responsibility. When learners overly rely on the argument framework and conclusion premises provided by AI, the responsibility of value judgment and logical verification that should be borne by themselves quietly shifts to algorithms. Learners no longer ask whether this viewpoint is reasonable or whether the argument is reliable, but instead assume the correctness of the AI output, which is a typical manifestation of the failure of the "critical evaluation" dimension function. As a result, learners gradually lose their awareness and ability to make independent judgments, transforming from active constructors of textual meaning to passive screeners of generated results, and completely surrendering the right to question the "author" to technology.

The above three risks are interrelated and progressive. The algorithm blurs the individual

expression of learners, and the drift of learners' positions causes their cultural status to unconsciously slide towards technological assumptions, while the outsourcing of responsibility completely abandons their final judgment of the resulting outcomes. These three points point to a core proposition that, in the context of the deep involvement of AI in writing teaching, if critical literacy is not consciously cultivated, "technological empowerment" may quietly slide towards "subjectivity transfer". This is precisely the practical urgency of the aforementioned "transition from technological rationality to humanistic rationality" era proposition.

3.1 Critical Awareness and Critical Positionality: Identifying and Resisting Algorithmic

The common problem revealed by the previous analysis of triple risk is learners' lack of concentration and passive acceptance of "algorithm default values". The default value of the algorithm refers to the biased output paradigm formed by generative AI in text production based on large-scale corpus training and probabilistic generation mechanisms. It specifies 'what is reasonable expression' and 'what is appropriate argumentation'. But it often hides in a neutral, fluent, and natural textual form, causing learners to unconsciously internalize the cultural assumptions, value biases, and argumentative inertia of algorithms into their writing norms in the pursuit of expression efficiency [10]. To resist this hidden technological discipline, we must seek the synergy of critical awareness and critical status within the APSE framework. We discuss these two dimensions side by side because they together constitute the initial stage of the relationship between learners and AI. Consciousness undertakes the cognitive mission of "penetrating the cover", and this position completes the value choice of "subject status". Both are indispensable, laying the foundation for subsequent interaction strategies and evaluation abilities.

The core of critical awareness is to enable learners to 'see' the default values of algorithms. This kind of 'seeing' is far more than just identifying obvious errors in the output of AI, but a deeper insight. Even if the text appears flawless on the surface, AI may still impose potential constraints on learners' thinking and

expression through micro mechanisms, such as the use of positive or negative language, cultural attribution of examples, logical paths of argumentation, and value orientations of styles. Tacelosky et al. have shown that learners' ability to identify cultural biases in AI outputs is not a natural acquisition, but depends on specific reflective training [15]. Liu Fang and Dong Yu compare critical thinking to the "searchlight, perspective mirror, firewall, and safety net" of understanding and applying generative AI [16]. This metaphor accurately reveals the multiple functions of critical consciousness, illuminates the black box of algorithm operations, reveals the cultural code of text loading, constructs cognitive barriers to discipline, and provides the last line of defense for learners' expression security.

On the basis of "seeing", the critical position further answers the question of how learners should face the default values of the algorithm. It opposes the worship of AI as an absolute authority in technology, which makes learners passive recipients of algorithms. We also oppose the fear of technology, where AI is seen as a harmful tool and this attitude leads learners to miss out on the possibility of technological empowerment. The critical stance advocates maintaining a constructive connection between open use and sustained vigilance, where learners not only make full use of AI to improve expression efficiency, but also consistently position themselves as the ultimate bearers of writing meaning. As for writing instruction, this means that students cannot automatically equate AI output with "best expression" because it is more "authentic". On the contrary, they should maintain necessary skepticism and reprocessing awareness towards the output of AI, and confirm "who is speaking" and "who is responsible for expressing" in each human-computer interaction [10,14].

Consciousness and status are the "eyes of cognition" and the "anchors of the subject". The collaborative operation of the two enables learners to not only perceive the implicit operations of the algorithm's default values, but also firmly stand in a dominant position under the influence of technology. This is the first line of defense against algorithmic discipline and protecting subjectivity in writing. It is also a common prerequisite and value

foundation for the subsequent development of critical interaction strategies and critical evaluation abilities.

3.2 Critical Interactive Strategies and Critical Evaluation: From Prompt Generation to Textual Review

The previous analysis of the triple risk revealed three ways in which subjectivity is lost in AI assisted writing, namely the weakening of the author's voice, the drift of cultural status, and the outsourcing of judgment responsibility. The synergistic effect of key interaction strategies and key assessments is a systematic response to these three risks. The former resists the gradual erosion of weakened voices and status drift through process negotiations, while the latter hinders the ultimate completion of responsibility outsourcing through outcome judgments. These two dimensions cover the complete chain from the writing process to the writing product, expanding the APSE framework from the initial defense of "seeing" and "standing" to the practical depth of "control" and "judgment".

The core task of the key interaction strategy is to reconstruct human-computer interaction from a one-way mode of "one-time instructions" to a two-way dialogue of "continuous negotiation and calibration". In AI assisted writing, cues are not insignificant technical details, but key interfaces that determine the form of human-machine relationships. It may not only make learners passive recipients of algorithms, but also transform AI into a cognitive coordinator for writers [13]. Jiao and Chen systematically elaborated on the strategic value of cue design in large-scale language model English teaching, and revealed how high-quality cues create space for learners' subjective expression [17]. The true maturity of AI assisted writing should reach the ideal state of 'the deeper the interaction, the clearer the author's own position'. For this, learners need to internalize a set of negotiation writing mechanisms. The first step is to identify which parts of the AI output can be used as resources and which parts require careful algorithm presets. Next is to compare and examine the differences in logical strength and cultural bias between different generated versions. The third is to organically combine and restructure the language resources provided by AI with one's

own ideological framework. The fourth is to transform the negotiated resources into a unique expression that truly carries the author's voice [2]. Knoth et al empirically revealed the intrinsic relationship between AI literacy and cue strategies, confirming the critical significance of this negotiation mechanism in protecting writing subjectivity [18].

However, the negotiation process does not automatically guarantee the accuracy of the results. Even if learners maintain a certain degree of dominance through strategic interaction during the generation stage, the final text may still unknowingly fall into the cultural assumptions and value inertia of algorithms, which is a potential channel for "technological empowerment" to slide towards the subjective transfer mentioned above. Therefore, critical evaluation is an essential "final judgment" component in the APSE framework, which is the ultimate confirmation of "process differences" (from passive acceptance to ongoing negotiation).

In the context of AI assisted writing, evaluation involves at least four gradually layered judgments, corresponding to the actual implementation of the core issue of "author judgment" discussed earlier. The first is authenticity assessment, which involves fact checking and source tracing of content generated by AI to prevent misleading and deceptive practices. The second is logical coherence judgment, which checks whether the text only follows the traditional argumentative template of the algorithm, rather than presenting a unique reasoning route. The third is cultural suitability assessment, analyzing whether the text is excessively constrained by a single cultural perspective, thereby blurring local experiences and cultural self-awareness. The fourth is the author's judgment, questioning whether the text ultimately still carries the author's own sense of problem awareness, expression of stance, and value responsibility [6,14]. These four judgments form a subjective filter from "information credibility" to "expression credibility", returning the right to question "author identity" to the learner themselves. In practical teaching practice, teachers can guide students to create an "AI output review checklist" that transforms quadruple judgments into actionable and verifiable questions, such as "Is this information reliable?" and "Is this reasoning

route my own idea.

Strategy and evaluation are the 'rudder of the process' and the 'ruler of the outcome'. The collaborative operation of the two not only enables learners to control their writing direction in continuous dialogue with AI, but also confirms the true existence of subjectivity in the final text. This is the complete implementation of the "triple alignment transformation" of different goals, objects, and processes in writing practice, and it is also the logical hub of the APSE framework from theoretical construction to teaching application.

3.3 Integrated Analysis Matrix of Risk, Dimension, Task and Evaluation

Table 1. Critical AI Literacy Analysis Matrix in AI-Assisted Writing

Risk Type	Corresponding APSE Dimension	Typical Classroom Manifestation	Instructional Intervention	Observable Evaluation Indicator
Erosion of authorial voice	Critical awareness Critical evaluation	Directly adopting AI-generated opening statements or arguments; minimal differences between revisions and AI output	Implementing "position-first, collaboration-after" tasks; requiring justifications for accepting or rejecting AI suggestions	Personal stance reflected in revision notes; degree of divergence between final draft and AI output
Cultural stance drift	Critical awareness Critical evaluation	Using AI-generated Western examples; failing to detect cultural bias	Introducing cultural bias identification training; requiring the addition of local/native cases	Proportion of local cultural cases in the text; accuracy in identifying cultural stance
Outsourcing of judgmental responsibility	Critical interactive strategies Critical evaluation	Directly using AI-provided argumentation frameworks; failing to question AI suggestions	Establishing source verification procedures; guiding students to reflect on judgmental responsibility	Completeness of source verification records; critical analysis of responsibility attribution in reflections

The construction logic of a matrix is derived from the previous theoretical derivation. The risk of weakening the author's voice is essentially the lack of learners in the "critical position" dimension. They did not intervene in writing as "intruders" but as "receivers", nor did they conduct "critical evaluations" to determine the author identity of the final text. Therefore, the core dimensions corresponding to this risk are key positions and key assessments. Teaching interventions should focus on "speech recognition and reinforcement". For example, learners are required to first independently complete a position summary, then compare it with the version generated by AI, and finally modify it to form a text with a unique personal voice. The evaluation indicators can be set as "consistency between the final text and the independent outline", "frequency of language markers reflecting personal positions in the

The previous explanation of the four dimensions of the APSE framework and the analysis of three subjective risks in Section 3 have paved the way for critical AI to move from theoretical construction to teaching practice. However, in order to make this framework truly operational and instructive, it is still necessary to establish a systematic correspondence between risks, dimensions, teaching interventions, and evaluation indicators. Therefore, this study constructed a comprehensive analysis matrix of "risk literacy dimension teaching intervention evaluation indicators" (see Table 1 below) as a hub connecting theoretical analysis and teaching design.

text", etc. The risk of cultural stance drift stems from the blind spot of "critical awareness", where learners fail to perceive the cultural biases and value assumptions hidden in the output of AI. They also need to examine the appropriateness of culture through "critical evaluation". Therefore, the corresponding dimensions are critical awareness and critical evaluation. Teaching interventions can be designed with the task of "cultural bias detection". For example, learners are required to identify cultural centric expressions implicit in the output of AI and attempt to rewrite them from a local perspective. Evaluation indicators can include "accuracy of AI cultural preset recognition" and "clarity of rewriting the local cultural status in the text". The risk of outsourcing responsibility stems from the lack of a 'critical interaction strategy'. Learners simplified human-computer interaction into one-time instructions, abandoning the

negotiation process of "recognition comparison recombination conversion", and ultimately failed to recover responsibility through "critical evaluation". Therefore, the corresponding dimensions are critical interaction strategies and critical evaluation. The teaching intervention should adopt the process of "multi round prompt design negotiation process recording responsibility attribution statement", as shown in the example in Table 2 below. The evaluation criteria can focus on "reasonable explanations for rapid iteration", "analyzing the differences between the final text and AI output in terms of argumentation, structure, and examples", and "consciously explaining the reasons for adopting/rejecting AI suggestions in reflective articles".

The above matrix design establishes a strict logical cycle between three types of risks, four dimensions, a series of intervention measures, and measurable indicators, thereby addressing the theoretical necessity of the "triple paradigm shift" discussed in Section 2. At the same time, a clear blueprint was provided for the teaching approach to be developed in Section 4. The specific structure of Table 1 and its application in teaching experiments will be elaborated in Section 4, with reference to case studies.

4. Teaching Practice Path

4.1 Teaching Objectives: From "Being Able to Use AI" to "Thinking AI"

The above analysis of the three types of subjectivity risks reveals a profound educational paradox. When the ultimate goal of writing instruction is to be able to use AI, technological empowerment may only evolve into a channel for subjectivity transfer. In the pursuit of expression efficiency, students unconsciously pass on the author's voice, cultural status, and judgment responsibility to algorithms [4]. Therefore, AI-assisted writing teaching must undergo a fundamental transformation at the goal level, from instrumental rationality of "being able to use AI" to humanistic rationality of "thinking about AI".

The establishment of the goal of "AI thinking" is not only the implementation of the "triple paradigm shift" in previous teaching, but also a specific response to Pavlenko's (2025) concept of "holistic education". It requires learners to develop three progressive core competencies in

AI assisted writing [14]. Firstly, it is the ability to ask questions, treating the output of AI as a 'dialogue object', presenting continuous questions instead of passively accepting 'standard answers', and activating critical awareness in the cycle of cognition inquiry question'. The second is deliberative writing ability, which transforms human-computer interaction from a one-way mode of "instruction output" to a two-way cycle of "dialogue negotiation". Through strategic interventions such as prompt design, version comparison, and resource reorganization, it firmly controls the dominant position of discourse in the writing process. The third is responsible expression ability. When the final text is formed, it can clearly distinguish the contribution boundary between itself and AI, and provide effective explanations for adopting or rejecting AI suggestions, making "author identity" a conscious pursuit of writing [2,10].

4.2 Task Design: Prompt Ethics, Traceability Verification and Author Preservation

Given the subjective risks of the three types mentioned above, this study designed a "three-stage progressive" AI assisted argumentative writing task list (see Table 2 below). The task list focuses on "timely ethical traceability verification of copyright protection", ensuring that each stage of task design corresponds to active defense against specific risks, thereby transforming the goal of "thinking AI" into actionable teaching practices.

In the first stage, learners independently complete a position summary and outline of core arguments. This design directly addresses the risk of 'weakening the author's voice'. Before engaging with AI, learners must first complete independent stance expression and argument construction, forming their own problem awareness and expression intention. This' pre-positioned 'design aims to prevent learners from becoming' passive recipients' in subsequent interactions with AI and ensure that they always maintain the primary posture of 'interlocutors'. The submission of materials (position summary, argument outline) and ability orientation (author protection, problem-based competence) are special training for key position dimensions in the APSE framework.

In the second stage, learners use AI to unfold text and perform at least three rounds of prompt optimization. This design system should address the risk of outsourcing judgment responsibility. Learners need to record each round of prompt text, AI output, and prompt modification reasons, in order to transform human-computer interaction from "one-time instructions" to "continuous negotiation and calibration". Through multiple rounds of rapid iteration, learners gradually control their writing initiative in the negotiation mechanism of "recognition comparison recombination transformation", avoiding presetting complete delivery algorithms for argumentation frameworks and conclusions. The submission of materials (prompt records, reasons for modification) and ability orientation (key interaction strategies, prompt ethics) are the actual implementation of key interaction strategy dimensions in the APSE framework.

In the third stage, learners submit their final texts and reflective articles. This design fully conforms to the cumulative effects of "cultural positioning drift" and "responsibility outsourcing". Learners need to write reflective essays when submitting their final compositions, explaining the reasons for

adopting or rejecting AI suggestions, in order to cultivate their ability to critically examine the resulting outcomes. Learners need to conduct a retrospective evaluation of their collaboration with AI and make comprehensive judgments on the authenticity, logicity, cultural appropriateness, and authorship of the final text [6,14]. Reflective prose transforms "responsible expression" from an abstract concept to observable behavior, and learners realize the greatest ownership of the "author" in the process of explaining the reasons. The current ability orientation (critical evaluation and responsible expression) is the concentrated cultivation of critical evaluation dimensions in the APSE framework.

The three-stage task takes "risk prevention" as the logical mainline, gradually progressing and connecting layer by layer. In the first stage, establish a stance to resist the weakening of voices. In the second stage, negotiation takes the lead and prevents responsibility outsourcing. In the third stage, the power of adjudication is used to correct cultural drift. This design makes the cultivation of critical AI literacy a systematic response to three types of risks, achieving a comprehensive transformation from theoretical analysis to teaching practice.

Table 2. Task Sheet for AI-Assisted Argumentative Essay Writing

Phase	Task Description	Submission Materials	Competency Focus
Phase 1	Independently complete a position summary and a core argument outline	Position summary; argument outline	Authorial agency preservation; problematizing ability
Phase 2	Use AI to develop the main body, with at least three rounds of prompt optimization	Records of prompts and AI outputs for each round; justifications for prompt modifications	Critical interactive strategies; prompt ethics
Phase 3	Submit the final text and a reflective essay	Final essay; reflective essay (explaining reasons for accepting or rejecting AI suggestions)	Critical evaluation; responsible expression

4.3 Evaluation Optimization: From Result Scoring to Process Human-Computer Collaborative Evaluation

It is difficult to accurately assess students' actual abilities and development level in AI assisted writing solely based on their final written grades [4,6]. Although the final text can reflect the quality of language output, it cannot demonstrate how learners recognize the default values of algorithms, maintain dominant writing ability through negotiation, and confirm their position with multiple voices during interaction with AI. This is precisely the

core essence of the goal of "thinking about AI". Therefore, the evaluation method needs to shift from a single result-oriented approach to a human-machine collaborative evaluation process, and expand the evaluation focus from "what text is" to "how to generate text" and "how learners think".

Specifically, a "three-level progressive" evaluation system can be constructed. The first level is text outcome evaluation, which retains basic checks on traditional dimensions such as language accuracy, text organization, and logical reasoning to ensure the fundamental objectives of writing instruction. The second

level is process trajectory assessment, which focuses on the process evidence of learners in AI collaboration, such as prompting design strategies, version comparison tracking, modification reasons, and location statements, to determine the application level of key interaction strategies. The third level is metacognitive assessment, which examines whether learners can clearly distinguish the role boundaries of AI in writing, and whether they can provide effective explanations for

accepting or rejecting AI suggestions to verify whether they have formed a self-awareness of "authorship"[4,14]. The layered scaffold model proposed by Laflamme provides an operational theoretical framework for evaluating this process [19]. It emphasizes that evaluation should be embedded in the learning process and provide progressive feedback, which is highly consistent with the "three-level progressive" system of this study.

Table 3. Example of a Process-Oriented Assessment Rubric for AI-Assisted Writing

Assessment Dimension	Core Connotation	Correspondence to APSE Framework	Key Points for Evaluation	Score Level
Prompt Design	Whether learners can engage in sustained negotiation and calibration with AI through iterative prompt design, thereby maintaining control over the writing process within a cycle of "identifying-comparing-reorganizing-transforming"	Critical Interactive Strategies	<ul style="list-style-type: none"> Depth and progressiveness of prompt iteration Whether learners demonstrate critical examination and follow-up questioning of AI outputs <ul style="list-style-type: none"> Reasonableness and reflectiveness of justifications for modifications 	1-5
Source Verification	Whether learners can fact-check, critically examine the logic of, and trace the sources of AI-generated content, thereby resisting misinformation and misleading claims	Critical Evaluation (authenticity judgment, logical judgment)	<ul style="list-style-type: none"> Ability to identify factual errors in AI outputs Sensitivity to logical fallacies in arguments Whether learners proactively verify information sources 	1-5
Authorship Preservation	Whether learners can maintain their individual voice and stance in the final text, ensuring that AI outputs serve their own expression rather than replace their own thinking	Critical Positionality Critical Evaluation (authorial judgment)	<ul style="list-style-type: none"> Consistency between the final text and the independently developed outline Linguistic markers in the text that reflect the learner's personal stance <ul style="list-style-type: none"> Whether learners mechanically adopt formulaic expressions generated by AI 	1-5
Depth of Reflection	Whether learners can critically reflect on their collaborative relationship with AI, clearly define the boundaries of AI's role, and provide well-grounded justifications for accepting or rejecting AI suggestions	Critical Integrative Competence at the Metacognitive Level	<ul style="list-style-type: none"> Depth of explanation in the reflective essay regarding justifications for acceptance or rejection <ul style="list-style-type: none"> Clarity of understanding regarding the boundaries of AI's role Whether learners demonstrate conscious awareness of "authorship attribution" 	1-5

At the practical level, the "process portfolio" mechanism can be introduced, requiring students to systematically record the entire process of AI assisted writing, including independent outlines, multiple rounds of prompt texts and AI outputs, explanation of reasons for revisions, final texts, and reflective papers. The portfolio not only provides empirical evidence for process evaluation, but also encourages students to reflect on their learning trajectory while organizing materials,

thus achieving the educational function of "evaluation is learning".

Based on the above ideas, this study designed an AI assisted writing process evaluation scale (see Table 3 below), which concretizes the four-dimensional ability of APSE into observable and feedback teaching indicators. Each evaluation dimension of the gauge is clearly mapped to the APSE framework. Design corresponding critical interaction strategies and check whether learners have

achieved the negotiation process of "recognition comparison recombination transformation" through multiple prompts. Traceability verification corresponds to the authenticity and logical judgment in critical evaluation, examining learners' ability to check facts and verify the logic of content generated by AI. The author maintains a corresponding the entire writing process. It not only responds to the era proposition of "transitioning from technological rationality to humanistic rationality", but also provides scientific basis for teachers to diagnose teaching effectiveness and adjust intervention strategies, ultimately making the cultivation of critical AI literacy truly fall into the closed loop of "target task evaluation".

Critical stance and critical evaluation of the author's judgment, and checks whether the learner adheres to their personal voice and stance in the final text. The depth of reflection corresponds to critical integration ability at the metacognitive level, examining whether learners can critically reflect on their collaborative relationship with AI and clearly explain the reasons for adopting or rejecting AI suggestions. These four dimensions support each other and together constitute the integrity evaluation of the goal of "thinking AI".

Through the collaborative use of the "three-level progressive" evaluation system, process portfolio, and evaluation scale, evaluation is no longer the final judgment of learning outcomes, but a "peer" feedback mechanism that runs through

5. Conclusions

How to protect the authorship of learners in an era where algorithms are increasingly involved in language expression and cultural meaning generation? This is a fundamental issue that arises when generative AI is embedded in writing instruction at an unprecedented depth. Based on this issue, this study proposes the concept of "critical AI literacy" through theoretical construction and teaching path design, and systematically elaborates its theoretical connotation, risk response mechanism, and teaching implementation path using the APSE framework.

In order to achieve this transformation, this study has made three core contributions at the theoretical level. Firstly, the disciplinary integration of theoretical resources involves

incorporating critical language awareness and critical information literacy into the research of AI literacy, clarifying their respective analytical functions of "identifying technological biases" and "verifying information authenticity", laying a special disciplinary foundation for critical AI literacy in foreign language education. This integration directly responds to the introduction's focus on the dual status of "language non neutrality" and "technology non neutrality". Secondly, the functional extension of the analysis framework is based on APSE model, which expands the static "ability analysis" framework into a dynamic "teaching design" tool, enabling it to systematically integrate practical aspects such as risk identification, teaching intervention, and evaluation indicators. This exposition enables the four dimensions of APSE to not only explain the mechanism of subjective risk generation (as described in Section 3), but also guide the systematic design of teaching practice (as shown in Section 4). Thirdly, by fundamentally distinguishing between discriminative AI literacy and general AI literacy, the ontological reconstruction of the literacy paradigm shows that the latter is only a linear extension of the "skill list", while the former itself represents the reconstruction of problem awareness. Therefore, its core focus has shifted from "whether AI can be effectively used" to "whether the author's judgment can be maintained when collaborating with AI". This transformation is the concrete implementation of the "three person" exploration in terms of literacy level (who is speaking, who is determining meaning, and who is not taking responsibility), which also forms a close loop from problem to theoretical response.

This study has several limitations, which also point the way for future exploration. From the perspective of research scenarios, this article mainly focuses on AI assisted writing teaching, and the applicability of the APSE framework in other language skills such as speaking, translation, and reading still needs to be tested. Future research can explore the adjustment path of this framework in different scenarios and test its cross scenario explanatory ability. From the nature of the research, this study belongs to theoretical construction, and the proposed critical AI literacy framework and teaching path design still require empirical methods such as classroom observation, action

research, and quasi experiments to verify its effectiveness, feasibility, and educational value in practical teaching contexts. From the perspective of research subjects, teachers, as key participants in teaching implementation, need to further study their critical AI literacy and training paths. Future research should focus on how teachers can take the lead in transitioning from "technology users" to "key leaders" in human-computer collaborative teaching practices, and how to construct corresponding teacher development support systems.

The iteration of technology is endless, but the fundamental exploration of education is consistent. This study's exploration of critical AI literacy is a response to the issues of this era. Today, as technology is deeply rooted in the educational context, we need to ask more questions than ever before. We hope that this research can receive richer feedback in future research and practice on how to protect human educational ideals beyond technology.

Acknowledgments

This research was funded by Higher Education Science Research Project of Chongqing Higher Education Society, 2025–2026 (Project No.: cqgj25043C), Special Project on “Reading and Teacher Development” of the 14th Five-Year Plan of the Chinese Tao Xingzhi Research Association (Project No.: CTAO 202512926CQ) and Language and Writing Research Project of Chongqing Normal University, 2025 (Supranational English Language Order and Local Discourse Resistance: The Generative Mechanism and Adaptive Strategies of Language Ideologies among Pre-service English Teachers , Project No.: 202504).

References

- [1] Wang C, Aguilar S J, Bankard J S, et al. Writing with AI: What college students learned from utilizing ChatGPT for a writing assignment. *Education Sciences*, 2024, 14(9):976.
- [2] Xu F. AI empowered cultivation of ideological depth in English writing: A teaching action research. *Foreign Language World*, 2025, (6):59-66, 82.
- [3] Sperling K, Stenberg C J, McGrath C, et al. In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open*, 2024, 6:100169.
- [4] Zhang J. College students' AI-empowered foreign language learning literacy: Framework construction and scale development. *Foreign Language Audio-Visual Teaching*, 2025, (1):69-78, 122.
- [5] Wang H, Xie Y, Zhang P, et al. Pre-task planning in L2 writing: Does generative AI assistance make a difference? *Journal of Second Language Writing*, 2026, 72:101320.
- [6] Zou M L, Reinders H, Amjad F. Understanding the potential role of GenAI-mediated informal digital learning of English (GenAI-IDLE) in the Global South: AI literacy, emotions, and willingness to communicate as outcomes. *ReCALL*, 2026, 38(1):131-149.
- [7] Albeih H H M, Rice M F. Generative AI and language diversity: Implications for teachers and learners. *Arab World English Journal*, 2025, 16(1):43.
- [8] Sun Y. Resisting homogenization in EAL writing education: Translingual practices in the rise of generative AI. *TESOL Quarterly*, 2026.
- [9] Abrams S S, Hanghøj T. Reflexive AI literacies: Conceptualizing a model for learners' critical engagement with generative AI. *Pedagogies: An International Journal*, 2025, 21(2):214-233.
- [10] Wang C, Wang Z. Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing*, 2025, 67:101187.
- [11] Darvin R. Identity and investment in the age of generative AI. *Annual Review of Applied Linguistics*, 2025, 45:1-18.
- [12] Veldhuis A. Critical artificial intelligence literacy: A scoping review and framework synthesis. *International Journal of Child-Computer Interaction*, 2025:100708.
- [13] Agirdag O. Beyond prompt engineering: Prompting (l)iteracy, linguistic capital, and educational inequality. *Educational Theory*, 2026, 76(2):206-223.
- [14] Pavlenko O. Who CAREs? A dialogic framework for critical AI literacy in multilingual classrooms. *Inquiry: Critical Thinking Across the Disciplines*, 2026, 35(1):215-238.
- [15] Tacelosky K, Kasun G S, Shapiro B R, et

- al. Exploring critical AI literacy in language education: A case study. *Foreign Language Annals*, 2025, 58(4):966-994.
- [16]Liu F, Dong Y. Why is critical thinking becoming increasingly important in the era of generative artificial intelligence? *Foreign Language Audio-Visual Teaching*, 2025, (1):3-10, 113.
- [17]Jiao J, Chen T. Large language models empowering English teaching: Four scenarios. *Foreign Language Audio-Visual Teaching*, 2023, (2):12-17, 106.
- [18]Knoth N, Tolzin A, Janson A, et al. AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, 2024, 6:100225.
- [19]Laflamme K A. Scaffolding AI literacy: An instructional model for academic librarianship. *The Journal of Academic Librarianship*, 2025, 51(3):103041.