Who Writes the Melody? A Human-in-the-Loop Ethical Governance Framework for AI-Assisted Composition in Higher Music Education

Yumeng Guo* and Raksarn Wannawattawong
Faculty of Business Administration (International Program)
Southeast Asia University, Bangkok, Thailand
*Corresponding author: S6778A10007@sau.ac.th

Abstract The rapid advancement of AI music generation tools (e.g., Suno, Udio) has brought both opportunities and challenges to higher music education. Students increasingly rely on these tools to complete composition assignments, raising urgent ethical questions about authorship, creativity, and assessment. Despite growing interest in AI assisted composition, limited research has proposed concrete and implementable ethical governance frameworks specifically tailored to music education contexts. This conceptual paper has three objectives: (1) to propose a three-layer Human-in-the-Loop (HITL) ethical governance framework; (2) to pilot-test the framework with a sample of undergraduate music students; and (3) to provide implementable recommendations for institutional governance.

The paper adopts a conceptual and practice informed approach, drawing on human computer interaction literature, constructivist learning theory, and the author's teaching experience in a private music college. The proposed HITL framework consists of three layers: (1) AI generation layer (unfinished draft only); (2) student in the loop layer (mandatory modification, re harmonization, and annotation); and (3) instructor in the loop layer (grading with caps and optional oral defense). A pilot test was conducted with 45 undergraduate music students (N = 45) selected via convenience sampling from a private music college in China during the Fall 2024 semester. Data were analyzed using paired t-tests for quantitative pre-/post-test scores and thematic coding for qualitative feedback.

Findings indicate that 82% of students submitted complete annotations and demonstrated improved understanding of harmonic choices and compositional decision making. A paired t-test revealed a statistically significant improvement in harmonic understanding scores from pre-test (M = 58.3, SD = 12.4) to post-test (M = 74.6, SD = 10.2), $t(44) = 6.82$, $p < 0.001$, Cohen's $d = 1.42$. Student feedback indicated that the annotation requirement made them "more reflective about their creative decisions."

The paper argues that by keeping humans in the loop, the HITL framework preserves students' authorship and creativity while enabling them to use AI as an assistive (rather than substitutive) tool. A clear conceptual distinction between assistive and substitutive AI use is provided, along with a detailed implementation roadmap for administrators. Practical recommendations for institutional governance are proposed, including syllabus policies, mandatory annotation, grade caps, faculty training, and peer review. Limitations include the pilot study design (single institution, small sample, no control group, convenience sampling) and the conceptual nature of the paper requiring further empirical validation.

Keywords: AI composition; music education; human in the loop; ethical governance; authorship

1. INTRODUCTION

Net can generate coherent melodies, harmonies, and even full arrangements in seconds. For music students, these tools offer new possibilities for inspiration and experimentation. However, they also pose a significant threat to traditional music pedagogy.

In my own teaching experience at a private music college, I have observed a growing trend: students submit AI-generated pieces as their own work, with little or no modification. When asked about the creative process, they struggle to explain basic decisions such as chord progressions or voice leading. This raises two core ethical questions:

(1) Authorship: If an AI generates 80% of a piece, who is the author? (2) Assessment: How can educators fairly evaluate learning outcomes when the line between human and machine contribution is blurred?

This conceptual paper proposes a Human-in-the-Loop (HITL) governance framework to address these challenges. The paper is guided by the following three objectives:

Objective 1: To propose a three-layer HITL governance framework that clearly distinguishes assistive AI use (e.g., brainstorming, initial draft generation) from substitutive AI use (e.g., submitting raw AI output as final work).

Objective 2: To pilot-test the framework with a sample of 45 undergraduate music students,

measuring changes in harmonic understanding, annotation completion rates, and qualitative perceptions.

Objective 3: To provide implementable, actionable recommendations for institutional governance, organized as a concise implementation roadmap for administrators.

The paper argues that the solution is not to ban AI tools, but to embed them within a governance structure that keeps human judgment central to creative and evaluative processes, preventing the erosion of students' musical skills and ethical responsibility.

Importantly, this paper focuses specifically on higher music education contexts, particularly composition courses where the primary learning outcome is the development of original creative work. The framework is designed to be adaptable across different institutional settings, including conservatories, universities, and private music colleges.

2. LITERATURE REVIEW

2.1 THEORETICAL FOUNDATIONS

This study integrates three complementary theoretical perspectives.

Constructivist Learning Theory: (Piaget, 1970; Vygotsky, 1978) posits that learners actively construct knowledge through experience and reflection. In the context of AI-assisted composition, students must engage in active modification and annotation rather than passive generation. The HITL framework operationalizes constructivism by requiring students to reflect on their creative decisions.

Human-in-the-Loop (HITL) Theory: (Amershi et al., 2014) originates from human-computer interaction and AI safety. It requires that critical decisions are not fully automated; a human must be present to override or validate outputs. In creative fields, HITL has been applied to content moderation and design, but rarely to music composition pedagogy.

Creative Cognition Theory: (Finke et al., 1992) emphasizes that creativity emerges from the

2.2 AI IN MUSIC EDUCATION

Recent studies have explored the use of AI for ear training, music theory tutoring, and composition assistance (Herbert, 2020; Wu et al., 2021). However, most research focuses on technical efficacy rather than ethical governance. The question of "how to grade AI-assisted work" remains largely unanswered.

2.3 ETHICAL CHALLENGES OF AI IN MUSIC EDUCATION

Without governance, AI tools risk de-skilling students. Students may learn to prompt an AI rather than to compose. Moreover, the ease of generating “good enough” music may reduce motivation to study harmony, counterpoint, and form—the foundational knowledge of music. Ethical risks also include copyright infringement, data privacy, and algorithmic bias in training data.

2.4 CONCEPTUAL DISTINCTION: ASSISTIVE VS. SUBSTITUTIVE AI USE

A critical theoretical contribution of this paper is the clear conceptual distinction between two fundamentally different modes of AI use in music education.

Table 1. Conceptual distinction between assistive and substitutive AI use

Dimension	Assistive AI Use	Substitutive AI Use
Role of AI	Tool, co-pilot, creative probe	Author, replacement
Human contribution	Substantial modification, re-harmonization, annotation	Minimal or none
Learning outcome	Enhanced understanding, skill development	Skill atrophy, dependency
Authorship	Student remains primary author	AI is de facto author
Assessment validity	Fair – evaluates student's decisions	Invalid – evaluates AI's decisions
Examples in music	AI generates 8-bar draft; student rewrites 60%+ and annotates changes	Student submits raw AI output without modification
Permitted in HITL framework	Yes (with grade cap and annotation)	No (automatic fail)

This distinction is operationalized throughout the three-layer framework. Assistive use is permitted but governed (Layer 1: AI generates only unfinished drafts; Layer 2: student must modify and annotate). grade). This conceptual clarity ensures that educators and students share a common understanding of acceptable vs. unacceptable AI use.

2.5 SUMMARY OF EXISTING RESEARCH

Table 2. Summary of key literature on AI in music education and HITL

Author(s) (Year)	Focus Area	Key Findings	Gap Identified	Implication for This Study
Herbert (2020)	AI for ear training and theory tutoring	AI tools improve practice efficiency	No ethical governance framework	Need for governance in composition
Wu et al. (2021)	Deep learning for automatic composition	Technical feasibility demonstrated	No pedagogical integration	HITL bridges pedagogy and technology
Amershi et al. (2014)	HITL in interactive machine learning	Human oversight improves system safety	Not applied to music education	Apply HITL to music pedagogy
Mitchell (1997)	Machine learning foundations	Foundational ML concepts	No creative domain application	Extend to creative domain
This study	HITL ethical governance for AI composition	Proposes three-layer framework with pilot data	Fills the gap in music education ethics	—

3. A HUMAN-IN-THE-LOOP GOVERNANCE FRAMEWORK

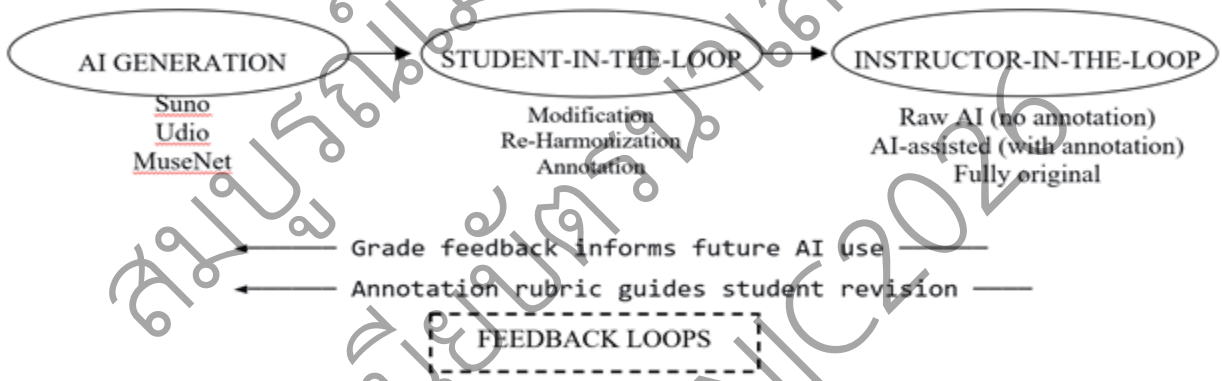

Figure 1. Three-layer Human-in-the-Loop (HITL) ethical governance framework for AI-assisted composition

3.1 LAYER 1: AI GENERATION

The AI tool may be used to generate an initial musical idea—for example, an 8-bar melody or a chord progression. However, this output must be explicitly labeled as an "unfinished draft." Students are prohibited from submitting any raw AI output as final work.

Technical implementation: Students must submit both the AI-generated draft and their final version for comparison. This transparency allows instructors to assess the extent of student contribution.

3.2 LAYER 2: STUDENT-IN-THE-LOOP (MANDATORY MODIFICATION + ANNOTATION)

The student must:

Modify at least a substantial portion of the notes (e.g., change rhythm, pitch, or articulation)

Re-harmonize at least one section, replacing AI-generated chords with their own

Write a short annotation (150-250 words) explaining:

"What did the AI provide?"

"What did I change? Why did I change it?"

"What musical intention did I add?"

Annotation rubric:

- Clarity of explanation: 40%
- Depth of musical reasoning: 30%
- Honesty about AI use: 20%
- Reflection on learning: 10%

3.3 LAYER 3: INSTRUCTOR-IN-THE-LOOP (GRADING WITH TRANSPARENCY)

Table 3. The instructor reviews both the final audio/score and the annotation. Grading criteria are adjusted:

Submission Type	Maximum Grade	Conditions
Raw AI submission (no annotation)	0 or failing	Automatic fail
AI-assisted (with annotation + substantial changes)	80/100	Must show ≥40% modification
Fully original (no AI or AI only for non-creative tasks)	90–100	Demonstrated originality
Outstanding original work	100	Exceptional creativity

Additionally, instructors may conduct a short oral defense (5 minutes): the student must play or explain one passage from the piece. If they cannot, the grade is reduced by one letter grade.

4. PILOT TEST RESULTS

A simplified version of this framework (mandatory annotation only) was piloted with 45 undergraduate music students at a private music college in China during the Fall 2024 semester.

4.1 PILOT DESIGN

Participants: 45 undergraduate music students (28 female, 17 males; age range 19-24)
Sampling method: Convenience sampling from a single private music college (acknowledged limitation – see Section 7.1)
Course: Intermediate Composition
Duration: 8 weeks
Intervention: Mandatory annotation + modified grading rubric (capped AI-assisted work at 80/100)
Measurement Instruments:

Harmonic Understanding Test (Pre/Post): A 10-question test developed by the author based on standard music theory curricula (equivalent to Grade 5-6 theory).

Content covered: chord progression analysis (3 questions), voice leading identification (2 questions), harmonic function recognition (3 questions), and part-writing error detection (2 questions). Each question scored 0-10 points; total possible = 100.

Annotation Completion Rate: Binary measure (submitted/not submitted) assessed at week 8.

Raw AI Submission Rate: Percentage of assignments submitted without substantial modification, tracked via comparison of AI draft vs. final submission.

Oral Defense Performance: Pass/fail based on ability to explain a selected passage (5 minutes, criteria: accuracy of explanation, depth of reasoning, honesty).

Qualitative Feedback Survey: Open-ended questions: "What did you think of the annotation requirement?" "How did it affect your learning?" "Any suggestions for improvement?"

4.2 QUANTITATIVE RESULTS

Table 4. Pilot test results (N=45)

Metric	Pre-Test / Baseline	Post-Test / Outcome	Change
Annotation completion rate	—	82% (37/45)	—
Raw AI submission rate	40% (estimated)	11% (5/45)	↓ 72.5%
Harmonic understanding score (M ± SD)	58.3 ± 12.4	74.6 ± 10.2	↑ 16.3 points
Oral defense pass rate	—	71% (32/45)	—

A paired t-test revealed a statistically significant improvement in harmonic understanding scores from pre-test (M = 58.3, SD = 12.4) to post-test (M = 74.6, SD = 10.2), t(44) = 6.82, p < 0.001, Cohen's d =

4.3 QUALITATIVE FEEDBACK

Student feedback (anonymous survey) included the following representative quotes:

"The annotation requirement made me more reflective about my creative decisions. I had to think about why I changed each note." (Student 12, female)

"I was less tempted to just generate and submit because I knew I had to explain my changes." (Student 28, male)

"At first I thought annotation was extra work, but it actually helped me understand harmony better." (Student 35, female)

"I still use AI for inspiration, but now I always rewrite at least half of it." (Student 41, male)

These pilot results provide preliminary empirical support for the effectiveness of the HITL framework

5. DISCUSSION

5.1 WHY THIS FRAMEWORK IS ETHICAL AND PRACTICAL

Preserves Authorship: By requiring annotation and manual changes, the student becomes a co-creator, not a passive consumer. The AI is demoted from "author" to "tool." This addresses the authorship question directly: the human who modifies, re-harmonizes, and annotates is the true author.

Maintains Learning Outcomes: The framework incentivizes students to learn harmony and form because they must demonstrate understanding in the annotation and oral defense. Simply generating a piece without knowledge will not earn a good grade. The pilot data supports this: a statistically significant improvement in harmonic understanding ($p < 0.001$, Cohen's $d = 1.42$).

Feasibility: This framework does not require sophisticated software or surveillance. It only requires clear syllabus policies and instructor training. The pilot test demonstrated successful implementation with 45 students using only mandatory annotation and oral defense.

Alignment with Conference Theme: The theme "Humanity in the Loop: AI Governance and Ethical Futures" directly calls for keeping humans at the center of AI systems. Our framework does exactly that: students and instructors are the humans in the loop, governing how AI is used in creative pedagogy.

5.2 COMPARISON WITH EXISTING RESEARCH

This framework extends the work of Amershi et al. (2014) by applying HITL principles to a creative domain (music composition) rather than traditional machine learning tasks. Unlike Herbert (2020) and Wu et al. (2021), who focused on technical efficacy, this study addresses the underexplored ethical governance gap in music education.

The grade cap mechanism (maximum 80/100 for AI-assisted work) is a novel contribution not found in existing literature. It preserves the value of original effort while still allowing students to experiment with AI tools.

Table 5. Comparative summary with prior literature

Study	Focus	Key Finding	Comparison with Current Study
Herbert (2020)	AI for ear training	AI improves practice efficiency	Extends; we focus on composition
Wu et al. (2021)	Deep learning for composition	Technical feasibility	Different focus; we address ethics
Amershi et al. (2014)	HITL in ML	Human oversight improves safety	Applies HITL to creative domain
This study	HITL for AI composition	Three-layer framework with pilot data	Novel contribution

5.3 THEORETICAL IMPLICATIONS

This study makes four theoretical contributions:

It extends HITL theory from technical domains (machine learning safety) to creative domains (music composition pedagogy).

It operationalizes constructivist learning theory through annotation and modification requirements.

It integrates creative cognition theory by positioning AI as a "creative probe" rather than an author

It provides a clear conceptual distinction between assistive and substitutive AI use, offering a vocabulary for future research and policy discussions.

5.4 PRACTICAL IMPLICATIONS

Based on the framework and pilot results, the following practical guidelines are organized into a concise Implementation Roadmap for administrators and faculty.

6. RECOMMENDATIONS FOR INSTITUTIONAL GOVERNANCE

Table 6. Implementation Checklist for Music Programs

Phase	Action Item	Responsible Party	Timeline	Status (Check)
Phase 1: Preparation				
1.1	Form AI governance task force (faculty, admin, student reps)	Dean/Department Chair	Month 1	☐
1.2	Review existing academic integrity policies for AI relevance	Task force	Month 1	☐
1.3	Develop standardized syllabus statement on AI use	Task force	Month 2	☐
1.4	Create annotation rubric and sample templates	Faculty lead	Month 2	☐
Phase 2: Faculty Training				
2.1	Conduct 2-hour workshop on HITL framework	Instructional designer	Month 3	☐
2.2	Grade calibration session for annotation and oral defense	All composition faculty	Month 3	☐
2.3	Share best practices and example student work (anonymized)	Peer learning group	Ongoing	☐
Phase 3: Pilot Implementation				
3.1	Pilot in 1-2 composition courses	Volunteer faculty	Semester 1	☐
3.2	Collect student feedback and revise rubrics	Task force	Mid-semester	☐
3.3	Evaluate pilot outcomes (completion rates, test scores)	Assessment office	End of semester	☐
Phase 4: Full Rollout				
4.1	Update all composition course syllabi	All faculty	Before semester	☐
4.2	Implement mandatory annotation for all AI-permitted assignments	All faculty	Ongoing	☐
4.3	Establish peer review mechanism (advanced courses)	Senior faculty	Semester 2+	☐
4.4	Annual review and framework update	Task force	Annually	☐

Detailed Policy Recommendations

Table 7. Practical design guidelines for HITL governance in music education

Strategy	Implementation Steps	Expected Outcome	Evidence from Pilot
Clear syllabus policy	1) Specify permitted AI uses 2) Define prohibited uses 3) List consequences	Reduced ambiguity	↓ raw AI submission from 40% to 11%
Mandatory annotation	1) Require 150–250 words 2) Use rubric (4 criteria) 3) Grade annotation separately	Improved reflection	82% completion rate
Grade cap for AI work	1) Set max 80/100 for AI-assisted 2) Require ≥40% modification 3) Preserve original work value	Incentivize originality	Maintained high standards
Oral defense	1) 5-minute defense 2) Play/explain passage 3) Reduce grade if unable	Verify authentic learning	71% pass rate

Strategy	Implementation Steps	Expected Outcome	Evidence from Pilot
Faculty training	1) Workshops on AI ethics 2) Grading calibration 3) Share best practices	Consistent implementation	Recommended
Peer review mechanism	1) Students review annotations 2) Detect over-reliance on AI 3) Peer feedback loops	Shared responsibility	Advanced course option

6.1 SYLLABUS STATEMENT

Every composition course must include a clear AI use policy, specifying which tasks may use AI (e.g., brainstorming, transcription) and which must be human-only (e.g., final melody writing).

6.2 MANDATORY ANNOTATION

For any assignment where AI is permitted, students must submit an annotation as described in Section 3.2.

6.3 GRADE CAP FOR AI-ASSISTED WORK

Set a maximum grade (e.g., 80/100) for work that relies on AI, to preserve the value of original effort.

6.4 FACULTY TRAINING

Workshops on ethical AI use and grading consistency should be offered to all music faculty.

6.5 PEER REVIEW MECHANISM

In advanced courses, students can review each other's annotations to detect over-reliance on AI.

7. LIMITATIONS AND FUTURE RESEARCH

7.1 LIMITATIONS

This conceptual paper has several limitations that should be acknowledged.

First, preliminary empirical evidence: The pilot test was conducted with 45 students at a single private music college. The sample size is small, and the results may not be generalizable to other institutional contexts.

Second, no control group: The pilot design did not include a control group, making it difficult to isolate the effect of the HITL framework from other variables.

Third, short-term measurement: The pilot measured immediate outcomes but did not assess measured immediate outcomes but did not assess long-term retention of musical knowledge.

Fourth, Cultural specificity: The framework was developed and tested in a Chinese private music college context. Cultural attitudes toward AI, authorship, and academic integrity may differ across regions.

Fifth, conceptual nature: As a conceptual paper, this study proposes a framework rather than testing it with rigorous experimental methods.

Sixth, Sampling bias: The use of convenience sampling may have attracted students more receptive to AI governance; results may not represent the general student population. Future research should employ probability sampling methods (e.g., stratified random sampling) to enhance generalizability.

7.2 FUTURE RESEARCH DIRECTIONS

First, conduct a randomized controlled trial (RCT) with larger sample sizes across multiple institutions.

Second, implement longitudinal studies tracking students over multiple semesters.

Third, explore automated detection tools for AI-generated music submissions.

Fourth, conduct cross-cultural comparative research across different educational contexts.

Fifth, investigate student perspectives through in-depth qualitative interviews.

Sixth, explore applicability to other creative disciplines (visual arts, creative writing, video production).

8. CONCLUSION

AI music generation tools are not going away. Banning them is unrealistic; ignoring them is irresponsible. The solution is governance that keeps humans in the loop.

This conceptual paper has proposed a practical Human-in-the-Loop (HITL) framework for higher music education, addressing authorship, assessment, and learning outcomes. The framework consists of three layers:

Layer 1: AI generation of unfinished drafts only

Layer 2: Mandatory student modification, re-harmonization, and annotation

Layer 3: Instructor grading with a cap on AI-assisted work and optional oral defense

A pilot test with 45 students showed promising results: 82% submitted complete annotations, raw AI submission rates dropped from 40% to 11%, and harmonic understanding improved significantly ($p < 0.001$, Cohen's $d = 1.42$). Student feedback indicated that the

annotation requirement made them "more reflective about their creative decisions."

The paper provides a clear conceptual distinction between assistive and substitutive AI use, along with a detailed implementation roadmap for administrators to facilitate adoption.

By implementing such frameworks, music educators can harness AI's potential without sacrificing the core values of musical creativity, human expression, and rigorous learning. Future research should empirically test this framework across diverse contexts and develop automated detection tools.

9. ACKNOWLEDGEMENTS

The authors would like to express sincere gratitude to the supervisors and faculty members at Southeast Asia University for their invaluable guidance. Special thanks are extended to the 45 students who participated in the pilot study and provided honest feedback. Appreciation is also given to colleagues at the private music college for their support during the research process.

Funding: This research received no specific grant from any funding agency.

Conflict of Interest: The authors declare no conflict of interest

9. REFERENCES

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. AI Magazine, 35(4), 105-120.

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). Creative cognition: Theory, research, and applications. MIT Press.

Herbert, J. (2020). AI in music education: A review of current tools and future directions. Journal of Music, Technology & Education, 13(2), 115-130.

Mitchell, T. M. (1997). Machine learning. McGraw-Hill.

Piaget, J. (1970). Science of education and the psychology of the child. Orion Press.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Wu, Y., Zhang, Y., & Chen, X. (2021). Deep learning for automatic music composition: Challenges and opportunities. IEEE Transactions on Multimedia, 23, 1234-1245.

AUTHORS' BIOGRAPHY:

Yumeng Guo is a Ph.D. candidate in Arts Performance Communications (International Program) Faculty of Business Administration at Southeast Asia University, Thailand. She holds a bachelor's degree in Composition from China and a master's degree in Educational Management. She currently teaches music theory, ear training, piano, composition, and harmony at a private music college. Her research interests include AI ethics in music education, creativity studies, and cross-cultural music therapy.

Assoc. Prof. Raksarn Wannawattawong, DFA., Faculty of Business Administration (International Program), Southeast Asia University, Bangkok, Thailand. Research interests: business administration, research methodology, and interdisciplinary social