müzikist akademik dergisi musicist academic journal

ARTIFICIAL INTELLIGENCE IN VOCAL PEDAGOGY: A CRITICAL REVIEW OF TOOLS, PEDAGOGICAL FUNCTIONS, AND RESEARCH TRENDS

SES EĞİTİMİNDE YAPAY ZEKÂ: ARAÇLAR, PEDAGOJİK İŞLEVLER VE ARAŞTIRMA EĞİLİMLERİ ÜZERİNE ELEŞTİREL BİR DERLEME

Bilge ATAY KARLIDAĞ

Cilt • Vol. 1 | Sayı • Issue 2
Ekim • October 2026 | s. • pp. 6-19
e-ISSN: 3149-8264
www.muzikist.org

Research Article - Araştırma Makalesi

Cilt • Vol. 1 | Sayı • Issue 2
Ekim • October 2026 | s. • pp. 6-19
e-ISSN: 3149-8264
www.muzikist.org

Except where otherwise noted, content is licensed under CC BY-NC 4.0. Icons by Font Awesome.

Research Article - Araştırma Makalesi

Artificial Intelligence in Vocal Pedagogy: A Critical Review of Tools, Pedagogical Functions, and Research Trends

Ses Eğitiminde Yapay Zekâ: Araçlar, Pedagojik İşlevler ve Araştırma Eğilimleri Üzerine Eleştirel Bir Derleme

Bilge ATAY KARLIDAĞ

Lecturer Dr., Gaziantep University, Turkish Music State Conservatory, Department of Voice Education
bilgeatay@gantep.edu.tr
ORCID: 0000-0002-1689-7575

Abstract

Artificial intelligence (AI) is now used in vocal pedagogy for performance analysis, automated assessment, feedback, and individualized practice. The literature, however, does not support treating computational output as self-evidently pedagogical. This critical review examines 15 verified English-language academic studies published between 2022 and 2026, comparing the technologies employed, the vocal skills targeted, the forms of feedback provided, and the pedagogical roles assigned to AI. The studies show a field broadening from classification and automated scoring toward deep learning, large-language-model dialogue, explainable AI, adaptive recommendation, and multimodal systems. Pitch, rhythm, timbre, and pronunciation receive the greatest attention because they are comparatively tractable computationally; reflective learning and self-regulation are less developed. Cultural-stylistic validity is treated here not merely as an inclusivity concern but as a measurement-validity problem whenever the system's reference norm does not match the musical tradition being evaluated. Drawing on these patterns, the review proposes the Vocal-AI Pedagogical Function Framework (VAPFF), which treats human interpretation and pedagogical action as necessary links between technical analysis and learning. AI in vocal pedagogy should therefore be judged not only by how accurately it measures performance, but by whether its output can be interpreted, situated within a pedagogical context, and used without weakening learner or teacher agency.

Keywords: *artificial intelligence, vocal pedagogy, singing education, feedback, music education, human centered AI*

Öz

Yapay zekâ (YZ) destekli araçlar ses eğitiminde performans analizi, otomatik değerlendirme, geri bildirim ve kişiselleştirilmiş çalışma desteği amacıyla giderek daha fazla kullanılmaktadır. Ancak vokal performansın hesaplanabilir hâle gelmesi, üretilen her çıktının pedagojik olarak anlamlı olduğu anlamına gelmez. Bu eleştirel derleme, 2022-2026 yılları arasında yayımlanan ve tam metni doğrulanabilen 15 İngilizce akademik çalışmayı teknoloji, hedeflenen vokal beceriler, geri bildirim biçimleri ve pedagojik işlevler bakımından karşılaştırmaktadır. İncelenen çalışmalar, alanın sınıflandırma ve otomatik puanlamadan derin öğrenme, büyük dil modeli tabanlı diyalog, açıklanabilir YZ, adaptif öneri ve çok modlu sistemlere doğru genişlediğini göstermektedir. Perde, ritim, tını ve telaffuz gibi hesaplamalı olarak daha kolay temsil edilebilen boyutlar belirginken; yansıtıcı öğrenme ve öz-düzenleme daha sınırlı biçimde ele alınmaktadır. Kültürel-stilistik geçerlik ise bu çalışmada yalnızca bir kapsayıcılık meselesi olarak değil, sistemin kullandığı referans norm ile değerlendirilen müzik geleneği uyuşmadığında doğrudan bir ölçme geçerliği sorunu olarak ele alınmaktadır.

Geliş Tarihi/Received: 25.08.2026 - **Kabul Tarihi/**Accepted: 21.09.2026 - **Yayın Tarihi/**Published: 30.10.2026

Atay Karlıdağ, B. (2026). Artificial intelligence in vocal pedagogy: A critical review of tools, pedagogical functions, and research trends. Müzikist Akademik Dergisi, 1(2), 6-19.

Bu bulgulardan hareketle Vokal-Yapay Zekâ Pedagojik İşlev Çerçevesi (VAPFF) önerilmektedir. Çerçeve, teknik analiz ile öğrenme arasında insan yorumunu ve pedagojik eylemi zorunlu bir ara katman olarak konumlandırmaktadır. Derlemenin temel savı, ses eğitiminde YZ'nin değerinin yalnızca ölçüm doğruluğuyla değil, üretilen bilginin öğrencinin ve öğretmenin kullanabileceği, bağlama duyarlı ve özerkliği destekleyen bir pedagojik kaynağa dönüşme kapasitesiyle değerlendirilmesi gerektiğidir.

Anahtar kelimeler: *yapay zekâ, ses eğitimi, vokal pedagoji, geri bildirim, müzik eğitimi, insan merkezli yapay zekâ*

Introduction

Artificial intelligence is no longer a marginal topic in music education. It is already present in systems that classify performance, detect deviations, generate scores, recommend practice material, and produce conversational or adaptive feedback. Reviews of the wider music-education literature describe substantial growth in performance assessment, personalization, intelligent tutoring, and generative applications, while also noting unresolved questions about pedagogical validity, cultural sensitivity, and the role of the teacher (Mazlan et al., 2026). Those questions become especially important in voice education, where a sounding result is inseparable from bodily coordination, sensory feedback, stylistic judgment, and the ongoing work of the teacher.

For that reason, singing cannot be reduced to a sequence of acoustic measurements. Producing a sung phrase requires auditory perception, motor planning, respiration, phonation, resonance, articulation, and somatosensory information to operate together. The neuroscience of singing has long emphasized auditory-motor control (Zarate, 2013), while embodied accounts of vocal and instrumental teaching describe learning as something that emerges through the interaction of learner, teacher, body, task, and environment (Bremmer & Nijs, 2020). A technically desirable sound, then, does not tell us on its own how it was produced, whether the coordination was efficient, or whether the singer can reproduce it without external guidance.

This distinction matters because recent AI systems do more than recognize features; they increasingly evaluate and recommend. Machine-learning studies have identified paralinguistic singing attributes and singing techniques (Stadler et al., 2023; Xu et al., 2022), deep-learning models have been applied to the assessment of vocal technique (Hou et al., 2025), and newer systems combine real-time, adaptive, explainable, or multimodal feedback (Deng et al., 2026; Sun et al., 2026). These developments give students and teachers access to information that was previously difficult to capture consistently. At the same time, they make it easy to mistake a strong computational result for a strong educational result.

Feedback research helps clarify the difference. In music performance teaching, information about correctness is only one level of feedback; more consequential feedback helps the learner understand a process and eventually regulate it independently (McPherson et al., 2022). A display that marks a note as flat may be useful, but it does not yet explain why the problem recurs or what the singer should change in the next attempt. That step involves pedagogical judgment. It also involves metacognition, because effective practice requires musicians to plan, monitor, evaluate, and revise their own work rather than simply wait for correction (Concina, 2019). AI can support that process, but poorly framed feedback may also encourage students to defer too readily to the system.

The same caution applies to human control and, more fundamentally, to the validity of the reference norm itself. Research on human-centered educational AI repeatedly identifies transparency, reliability, stakeholder participation, and retained human decision-making as core design concerns (Alfredo et al., 2024). In singing, however, cultural-stylistic validity is also a measurement-validity issue: a model can be computationally consistent yet pedagogically invalid if it judges performance against a norm that does not belong to the repertoire or tradition being evaluated. [5]

Cross-cultural music-information-retrieval research shows that computational representations developed within dominant musical systems do not transfer neutrally across traditions (Papaioannou et al., 2026). Turkish makam music provides a concrete example. Its pitch organization is microtonal and makam-dependent, and computational research has documented the limitations of reducing its pitch behavior to conventional Western equal-tempered representations (Bozkurt et al., 2014). Intonation, ornamentation, timbre, articulation, and vibrato therefore cannot be interpreted through a universal numerical norm. An AI-detected "deviation" is pedagogically meaningful only if the target against which it is calculated is itself musically valid.

Broader reviews have addressed AI in music education and contemporary competencies in vocal pedagogy (Li et al., 2026; Mazlan et al., 2026), but the pedagogical roles of AI in voice education remain scattered across music education, audio engineering, machine learning, human-computer interaction, and educational psychology.

What is missing is a synthesis that keeps the technical and pedagogical questions in view at the same time: what can the system detect, and what must still happen before that output contributes to learning?

This review addresses that gap in two ways. First, it brings together accessible academic evidence on AI-supported vocal pedagogy, with attention to technological development, target vocal dimensions, feedback, pedagogical function, and unresolved research problems. Second, it proposes a compact framework for tracing the route from technical analysis to pedagogical use. The guiding question is therefore straightforward: How is AI currently being used in vocal pedagogy, and under what conditions does computational evidence become meaningful educational support?

Method

Review Design and Source Selection

The study used a structured critical-review design. This approach was selected because the relevant literature is dispersed across several disciplines and because the aim was interpretive as well as descriptive: to compare assumptions, connect findings, and build a pedagogical account of the field rather than calculate a pooled intervention effect (Grant & Booth, 2009; Snyder, 2019). The review does not claim exhaustive systematic coverage. Its emphasis is on full-text academic evidence that could be verified and that bears directly on AI-supported vocal teaching and learning.

Searches and citation chaining were conducted through accessible academic discovery environments and publisher platforms up to September 21, 2026. Search terms combined artificial intelligence and machine learning with singing, vocal pedagogy, vocal training, feedback, assessment, and education.

A study was retained when it concerned human singing or vocal pedagogy, used a learning-based computational approach such as machine learning, deep learning, generative AI, explainable AI, or adaptive recommendation, and gave that technology a direct role in teaching, practice, feedback, assessment, or learner support. Studies concerned only with clinical voice diagnosis, speech-language therapy, voice cloning, singing-voice synthesis, source separation, singer identification, or technical audio processing without a pedağogical application were excluded.

The review was limited to English-language academic publications for which both the full text and publication status could be checked. Retracted papers were excluded. Fifteen studies published between 2022 and 2026 formed the final analytical corpus. For each study, the analysis recorded the technological approach, the vocal or learning dimension addressed, the form of feedback, the relationship between learner and teacher, and the pedagogical role assigned to the system.

Analytical Procedure

The analysis proceeded in two passes. A descriptive comparison first mapped the technologies, target skills, feedback formats, and study designs.

A second, theory-informed reading examined what the systems were actually doing pedagogically. Five provisional functions - diagnostic, corrective, adaptive, reflective, and evaluative - were used as sensitizing categories rather than fixed labels. Because one system could serve several purposes, categories were allowed to overlap. Particular attention was paid to the boundary between technical operations, such as feature extraction or scoring, and pedagogical processes, such as interpretation, practice decisions, and self-evaluation. The Vocal-AI Pedagogical Function Framework (VAPFF) was refined through this comparison.

Results

Profile and Technological Development of the Literature

The 15 included studies span 2022-2026, but the literature is weighted heavily toward the most recent period: 10 studies appeared in 2025 or 2026. The geographical picture is similarly uneven. Most studies were conducted in Chinese institutions or educational settings. The corpus includes system-development papers, classification studies, platform evaluations, and controlled or quasi-experimental interventions, suggesting a field that is growing quickly but has not yet diversified methodologically or geographically to the same degree.

The technological trajectory is easier to see when the studies are read chronologically. Earlier work centered on AI-supported teaching platforms, supervised machine learning, technique recognition, and online learning environments (Stadler et al., 2023; Wang, 2022; Xu et al., 2022; Zhang & Zhang, 2024).

Later studies introduced deep convolutional assessment, LLM-supported dialogue, intelligent recommendation, explainable AI, voiceprint-based guidance, and multimodal systems that incorporate respiration, facial expression, or semantic information (Deng et al., 2026; Hou et al., 2025; Li et al., 2025; Su & Guo, 2026; Sun et al., 2026; Wei, 2026). In short, the research problem is shifting from whether AI can recognize vocal output to whether it can adapt its response and participate meaningfully in the learning process.

Table 1
Included studies and primary pedagogical emphasis

Study	Technology	Primary focus	Pedagogical function
Wang (2022)	AI teaching platform	Monitoring and personalized support	Adaptive; evaluative
Xu et al. (2022)	Supervised ML	Paralinguistic singing attributes	Diagnostic; evaluative
Stadler et al. (2023)	ML technique recognition	Singing-technique perception	Diagnostic; evaluative
Zhang & Zhang (2024)	AI-supported distance learning	Performance and self/peer evaluation	Adaptive; reflective; evaluative
T. Zhang (2024)	Smart vocal classroom	Real-time comparison and personalized paths	Diagnostic; corrective; adaptive
Chen (2025)	Computer-assisted AI training	Pitch, range, expression	Corrective; adaptive; evaluative
Y. Li (2025)	Intelligent performance matching	Targeted performance feedback	Diagnostic; corrective; evaluative
W. Li et al. (2025)	LLM feedback + reflection	Metacognition and singing performance	Reflective; corrective
Hou et al. (2025)	Deep CNN	Bel canto technique assessment	Diagnostic; evaluative
Su & Guo (2026)	Adaptive recommendation	Personalized instructional strategy	Adaptive; corrective
Ao & Ao (2026)	Automatic pronunciation scoring	Pronunciation and acoustic quality	Diagnostic; corrective; evaluative
Sun et al. (2026)	Explainable AI scaffolding	Cognitive load and embodied expression	Corrective; reflective; adaptive
Wu & Wen (2026)	Real-time precision feedback	Multidimensional vocal outcomes	Corrective; evaluative
Deng et al. (2026)	Multimodal MoE	Breath, expression, semantics	Diagnostic; adaptive; corrective
Wei (2026)	Voiceprint-based ML	Quantitative evaluation and guidance	Diagnostic; evaluative; adaptive

Vocal Dimensions and Feedback

The literature is still most comfortable with vocal features that can be represented computationally. Pitch and intonation, timing, timbre, pronunciation, spectral characteristics, and vocal-range behavior recur across the corpus. Recent studies have widened the lens to include emotional expression, cognitive load, breathing, facial expression, and embodied performance (Chen, 2025; Deng et al., 2026; Sun et al., 2026). Even so, expressive and aesthetic qualities are usually modeled less directly than acoustic variables and often depend on human ratings or composite measures. This imbalance is important because it marks the point at which technical measurement becomes increasingly dependent on interpretation.

Feedback formats also vary substantially. Some systems return scores or classifications; others offer real-time corrections, audio-visual comparison, personalized recommendations, explanatory scaffolds, or dialogue. The more pedagogically ambitious studies move beyond telling the learner what happened and try to support sense-making. Li et al. (2025), for example, combined audio comparison with LLM-supported dialogue and reflective journaling, while Sun et al. (2026) used explainable AI to make system guidance easier to interpret. By contrast, several technically oriented studies stopped at classification or scoring accuracy and did not investigate how learners understood the output or what they did with it afterward.

Pedagogical Functions and Learning Outcomes

Across the corpus, diagnostic and evaluative uses of AI were the most established. Systems commonly identified vocal attributes, classified technique, quantified deviation, or produced performance scores. Corrective and adaptive functions became more visible in later work through real-time guidance, personalized learning paths, and recommendation systems. Reflective uses were less common. They appeared most clearly when students compared performances, engaged in self-assessment, or were asked to reflect through dialogue or writing. This distribution suggests that the field has progressed further in automating judgment than in studying how learners appropriate that judgment for their own practice.

The intervention studies make the same point from another angle. Chen (2025) reported stronger performance gains with computer-assisted training, whereas Li et al. (2025) found that AI-assisted feedback improved metacognition without producing a statistically superior change in singing performance relative to conventional instruction. Wu and Wen (2026) linked precision feedback to learning outcomes through perceived skill mastery and self-efficacy, and Sun et al. (2026) found that explainable scaffolding could reduce cognitive load while supporting embodied expression. These are not interchangeable outcomes. Performance, metacognition, self-efficacy, cognitive load, and learner autonomy describe different parts of the learning process and should not be collapsed into a single claim that AI 'works.'

Human Mediation and Cultural-Stylistic Validity

Teacher involvement remained visible in much of the corpus. In most studies, AI supplied evidence, comparison, or individualized guidance rather than replacing instruction altogether. What was rarely examined directly was the negotiation that follows: whether learners or teachers accepted the system's judgment, questioned it, overrode it, or used it selectively. Human mediation was therefore present in practice more often than it was treated as an explicit research variable.

Cultural-stylistic validity was even less developed. The studies covered Bel Canto, classical and popular singing, Chinese and European art song, and multilingual settings, but none directly tested whether AI judgments remained valid across fundamentally different intonational systems or microtonal traditions. Repertoire variety should therefore not be mistaken for cross-cultural validation. The current evidence says little about what happens when the reference model itself belongs to a different musical logic from the repertoire being taught.

Table 2
Vocal-AI Pedagogical Function Framework (VAPFF)

Stage	Analytical meaning
1. Vocal/multimodal input	Audio, video, sensor, text, or combined performance evidence
2. AI detection and analysis	Feature extraction and state detection
3. Computational judgment	Classification, scoring, comparison, prediction, or error detection
4. Feedback representation	Visual, numerical, auditory, textual, dialogic, or multimodal output
5. Human interpretation/mediation	Learner and/or teacher makes sense of the output
6. Pedagogical action	Practice, correction, reflection, strategy change, or teacher intervention
7. Learning function	Diagnostic, corrective, adaptive, reflective, and/or evaluative role
8. Cross-cutting dimensions	Human agency; pedagogical validity; cultural-stylistic validity; responsible AI

Illustrative VAPFF Walkthrough

Consider a student singing a tonal exercise in which several ascending notes are consistently below the intended target. At the input and detection stages, the system receives the audio signal, extracts fundamental-frequency information, and identifies a recurring negative pitch deviation. It may then encode that result as a trace, score, or warning. Up to this point, the process is computational. At the human-interpretation stage, the teacher and learner must decide what the pattern means: whether it reflects an unstable auditory target, a coordination issue around a register or vowel change, fatigue, or simply an inappropriate reference. The pedagogical response may involve reducing the phrase, clarifying the intended target, repeating the passage, and comparing successive attempts. Only when the output changes practice and supports later self-monitoring does the system move from measurement toward corrective or reflective learning. The same numerical deviation could require the opposite decision in a culturally specific repertoire. If an equal-tempered reference is applied to an intended microtonal pitch in Turkish makam music, the appropriate pedagogical action may be to reject or recalibrate the system's judgment rather than "correct" the singer. This example illustrates why VAPFF places human mediation and cultural-stylistic validity between computational judgment and pedagogical action.

Discussion

From Measurement to Pedagogy

The most important conclusion from the reviewed studies is that computational capability and pedagogical value are related but not equivalent. The field can now extract, classify, and score a growing range of vocal information. That is useful, but learning begins at a different point: when the learner or teacher can make sense of the information and decide what to do next. This distinction is consistent with research on music-performance feedback, where the pedagogical value of feedback increases as it moves from simple task correction toward process understanding and self-regulation (McPherson et al., 2022). A precise score can inform a lesson; it does not conduct the lesson by itself.

Embodied Learning, Metacognition, and the Role of the Teacher

Voice education makes this boundary particularly visible because singing is embodied. Acoustic output does not fully reveal respiratory coordination, effort, sensation, or the strategy a singer is using (Bremmer & Nijs, 2020; Zarate, 2013). The recent interest in multimodal systems is therefore well founded: combining sound with information about breathing, gesture, or expression may offer a richer account of performance. Still, additional data are only useful if they are pedagogically interpretable. A beginner and an advanced singer may need different explanations of the same measured deviation, and a technically relevant signal may be irrelevant to the immediate instructional goal.

Metacognition is what connects external evidence with increasingly independent practice. Students of singing eventually need to notice patterns, evaluate attempts, and adjust strategies without waiting for constant external correction (Concina, 2019). The findings of Li et al. (2025) are instructive here. Their AI-supported condition strengthened metacognitive development even though it did not produce a statistically superior short-term gain in singing performance. That result is not a failure of AI; it is a reminder that reflection and performance are different outcomes. The design question is therefore not how much feedback a system can generate, but whether its feedback helps the student become more capable of listening, judging, and deciding.

The corpus also points toward augmentation rather than replacement. Human-centered AI research treats retained human control, transparency, and stakeholder participation as basic conditions of responsible educational technology (Alfredo et al., 2024).

Those conditions are especially relevant in vocal pedagogy. The same measured feature may call for different responses depending on repertoire, developmental stage, technical intention, or style. Expert mediation is not an inconvenience to be engineered away; it is part of the pedagogical process. VAPFF therefore places human interpretation between the system's representation of feedback and the action that follows.

Cultural-Stylistic Validity

Cultural-stylistic validity should be treated as a measurement-validity problem, not as an optional diversity add-on or a future refinement. A score is only as valid as the reference model against which performance is judged. Cross-cultural music-information-retrieval research has shown that computational representations do not perform uniformly across musical traditions (Papaioannou et al., 2026). In singing, the implications are immediate: intonation, ornamentation, vibrato, timbre, and articulation are shaped by stylistic and cultural conventions, so an algorithmically detected deviation cannot automatically be treated as a pedagogical error.

The present corpus offers almost no direct evidence concerning microtonal or culturally specific vocal systems. This is particularly consequential for traditions such as Turkish makam music, where pitch organization and melodic behavior are makam-dependent and where computational studies have had to adopt representations finer than conventional twelve-tone equal temperament (Bozkurt et al., 2014). The issue is therefore not simply whether future systems become more culturally inclusive; it is whether the measurement target itself is valid for the repertoire being assessed. Repertoire diversity should not be confused with cross-cultural validation. AI assessment models should be validated within the musical systems in which they are intended to operate, with their reference assumptions made explicit to teachers and learners.

Implications for Vocal Educators and AI Design

For teachers, the practical issue is not whether a commercial AI tool contains many features, but whether its judgments can be pedagogically inspected and used. Before adopting an application, educators should ask what vocal variable is being measured, which reference model defines correctness, whether the output explains a pattern or merely produces a score, whether the teacher or learner can question or override the system, whether stylistic limitations are declared, and how voice data are handled. Table 3 translates these questions into a concise vetting guide. In general, transparent, process-oriented tools are easier to integrate responsibly than opaque applications that compress vocal quality into a single unexplained metric.

Table 3
Practical criteria for vetting AI-supported vocal tools

Criterion	Transparent / process-oriented	Opaque / warning sign
Target construct	Names the vocal feature being assessed and its pedagogical purpose.	Uses a global “singing quality” score without defining what it represents.
Reference norm	States the target, model, tuning, repertoire, or comparison basis.	Treats one reference norm as universally correct or leaves the norm unspecified.
Feedback quality	Shows patterns and provides information that can support a next practice decision.	Returns only a grade, color, or ranking with no interpretable process information.
Human control	Allows teachers or learners to inspect, contextualize, customize, or override feedback.	Presents system judgment as authoritative and non-negotiable.
Stylistic validity	Declares repertoire/style limits or permits context-sensitive targets.	Makes broad claims across styles without evidence of cross-cultural validation.
Data governance	Explains recording storage, privacy, consent, and data use.	Provides unclear or inaccessible information about how voice data are retained or reused.

These criteria do not imply that a simple scoring tool is unusable. A single metric can support motivation, monitoring, or between-lesson practice when it is treated as a limited indicator. The risk arises when the system is granted authority beyond the construct it actually measures. In VAPFF terms, a pedagogically stronger tool makes enough of the pathway from detection to feedback visible for teachers and learners to make a defensible decision about what should happen next.

Theoretical Contribution and Limitations

VAPFF is intended to make that division of labor analytically explicit. It separates seven stages: vocal or multimodal input; AI detection and analysis; computational judgment; feedback representation; human interpretation or mediation; pedagogical action; and learning function. Diagnostic, corrective, adaptive, reflective, and evaluative roles can occur at different points in this chain, while human agency, pedagogical validity, cultural-stylistic validity, and responsible AI cut across the whole process. The framework is deliberately simple. Its purpose is not to claim a new grand theory of vocal learning, but to stop technical success at one stage from being treated as evidence of success at another.

That distinction also has methodological value. A study that reports high classification accuracy has demonstrated something important about computational performance, but it has not yet shown that students can interpret the result, change their practice, or retain the resulting skill. Likewise, an adaptive system should not be called pedagogically personalized merely because an algorithm changes content; the adaptation needs to respond to a defensible learning need. By naming these stages separately, VAPFF offers voice educators, music-education researchers, and AI developers a common vocabulary for reporting what enters the system, what judgment is produced, how the output is communicated, who retains decision authority, and which learning outcome is actually measured.

This review also has clear limits. It includes only English-language academic studies whose full text and publication status could be verified, and it uses a critical-review rather than an exhaustive systematic-review design. The resulting corpus is analytically focused, geographically concentrated, and methodologically heterogeneous. These constraints make broad generalization inappropriate. They also point directly to the next research priorities: longitudinal work, comparisons among AI-only, teacher-only, and teacher-mediated conditions, measures of retention and transfer, and validation across languages, repertoires, and vocal traditions.

Conclusion

Artificial intelligence is changing the informational environment in which singing is taught and practiced. The reviewed studies show a clear movement from recognition and scoring toward systems that are adaptive, explainable, dialogic, and multimodal. That development matters because it expands the kinds of evidence that can be brought into a voice lesson or an individual practice session.

The more important shift, however, is conceptual. Vocal performance cannot be reduced to acoustic output, and information generated by AI is not automatically pedagogical feedback. Its educational value depends on interpretation, context, and the action that follows. Used well, AI can strengthen observation and reflection; used uncritically, it can turn a provisional computational judgment into a false pedagogical certainty. The appropriate goal is therefore not to transfer authority from teacher to system, but to use computational evidence in ways that improve the quality of human judgment.

VAPFF offers one way to examine that process without conflating its stages. It also makes the field's unresolved questions visible: how learner autonomy can be supported, how teacher mediation should be preserved, and how pedagogical and cultural-stylistic validity can be established across different vocal traditions. The next generation of research should therefore ask less often whether AI can evaluate singing and more precisely when, for whom, and under what musical conditions AI-generated evidence contributes to durable vocal learning.

Declarations

Ethical Approval: This review used published academic literature and involved no human participants; ethics committee approval was not required.

Conflict of Interest: The author serves on the Advisory Board of Müzikist Academic Journal. The author should have no role in the editorial handling or peer-review process of this manuscript.

References

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence, 6*, 100215. https://doi.org/10.1016/j.caeai.2024.100215

Ao, Y., & Ao, W. (2026). A method for evaluating and improving vocal pronunciation quality of vocal music students based on artificial intelligence technology. *International Journal of Web-Based Learning and Teaching Technologies, 21(1)*, 1-22. https://doi.org/10.4018/IJWLTT.402021

Bozkurt, B., Ayangil, R., & Holzapfel, A. (2014). Computational analysis of Turkish makam music: Review of state-of-the-art and challenges. *Journal of New Music Research, 43*(1), 3-23. https://doi.org/10.1080/09298215.2013.865760

Bremmer, M., & Nijs, L. (2020). The role of the body in instrumental and vocal music pedagogy: A dynamical systems theory perspective on the music teacher's bodily engagement in teaching and learning. *Frontiers in Education, 5*, 79. https://doi.org/10.3389/feduc.2020.00079

Chen, L. (2025). Enhancing vocal performance through computer-assisted training. International Journal of Web-Based *Learning and Teaching Technologies, 20*(1), 1-17. https://doi.org/10.4018/IJWLTT.381310

Concina, E. (2019). The role of metacognitive skills in music learning and performing: Theoretical features and educational implications. *Frontiers in Psychology, 10*, 1583. https://doi.org/10.3389/fpsyg.2019.01583

Deng, M., Liu, C., & Yang, Z. (2026). MoE-based multimodal alignment for vocal metaverse classrooms. *Scientific Reports, 16*, 17906. https://doi.org/10.1038/s41598-026-47778-y

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91-108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Hou, Z., Zhao, X., Jiang, S., Luo, D., Sheng, X., Geng, K., Ye, K., Xia, J., Zhang, Y., Ban, C., Chen, J., Zou, Y., Feng, Y., Yuan, X., & Fan, G. (2025). Dense dynamic convolutional network for Bel canto vocal technique assessment. *Scientific Reports, 15*, 15666. https://doi.org/10.1038/s41598-025-98726-1

Li, W., Cui, X., Manoharan, P., Dai, L., Liu, K., & Huang, L. (2025). AI-assisted feedback and reflection in vocal music training: Effects on metacognition and singing performance. *Frontiers in Psychology, 16*, 1598867. https://doi.org/10.3389/fpsyg.2025.1598867

Li, W., Mazlan, C. A. N., & Safian, A. R. (2026). 21st-century skills in vocal pedagogy: A scoping review. *Frontiers in Education, 11*, 1744952. https://doi.org/10.3389/feduc.2026.1744952

Li, Y. (2025). Diversified intelligent teaching methods and their impact on vocal music instruction. *International Journal of Web-Based Learning and Teaching Technologies, 20*(1), 1-19. https://doi.org/10.4018/IJWLTT.381090

Li, Y. (2026). Beyond output metrics: Reframing AI-assisted vocal pedagogy through human learning and educational value. *Frontiers in Psychology, 17*, 1844841. https://doi.org/10.3389/fpsyg.2026.1844841

Mazlan, C. A. N., Hanafi, H. F., Sarifin, M. R., & Md Noor, A. R. (2026). Artificial intelligence applications and pedagogical challenges in music education. *Discover Education, 5*, 140. https://doi.org/10.1007/s44217-026-01127-3

McPherson, G. E., Blackwell, J., & Hattie, J. (2022). Feedback in music performance teaching. *Frontiers in Psychology, 13*, 891025. https://doi.org/10.3389/fpsyg.2022.891025

Papaioannou, C., Benetos, E., & Potamianos, A. (2026). Cross-cultural music similarity: Bridging human perception, signal processing, and foundation models. *Transactions of the International Society for Music Information Retrieval, 9*(1), 347-368. https://doi.org/10.5334/tismir.341

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039

Stadler, A., Parada-Cabaleiro, E., & Schedl, M. (2023). Towards potential applications of machine learning in computer-assisted vocal training. In *Proceedings of the 16th International Symposium on Computer Music Multidisciplinary Research* (pp. 430-441). Zenodo. https://doi.org/10.5281/zenodo.10125347

Su, F., & Guo, Q. (2026). The optimization of vocal music teaching by integrating the STEAM concept with the intelligent recommendation system. *Scientific Reports, 16*, 739. https://doi.org/10.1038/s41598-025-30288-8

Sun, Y., Wang, X., & Wei, X. (2026). Can explainable AI scaffolding reduce cognitive load and enhance embodied expression? A quasi-experimental study of multilingual vocal learning in Bel Canto education. *Frontiers in Psychology, 17*, 1835027. https://doi.org/10.3389/fpsyg.2026.1835027

Wang, X. (2022). Design of vocal music teaching system platform for music majors based on artificial intelligence. *Wireless Communications and Mobile Computing, 2022*, 5503834. https://doi.org/10.1155/2022/5503834

Wei, Y. Z. (2026). Optimization of vocal teaching empowered by AI: A comprehensive study on quantitative evaluation and personalized guidance system for singing skills based on voiceprint feature extraction. *Advanced Electromagnetics, 15*(3), 3394-3403. https://doi.org/10.7716/aem.v15i3.3406

Wu, R., & Wen, J. (2026). The impact of precise feedback in AI vocal courses on learning outcomes: A chain mediation effect of skill mastery and self-efficacy. *Frontiers in Psychology, 17*, 1811130. https://doi.org/10.3389/fpsyg.2026.1811130

Xu, Y., Wang, W., Cui, H., Xu, M., & Li, M. (2022). Paralinguistic singing attribute recognition using supervised machine learning for describing the classical tenor solo singing voice in vocal pedagogy. *EURASIP Journal on Audio, Speech, and Music Processing, 2022*, 8. https://doi.org/10.1186/s13636-022-00240-z

Zarate, J. M. (2013). The neural control of singing. *Frontiers in Human Neuroscience, 7*, 237. https://doi.org/10.3389/fnhum.2013.00237

Zhang, T. (2024). Research on online vocal music smart classroom-assisted teaching based on wireless network combined with artificial intelligence. *International Journal of Web-Based Learning and Teaching Technologies, 19*(1), 1-19. https://doi.org/10.4018/IJWLTT.353156

Zhang, X., & Zhang, J. (2024). A new model of vocal music teaching in the context of Internet distance learning. *International Journal of Web-Based Learning and Teaching Technologies, 19*(1), 1-12. https://doi.org/10.4018/IJWLTT.348336