TRIBUNAL EXAMINADOR

Este proyecto fue aprobado por el Tribunal Examinador de la carrera: Maestría Profesional en Ingeniería del Software con Énfasis en Inteligencia Artificial, requisito para optar por el título de grado de Maestría Profesional en Ingeniería del Software con Énfasis en Inteligencia Artificial, el estudiante Elmer Mondragón Prieto.

DANIEL ENRIQUE FLORES CORDERO (FIRMA)
Firmado digitalmente por DANIEL ENRIQUE FLORES CORDERO (FIRMA)
Fecha: 2026.10.27 14:17:04 -06'00'

MBA Daniel Enrique Flores Cordero
Tutor



Digitally signed by MIGUEL PEREZ MONTERO (FIRMA)
Date: 2026.10.29 18:57:16 -06'00'

M.Sc. Miguel Pérez Montero
Lector 1

MARLON ALBERTO DURAN ALFARO (FIRMA)
PERSONA FISICA, CPF-01-0754-0182.
Fecha declarada: 29/10/2026 07:11:18 PM
Razón: Aprobador
Lugar: San Jose Costa Rica Contacto: Marlon Durán Alfaro

M.Sc. Marlon Durán Alfaro
Lector 2

San José, Costa Rica, 16 de Octubre 2026

Firmada digitalmente, de conformidad con la Ley de Certificados, Firmas Digitales y Documentos Electrónicos N° 8454, destacando el artículo 9°-

Marco evaluativo de sistemas de IA musical para accesibilidad y autonomía en discapacidad motora (2021-2026)

Elmer Mondragón Prieto
Maestría Profesional en Ingeniería del Software con énfasis en IA Aplicada
Universidad CENFOTEC, Costa Rica

Resumen—Los sistemas de inteligencia artificial (IA) musical pueden ampliar las formas de interacción y participación de personas con discapacidad motora; sin embargo, la literatura suele evaluarlos mediante criterios fragmentados, concentrados en novedad técnica, usabilidad aislada o calidad del resultado. Este artículo presenta un marco evaluativo para analizar y comparar dichos sistemas desde una perspectiva integral y trazable. Se realizó una revisión sistemática de literatura publicada entre 2021 y 2026, complementada con la integración de ISO/IEC 25010:2023, ISO/IEC 25059:2023, WCAG 2.2, CIF/ICF y NIST AI RMF. El marco resultante organiza la evaluación en seis dimensiones: interacción inclusiva y accesibilidad; confiabilidad y predictibilidad; seguridad y riesgo; interpretabilidad y transparencia; mantenibilidad y sostenibilidad; y contexto funcional y resultados. La propuesta emplea una escala ordinal de cuatro niveles basada en suficiencia de evidencia e incorpora la categoría N/A. Una aplicación ilustrativa a un caso documentado mostró capacidad inicial para organizar evidencia heterogénea, identificar fortalezas y vacíos, y construir perfiles por dimensión. El aporte principal es metodológico: una estructura común para valorar cuándo la IA mejora efectivamente la accesibilidad y la autonomía y cuándo introduce complejidad, opacidad o dependencia adicional.

Palabras clave—inteligencia artificial musical, accesibilidad, discapacidad motora, autonomía, calidad de software, marco evaluativo.

Abstract—Music-oriented artificial intelligence (AI) systems may expand interaction and participation opportunities for people with motor disabilities. Nevertheless, existing studies often assess them through fragmented criteria focused on technical novelty, isolated usability, or output quality. This paper presents an evaluative framework for the structured and traceable analysis of such systems. A systematic literature review covering 2021–2026 was combined with ISO/IEC 25010:2023, ISO/IEC 25059:2023, WCAG 2.2, the ICF, and the NIST AI RMF. The resulting framework comprises six dimensions: inclusive interaction and accessibility; reliability and predictability; safety and risk; interpretability and transparency; maintainability and sustainability; and functional context and outcomes. It uses a four-level ordinal evidence scale and a not-applicable category. An illustrative application to a documented case showed initial capacity to organize heterogeneous evidence, identify strengths and gaps, and build dimension-level profiles. The main contribution is methodological: a common structure to determine when AI effectively improves accessibility and autonomy and when it introduces complexity, opacity, or additional dependence.

Keywords—musical artificial intelligence, accessibility, motor disabilities, autonomy, software quality, evaluative framework.

1. Introducción

La incorporación de IA en herramientas musicales ha favorecido funciones de asistencia, automatización, adaptación y nuevas modalidades de control. Estas posibilidades son especialmente relevantes para personas con discapacidad motora, quienes pueden enfrentar barreras al utilizar instrumentos, aplicaciones o entornos de producción diseñados para formas tradicionales de interacción. No obstante, la presencia de IA no garantiza por sí sola una mejora en accesibilidad o autonomía. Un componente inteligente puede ofrecer adaptación útil, pero también introducir opacidad, errores, dependencia de calibración o dificultades de mantenimiento.

El problema central identificado es la ausencia de una estructura común para analizar el sistema completo. Muchos trabajos destacan desempeño técnico, experiencia de usuario o viabilidad del prototipo, pero no integran simultáneamente accesibilidad en tareas musicales reales, confiabilidad, riesgo, transparencia, sostenibilidad y resultados funcionales. Esta fragmentación dificulta comparar soluciones con IA frente a alternativas sin IA y establecer si la innovación tecnológica reduce barreras de manera verificable.

La pregunta que orienta el estudio es: ¿es posible diseñar un marco evaluativo para analizar y comparar sistemas de IA musical orientados a la accesibilidad y autonomía de personas con discapacidad motora, con base en evidencia publicada entre 2021 y 2026? El objetivo fue diseñar dicho marco mediante una revisión sistemática, la integración de estándares y la operacionalización de criterios en una rúbrica aplicable a casos documentados.

2. Fundamentos y trabajos relacionados

2.1 Sistema de IA musical como solución completa

El objeto de evaluación no es únicamente el modelo de IA. Se considera una solución sociotécnica compuesta por interfaz, modalidad de interacción, procesamiento musical, componente inteligente cuando exista, gestión de errores, contexto de uso y resultados funcionales. Esta perspectiva se alinea con ISO/IEC 25010:2023, que permite analizar calidad del producto de software, y con ISO/IEC 25059:2023, que añade propiedades específicas de los sistemas con IA [1], [2].

2.2 Accesibilidad, funcionamiento y autonomía

WCAG 2.2 aporta criterios verificables de percepción, operación, comprensión y robustez para interfaces digitales [3]. Sin embargo, en el dominio musical la accesibilidad debe expresarse en tareas reales: interpretar, colaborar, improvisar, producir, mezclar o controlar parámetros. La CIF/ICF permite relacionar estas tareas con actividad, participación, facilitadores y barreras del entorno [4]. En este artículo, autonomía significa realizar actividades musicales significativas con menor dependencia de apoyo externo.

2.3 Calidad, transparencia y riesgo

ISO/IEC 25010:2023 fundamenta confiabilidad, eficiencia, compatibilidad y mantenibilidad [1]. ISO/IEC 25059:2023 incorpora interpretabilidad, robustez y comportamiento bajo incertidumbre [2]. NIST AI RMF estructura el análisis de riesgos mediante las funciones GOVERN, MAP, MEASURE y MANAGE [5]. La literatura de XAI inclusiva también advierte que las explicaciones deben adaptarse a las necesidades de las personas usuarias y apoyar control informado, no limitarse a justificar el algoritmo ante especialistas [6], [7].

2.4 Patrones de diseño y evaluación identificados

La evidencia revisada muestra que las soluciones musicales accesibles no constituyen una categoría tecnológica única. Se identificaron interfaces digitales colaborativas, instrumentos gestuales, sistemas inmersivos, mecanismos de retroalimentación háptica o vocal, control mediante sensores y propuestas basadas en

bioseñales o interfaces cerebro-computadora. Esta diversidad confirma que la evaluación debe abstraerse de la tecnología específica y concentrarse en propiedades comunes del sistema: adecuación de la interacción, estabilidad, control, comprensión, continuidad de uso y resultados funcionales. Los trabajos sobre DuoRhythmo, improvisación inclusiva e interfaces colaborativas muestran que pequeñas adaptaciones de interacción pueden producir mejoras relevantes sin requerir necesariamente componentes inteligentes complejos [9], [10], [15].

Un segundo patrón es la importancia del contexto. La misma modalidad de control puede ser adecuada para una tarea y problemática para otra. Por ejemplo, una interfaz que reduce movimientos finos puede facilitar la ejecución, pero resultar insuficiente para tareas de mezcla que requieren acceso preciso a parámetros. De manera semejante, las tecnologías inmersivas o basadas en gestos amplían las posibilidades expresivas, aunque introducen exigencias de calibración, reconocimiento y latencia. Por ello, el marco no valora la tecnología de forma aislada, sino su correspondencia con la tarea musical, las condiciones funcionales de la persona y el entorno de uso.

Un tercer patrón es la desigualdad de la evidencia. Los estudios centrados en accesibilidad suelen documentar experiencia, aceptación y barreras observadas, mientras que los trabajos técnicos reportan precisión, estabilidad o desempeño de componentes. Con menor frecuencia se describen riesgos, mecanismos de recuperación, mantenibilidad, trazabilidad o comprensión del comportamiento automático. Esta asimetría explica por qué una solución puede parecer sólida en una dimensión y permanecer prácticamente desconocida en otras. El marco propuesto conserva esos vacíos como resultados explícitos, en lugar de tratarlos como cumplimiento implícito.

2.5 Brecha metodológica y contribución esperada

La revisión sugiere que el problema del campo no es la ausencia de propuestas, sino la falta de un lenguaje común para evaluarlas. Los sistemas se describen mediante métricas y conceptos difíciles de comparar, y la incorporación de IA suele analizarse desde el desempeño del componente antes que desde su efecto sobre la autonomía. La contribución esperada es, por tanto, un instrumento que conecte estándares de calidad, accesibilidad, funcionamiento humano y riesgo con evidencias observables en casos musicales. La intención no es sustituir pruebas con usuarios ni certificar productos, sino mejorar la consistencia del análisis documental y orientar futuras evaluaciones empíricas.

3. Metodología

La investigación fue aplicada, cualitativa-documental y descriptivo-analítica, con un componente propositivo. El procedimiento se organizó en cinco etapas: (1) revisión sistemática y síntesis de evidencia; (2) derivación de dimensiones conceptuales; (3) operacionalización en criterios e indicadores; (4) aplicación ilustrativa a un caso documentado; y (5) refinamiento del marco.

3.1 Revisión sistemática

La búsqueda cubrió publicaciones de 2021 a 2026 en Google Scholar y ACM Digital Library, siguiendo un protocolo inspirado en Biolchini et al. [8]. Las cadenas combinaron términos sobre interacción musical, IA, discapacidad motora, accesibilidad, usabilidad y evaluación. Se incluyeron sistemas con IA, alternativas comparables sin IA y fuentes metodológicas transferibles. Los estudios de otras discapacidades se conservaron únicamente cuando aportaban criterios aplicables a accesibilidad, interpretabilidad, riesgo, trazabilidad o resultados funcionales.

Cada fuente se registró en una matriz con objetivo, método, muestra o contexto, métricas, hallazgos, limitaciones y aporte a la rúbrica. La síntesis clasificó los sistemas por modalidad —interfaces digitales accesibles, XR, háptica, sensores, bioseñales y BCI— y comparó patrones, brechas y compensaciones entre accesibilidad, autonomía y complejidad tecnológica.

3.2 Construcción del instrumento

Los conceptos de los estándares y de la literatura se transformaron en dimensiones, criterios, indicadores observables, evidencia esperada y reglas de valoración. La trazabilidad se preservó vinculando cada puntuación con una descripción, métrica, observación o resultado explícito del caso. Cuando la evidencia no se encontraba reportada se asignó 0; cuando el criterio no correspondía al sistema se utilizó N/A.

3.3 Trazabilidad, reglas de decisión y control de sesgos

La calidad del instrumento depende de que cada valoración pueda rastrearse hasta una evidencia concreta. Por ello, la matriz de extracción registra el objetivo del estudio, el método, la población o contexto, las métricas, los hallazgos, las limitaciones y el aporte potencial a la rúbrica. Posteriormente, cada criterio del marco se vincula con resultados, observaciones o descripciones explícitas de la fuente. Cuando la documentación no permite sustentar el cumplimiento, el criterio se califica como evidencia insuficiente o no reportada, sin completar vacíos mediante inferencias favorables.

La escala distingue cuatro niveles: 3 para evidencia directa, suficiente y verificable; 2 para evidencia verificable pero limitada en alcance o consistencia; 1 para menciones o indicios que no demuestran el cumplimiento; y 0 para ausencia de información verificable. La categoría N/A se reserva para criterios que no corresponden al sistema evaluado y se excluye del cálculo del perfil. Ante duda entre dos niveles se adopta el nivel inferior, regla conservadora que reduce el riesgo de sobrevalorar prototipos con documentación incompleta.

El protocolo también reconoce amenazas de búsqueda, selección, publicación, disponibilidad y extracción. La cobertura se limita a Google Scholar y ACM Digital Library; además, los estudios presentan muestras pequeñas, tareas heterogéneas y distintos niveles de madurez tecnológica. Para mitigar estos riesgos se conservaron cadenas de búsqueda equivalentes, criterios explícitos de inclusión y exclusión, una plantilla única de extracción y una síntesis narrativa que diferencia evidencia empírica, aportes conceptuales y fuentes normativas [8].

3.4 Unidad de análisis y selección de evidencia

La unidad de análisis estuvo constituida por estudios y casos documentados de sistemas de software musical orientados a accesibilidad y autonomía. La selección fue intencional y guiada por criterios: descripción suficiente de la solución, presencia de evidencia de uso o evaluación y posibilidad de relacionar los hallazgos con las dimensiones del marco. Los sistemas con IA y discapacidad motora constituyeron el núcleo principal; las soluciones sin IA y los estudios de otras poblaciones se conservaron únicamente cuando aportaban evidencia comparativa o criterios metodológicamente transferibles.

La aplicación ilustrativa se concentró en DuoRhythmo por su pertinencia, nivel de documentación y correspondencia con la población delimitada. El caso describe una interfaz musical colaborativa accesible para personas con esclerosis lateral amiotrófica y reporta una evaluación remota de experiencia de usuario [9]. Su selección no implica que represente todo el dominio; se utilizó para examinar la claridad operativa de la rúbrica, la trazabilidad de las valoraciones y la capacidad de identificar tanto fortalezas como ausencia de evidencia.

3.5 Operacionalización de dimensiones e indicadores

La operacionalización siguió una secuencia de cuatro pasos. Primero, cada estándar o marco conceptual se tradujo en propiedades relevantes para el problema. Segundo, esas propiedades se formularon como criterios evaluables. Tercero, cada criterio se

asoció con un indicador observable y con tipos de evidencia esperada. Cuarto, se definieron reglas para asignar niveles sin confundir ausencia de reporte con incumplimiento. Por ejemplo, la confiabilidad se tradujo en consistencia de respuesta, estabilidad, latencia y manejo de errores; la autonomía se expresó mediante logro de tareas, control informado y reducción de dependencia.

La escala 0–3 se diseñó para valorar suficiencia de evidencia y no grado de acuerdo. Un valor 3 exige evidencia directa, consistente y verificable; 2 representa cumplimiento respaldado pero limitado; 1 corresponde a una mención o indicio insuficiente; y 0 indica que la fuente no permite valorar el criterio. N/A se utiliza únicamente cuando el criterio no corresponde al sistema. Esta separación es importante porque un artículo puede no documentar un aspecto que sí existe en el producto; el marco registra el vacío documental sin convertirlo automáticamente en una falla técnica.

3.6 Procedimiento de aplicación

El procedimiento inicia con la caracterización del sistema: propósito, población, modalidad de interacción, tarea musical, presencia de IA y tipo de evidencia disponible. Después, los hallazgos se distribuyen entre las seis dimensiones y se asignan valores con una justificación breve. El resultado no se reduce a un promedio global; se construye un perfil por dimensión que permite reconocer compensaciones. Una solución puede ser altamente accesible y, al mismo tiempo, presentar baja trazabilidad o escasa evidencia de sostenibilidad. Esta lectura evita que una fortaleza oculte riesgos relevantes.

Cuando el marco se aplique a dos o más casos, la comparación debe realizarse por dimensión y por tarea equivalente. La interpretación debe considerar el contexto y la calidad de la evidencia, evitando concluir que una diferencia numérica pequeña representa superioridad general. En esta investigación, la aplicación a un único caso se utilizó como prueba inicial del procedimiento y no como validación estadística ni comparación definitiva.

4. Resultados: marco evaluativo propuesto

El resultado principal es un instrumento multicriterio que organiza la evaluación en seis dimensiones complementarias. La Figura 1 resume su estructura.



4.1 Fundamento y alcance de las dimensiones

Las seis dimensiones no operan como compartimentos independientes. La interacción inclusiva determina si la modalidad de control, el flujo y la retroalimentación reducen barreras en tareas musicales reales. La confiabilidad y predictibilidad examinan estabilidad, latencia, robustez y recuperación ante fallos. Seguridad y riesgo registran límites, posibles efectos no deseados y medidas de mitigación. Interpretabilidad y transparencia valoran si los estados, automatizaciones y decisiones son comprensibles y permiten control informado. Mantenibilidad y sostenibilidad consideran dependencia tecnológica, documentación, compatibilidad y continuidad de uso. Finalmente, contexto funcional y resultados relaciona el sistema con actividad, participación y reducción de dependencia conforme a CIF/ICF [1]-[5].

Esta integración evita dos reducciones frecuentes: equiparar accesibilidad con navegación básica y equiparar calidad de IA con precisión del modelo. En un sistema musical accesible, una respuesta técnicamente correcta puede carecer de valor si exige movimientos incompatibles con la persona, introduce latencia que interrumpe la ejecución o elimina control sobre el resultado. Del mismo modo, un prototipo inclusivo en una sesión inicial puede ser poco sostenible si requiere calibración constante, hardware difícil de reemplazar o asistencia permanente. El perfil multidimensional permite visualizar estas compensaciones sin ocultarlas en una puntuación única.

Figura 1. Seis dimensiones del marco evaluativo propuesto.

Dimensión	Propósito principal
Interacción inclusiva y accesibilidad	Valorar adaptación de controles, claridad del flujo, retroalimentación y ejecución de tareas musicales significativas.
Confiabilidad y predictibilidad	Examinar estabilidad, consistencia, latencia, robustez y manejo de errores.
Seguridad y riesgo	Identificar límites, efectos no deseados, sobrecarga, fatiga y medidas de mitigación.
Interpretabilidad y transparencia	Determinar si el comportamiento y los estados del sistema son comprensibles y controlables.
Mantenibilidad y sostenibilidad	Valorar documentación, compatibilidad, dependencia tecnológica y continuidad de uso.
Contexto funcional y resultados	Relacionar el sistema con actividad, participación, reducción de barreras y autonomía.

Tabla 1. Dimensiones y propósito del marco evaluativo.

4.2 Escala de valoración

La rúbrica utiliza una escala ordinal de cuatro niveles orientada a la suficiencia de evidencia: 3 corresponde a evidencia sólida, directa y verificable; 2 a evidencia verificable pero limitada; 1 a una mención o indicio insuficiente; y 0 a ausencia de información verificable. La categoría N/A se utiliza cuando el criterio no corresponde a las características del sistema y se excluye del cálculo. La distinción entre 1 y 2 depende de la existencia de evidencia concreta de cumplimiento. Ante duda entre dos niveles se asigna el inferior, salvo justificación suficiente para el superior.

4.3 Procedimiento de aplicación

La aplicación comprende cinco pasos: caracterización del caso; organización de la evidencia por dimensión; puntuación justificada de criterios; construcción del perfil; e interpretación. Cuando se analizan dos o más sistemas, los perfiles permiten comparar similitudes, diferencias y compensaciones. Para la aplicación inicial de este estudio se utilizó un solo caso documentado y la interpretación se concentró en su perfil individual.

4.4 Producto evaluativo y forma de interpretación

El resultado de la aplicación no debe interpretarse como una certificación ni como una medida clínica. El marco produce un perfil de evidencia por dimensión que muestra fortalezas, debilidades, vacíos y criterios no aplicables. Las puntuaciones permiten ordenar la información, pero la interpretación final debe conservar la justificación cualitativa y las condiciones del caso. Cuando se comparen dos o más sistemas, la comparación debe realizarse criterio por criterio y dentro de tareas musicales equivalentes, evitando sumar valores de contextos funcionales incompatibles.

El procedimiento puede utilizarse en tres momentos: análisis retrospectivo de casos publicados, apoyo a la definición de requisitos durante el diseño y evaluación comparativa previa a una decisión de adopción. En cada modalidad, el evaluador debe declarar la fuente de evidencia, la fecha o versión del sistema, el contexto de uso, la población objetivo y las restricciones conocidas. Esta documentación permite repetir la valoración cuando aparezcan

nuevas versiones o evidencia adicional, reforzando la mantenibilidad del propio proceso evaluativo.

4.1 Caracterización de la base documental

La revisión reunió veinte fuentes organizadas en tres grupos. El primero corresponde al núcleo de interacción musical accesible e incluye interfaces colaborativas, instrumentos gestuales, realidad extendida, retroalimentación háptica, composición asistida y control mediante bioseñales. El segundo reúne trabajos sobre IA responsable, explicabilidad y gestión de riesgo. El tercero aporta soporte metodológico sobre trazabilidad, datasets, sostenibilidad y autonomía. Esta organización permitió conservar la delimitación en discapacidad motora y, al mismo tiempo, incorporar criterios que la literatura musical accesible suele omitir.

Los estudios mostraron distintos niveles de madurez. Algunos presentaron evaluaciones con usuarios y tareas concretas; otros fueron prototipos, demostraciones o contribuciones conceptuales. La evidencia fue principalmente cualitativa y se complementó con métricas como tiempo, errores, latencia, precisión, carga y aceptación. La heterogeneidad impidió una síntesis estadística, pero resultó adecuada para identificar patrones, brechas recurrentes y tipos de evidencia útiles para la rúbrica.

4.2 Hallazgos detallados por dimensión

4.2.1 Interacción inclusiva y accesibilidad

Las soluciones más sólidas describieron barreras concretas y adaptaciones vinculadas con tareas musicales. Se observaron modalidades alternativas de control, colaboración accesible, retroalimentación háptica o vocal y reducción de dependencia de destrezas motoras finas. Sin embargo, varios trabajos utilizaron los términos accesible o inclusivo sin detallar qué tarea se completó, qué apoyo fue necesario o qué barrera se redujo. El marco responde exigiendo evidencia de desempeño funcional y no únicamente intención de diseño.

4.2.2 Confiabilidad y predictibilidad

La confiabilidad apareció asociada con latencia, estabilidad de sensado, robustez de gestos, calibración y consistencia entre sesiones. En sistemas musicales interactivos, pequeñas variaciones pueden afectar sincronización, expresividad y confianza. Pese a ello, muchos estudios mencionaron problemas técnicos sin cuantificar su frecuencia o impacto. La rúbrica incorpora tanto métricas como observaciones de fallos, siempre que permitan relacionar el comportamiento del sistema con la continuidad de la tarea.

4.2.3 Seguridad y riesgo

Los riesgos fueron la dimensión menos sistemáticamente reportada. La evidencia incluyó fatiga, frustración, sobrecarga, falsos positivos, dependencia de calibración y necesidad de apoyo externo. Rara vez se describieron mitigaciones, límites operativos o monitoreo. Esta brecha es relevante porque una solución orientada a autonomía puede producir el efecto contrario si falla de manera impredecible o exige asistencia constante. NIST AI RMF permitió traducir estas preocupaciones en preguntas sobre identificación, medición, mitigación y documentación [5].

4.2.4 Interpretabilidad y transparencia

En los sistemas con automatización o adaptación, la documentación se centró más en capacidades que en comprensión del comportamiento. Se encontró poca evidencia sobre cómo la persona usuaria conoce el estado del sistema, corrige una decisión o comprende sus límites. La literatura de XAI inclusiva ayudó a ampliar la interpretabilidad desde la explicación técnica hacia la claridad operativa, la comunicación de incertidumbre y el control informado [6], [7].

4.2.5 Mantenibilidad y sostenibilidad

La mayoría de los trabajos describió prototipos y pruebas iniciales, pero ofreció información limitada sobre compatibilidad, actualizaciones, dependencia de hardware, documentación y continuidad. Esta ausencia dificulta estimar si una solución puede sostenerse fuera del entorno de investigación. La mantenibilidad se incorporó no solo como propiedad técnica, sino como condición para preservar accesibilidad: una actualización o cambio de plataforma puede eliminar adaptaciones y crear nuevas barreras.

4.2.6 Contexto funcional y resultados

Los resultados más valiosos fueron aquellos vinculados con participación, colaboración, logro de tareas y reducción de apoyo externo. Sin embargo, no todos los estudios distinguieron entre viabilidad técnica y beneficio funcional. CIF/ICF permitió organizar la interpretación alrededor de actividad, participación, barreras y facilitadores [4]. De este modo, el marco pregunta no solo si el sistema funciona, sino qué permite hacer, en qué condiciones y con qué grado de autonomía.

4.3 Síntesis de brechas

La síntesis reveló cinco brechas principales: fragmentación de criterios, predominio de evaluaciones parciales, trazabilidad débil entre evidencia y conclusión, atención insuficiente a riesgo e interpretabilidad y escasa comparación entre IA y alternativas sin IA. Estas brechas justifican una estructura multicriterio que conserve tanto los resultados positivos como los vacíos. La ausencia de información se convierte así en un hallazgo útil para orientar investigación, documentación y diseño.

5. Hallazgos y discusión

La revisión evidenció un campo tecnológicamente diverso, con interfaces colaborativas accesibles, sistemas inmersivos, retroalimentación háptica o vocal, instrumentos extendidos y soluciones basadas en sensores, bioseñales o BCI. Trabajos como DuoRhythmo documentan colaboración musical accesible para personas con ELA [9], mientras que otras propuestas muestran cómo cambios de diseño relativamente pequeños pueden aumentar inclusión sin añadir complejidad innecesaria [10]. Sistemas BCI y biofeedback amplían las posibilidades de control, pero presentan retos de ruido, entrenamiento y confiabilidad [11].

La principal brecha no es la falta de innovación, sino la debilidad comparativa. Los estudios emplean métricas y métodos heterogéneos; con frecuencia reportan experiencia de usuario o viabilidad técnica, pero omiten comportamiento sostenido, manejo de errores, mantenibilidad, interpretabilidad y riesgos. Esta situación impide determinar si una solución es realmente utilizable en condiciones funcionales concretas o si solo demuestra potencial tecnológico.

La accesibilidad debe evaluarse en tareas musicales reales. Poder abrir o navegar una interfaz no equivale a participar musicalmente. La evidencia debe mostrar si la persona puede interpretar, colaborar, componer, producir o controlar información con menor esfuerzo y dependencia. Estudios sobre interfaces colaborativas y accesibilidad del software musical refuerzan esta visión centrada en control, apropiación y continuidad [9], [12].

También se observó que la IA no constituye un indicador automático de calidad. Puede aportar adaptación, asistencia y nuevas modalidades de interacción, pero puede introducir opacidad, resultados impredecibles, calibración constante o dependencia de infraestructura. Por ello, la comparación con alternativas sin IA es esencial. El marco propuesto obliga a justificar el valor de la IA mediante resultados funcionales y evidencia verificable, en lugar de asumirlo por su novedad.

5.1 Síntesis de hallazgos por dimensión

La interacción inclusiva fue la dimensión con evidencia más frecuente: los estudios describieron controles alternativos, colaboración, retroalimentación háptica o vocal y adaptación de gestos. Sin embargo, la etiqueta "accesible" no siempre estuvo acompañada de tareas, métricas o barreras claramente documentadas. La confiabilidad apareció asociada con latencia, estabilidad, ruido de señal, reconocimiento de gestos y calibración, pero sin criterios uniformes para establecer cuándo esas limitaciones afectan de forma crítica la experiencia musical.

Seguridad, interpretabilidad y sostenibilidad fueron las dimensiones menos reportadas. Los riesgos se mencionaron indirectamente mediante fatiga, frustración, dependencia de apoyo o fallos de sensado, y pocas fuentes describieron mitigaciones sistemáticas. En sistemas con automatización, la documentación rara vez explicó qué comprende la persona usuaria, cómo puede corregir una decisión o qué ocurre cuando la IA falla. Asimismo, la mayoría de trabajos se concentró en el prototipo y la prueba inicial, con escasa información sobre mantenimiento, compatibilidad, actualizaciones y continuidad de accesibilidad.

La dimensión de contexto funcional permitió distinguir entre demostraciones técnicas y beneficios vinculados con participación. Los casos más informativos relacionaron la tecnología con colaboración, ejecución, mezcla, lectura musical o reducción de apoyo externo. Este hallazgo respalda el uso de CIF/ICF: la evaluación debe evidenciar no solo que el sistema funciona, sino que facilita una actividad musical significativa dentro de condiciones funcionales y ambientales concretas [4].

5.2 Aplicación ilustrativa

El marco se aplicó de manera ilustrativa al caso DuoRhythmo, una interfaz musical digital colaborativa orientada a personas con ELA [9]. La aplicación mostró capacidad inicial para redistribuir la evidencia del estudio entre las seis dimensiones, distinguir evidencia sólida de vacíos de información y construir un perfil interpretable. Las mayores fortalezas se relacionaron con interacción inclusiva, colaboración y contexto funcional; los principales vacíos se localizaron en riesgo explícito, interpretabilidad de componentes automatizados y continuidad técnica a largo plazo. Debido a que la aplicación se realizó sobre un único caso y con evidencia secundaria, estos resultados demuestran utilidad analítica inicial, no validación empírica definitiva.

Tabla 2. Perfil ilustrativo del caso DuoRhythmo.

Interacción inclusiva y accesibilidad: 3. Diseño y colaboración claramente orientados a accesibilidad motora.

Confiabilidad y predictibilidad: 2. Funcionamiento utilizable, limitado por el contexto remoto y el alcance de la prueba.

Seguridad y riesgo: 1. Se mencionan barreras, pero no existe evaluación sistemática de riesgos o mitigaciones.

Interpretabilidad y transparencia: 1. Evidencia limitada sobre comprensión del comportamiento interno y control informado.

Mantenibilidad y sostenibilidad: 1. Poca documentación sobre continuidad, mantenimiento y evolución del prototipo.

Contexto funcional y resultados: 3. Valor claro para participación y colaboración musical significativa.

El perfil evidencia dos fortalezas: accesibilidad de la interacción y resultados funcionales. Las puntuaciones intermedias o bajas no significan necesariamente que el sistema incumpla, sino que la publicación aporta evidencia limitada sobre confiabilidad, riesgo, transparencia y continuidad. Esta distinción es central para evitar confundir "no reportado" con "deficiente". La aplicación confirma que la rúbrica ayuda a visibilizar dimensiones que suelen quedar fuera de los reportes de experiencia de usuario [9].

5.3 Aporte metodológico

El marco integra en una sola estructura criterios que normalmente aparecen separados. Su originalidad es metodológica: conecta estándares de calidad, accesibilidad, funcionamiento humano y riesgo con evidencia de sistemas musicales accesibles. La matriz permite pasar de conceptos generales a preguntas verificables y conserva trazabilidad entre fuente, valoración e interpretación. Esto facilita auditoría académica, comparación y formulación de recomendaciones.

5.4 Implicaciones para investigación, diseño y adopción

Para investigación, el marco ofrece una estructura de reporte que puede mejorar comparabilidad entre estudios. Un artículo que documente población, tarea, modalidad de control, errores, latencia, apoyos requeridos, límites, mantenimiento y resultados funcionales proporciona evidencia más reutilizable que una descripción centrada únicamente en novedad técnica. Para diseño, la rúbrica puede utilizarse como lista de preguntas tempranas: qué barrera se reduce, qué tarea se habilita, cómo se comunica un fallo, quién conserva control y qué dependencias amenazan la continuidad.

Para adopción, el marco permite contrastar soluciones con IA y alternativas sin IA bajo la misma base. La decisión no debería premiar automáticamente una función inteligente, sino comprobar si mejora el desempeño o la autonomía sin introducir costos desproporcionados de configuración, opacidad o soporte. Este enfoque también aporta valor socioeconómico indirecto, pues ayuda a reconocer riesgos de abandono, retrabajo, asistencia constante y dependencia de tecnologías difíciles de mantener.

5.5 Amenazas a la validez del marco

La validez de contenido depende de la selección e interpretación de estándares y estudios, realizada por un único investigador. La validez externa es limitada porque el caso ilustrativo corresponde a una interfaz colaborativa para personas con ELA y no representa todas las condiciones motoras, tareas musicales o niveles de madurez tecnológica. La confiabilidad entre evaluadores aún no se ha medido; aunque los descriptores reducen ambigüedad, dos personas podrían asignar valores diferentes ante evidencia incompleta. Finalmente, la ventana 2021-2026 y las bases consultadas pueden excluir literatura relevante de otras disciplinas o idiomas.

5.6 Lectura e interpretación del perfil

El perfil evaluativo debe interpretarse como una representación de la evidencia disponible, no como una calificación absoluta del producto. Un valor bajo puede reflejar una debilidad del sistema, documentación insuficiente o ambas condiciones. Por esta razón, cada puntuación debe conservar la justificación y la fuente que la sustenta. La lectura recomendada combina tres preguntas: qué dimensiones muestran evidencia favorable, qué vacíos impiden una conclusión y qué riesgos o dependencias requieren atención antes de recomendar adopción.

No se propone sumar todas las dimensiones para obtener un único índice. Las dimensiones representan propiedades diferentes y pueden tener importancia distinta según el caso. En un instrumento controlado por BCI, la confiabilidad y la fatiga podrían ser críticas; en una interfaz colaborativa, la interacción inclusiva y el contexto funcional pueden tener mayor peso. El marco admite una interpretación contextual, pero exige que cualquier priorización se documente explícitamente.

6. Discusión

Los resultados confirman que evaluar accesibilidad en sistemas musicales requiere superar la lógica de cumplimiento técnico aislado. La interacción musical combina precisión temporal, expresividad, aprendizaje, colaboración y control, por lo que una interfaz aparentemente operable puede seguir siendo poco útil para

una tarea real. La primera implicación es que la accesibilidad debe formularse mediante actividades observables: ejecutar una secuencia, modificar un parámetro, colaborar, improvisar o completar una sesión con menor asistencia.

La segunda implicación es que la autonomía es una propiedad emergente del sistema completo. No depende solo de la facilidad de uso, sino también de estabilidad, comprensión, recuperación ante errores y continuidad. Una solución puede reducir el esfuerzo motor y simultáneamente aumentar dependencia de calibración o soporte. El enfoque multicriterio permite hacer visibles estas compensaciones y evita declarar éxito a partir de una única métrica.

La tercera implicación se relaciona con la evidencia. Los prototipos de accesibilidad suelen evaluarse con muestras pequeñas y contextos específicos, situación comprensible por la diversidad de las poblaciones y el costo de los estudios. Sin embargo, esta realidad aumenta la importancia de documentar tareas, apoyos, fallos, límites y condiciones de uso. El marco puede contribuir a mejorar la calidad del reporte y facilitar reutilización de resultados entre investigaciones.

La aplicación ilustrativa a DuoRhythmo mostró una fortaleza clara en interacción inclusiva, pero también niveles bajos en seguridad, interpretabilidad y sostenibilidad debido a información limitada. Este resultado no invalida el sistema; demuestra que una evaluación integral produce preguntas adicionales que una lectura centrada en experiencia de usuario podría omitir. La utilidad del marco reside precisamente en diferenciar evidencia favorable, evidencia limitada y ausencia de reporte.

Desde la perspectiva de ingeniería de software, la propuesta conecta calidad de producto con resultados de uso. ISO/IEC 25010 aporta estabilidad, eficiencia y mantenibilidad, mientras ISO/IEC 25059 y NIST AI RMF añaden propiedades necesarias cuando existe IA. WCAG y CIF/ICF evitan que la evaluación permanezca exclusivamente técnica. La integración ofrece una visión sociotécnica adecuada para soluciones donde el valor depende de la interacción entre persona, interfaz, tecnología y contexto.

El aporte debe interpretarse como marco inicial y no como instrumento validado. La consistencia entre evaluadores, la validez de contenido y la utilidad con usuarios requieren estudios posteriores. Aun así, la propuesta establece una base reproducible: define dimensiones, indicadores, evidencia esperada, reglas de decisión y una aplicación demostrativa. Estos elementos permiten someter el marco a crítica, refinamiento y validación de forma transparente.

7. Limitaciones y trabajo futuro

El estudio depende de la calidad de la documentación disponible y de una ventana temporal y repositorios definidos. La heterogeneidad de poblaciones, tareas, tecnologías y métricas limita la equivalencia estadística entre casos. No se contó con acceso directo a personas usuarias, especialistas, desarrolladores ni infraestructura especializada para reproducir sistemas XR, BCI o hápticos. Además, la aplicación del instrumento se concentró en un único caso documentado.

Como trabajo futuro se propone validar contenido y claridad mediante juicio de expertos en accesibilidad, ingeniería de software, IA, música y rehabilitación; aplicar la rúbrica a un conjunto mayor de sistemas; realizar pruebas con personas con discapacidad motora; examinar confiabilidad entre evaluadores; y desarrollar una herramienta digital que automatice registro de evidencia y perfiles comparativos. También se recomienda explorar versiones adaptadas a otras discapacidades sin modificar la delimitación del marco original.

Estas limitaciones orientan el trabajo futuro. Primero, se requiere validación de contenido mediante especialistas en accesibilidad, ingeniería de software, IA, interacción musical y rehabilitación. Segundo, la claridad y estabilidad de la escala deben evaluarse con varios calificadores, estimando acuerdo y revisando especialmente la distinción entre niveles 1 y 2. Tercero, el marco debe aplicarse a un conjunto diverso de sistemas con y sin IA, procurando tareas comparables y poblaciones con distintos perfiles motores. Cuarto, la evaluación documental debe complementarse con personas usuarias, observación en contexto real y estudios longitudinales de continuidad, aprendizaje, soporte y abandono.

También se propone desarrollar una herramienta digital que preserve la trazabilidad entre evidencia, puntuación y recomendación, facilite el versionado de evaluaciones y genere perfiles por dimensión sin reducir el análisis a un promedio global. Las extensiones hacia otras discapacidades deberían construirse como variantes explícitas del marco, ajustando indicadores y evidencias esperadas en lugar de asumir transferencia automática desde discapacidad motora.

7.1 Alcance de la generalización

Los resultados no permiten afirmar que el marco sea aplicable sin ajustes a toda discapacidad, tecnología o contexto musical. La delimitación en discapacidad motora y la aplicación a un caso establecen una base controlada, pero no cubren variaciones de edad, experiencia musical, severidad funcional, entorno cultural o recursos disponibles. La transferibilidad debe evaluarse mediante nuevas aplicaciones y no suponerse a partir de similitudes superficiales.

7.2 Agenda de investigación

Una agenda de investigación puede organizarse en tres niveles. El primero es instrumental: validación de contenido, acuerdo entre evaluadores y refinamiento de la escala. El segundo es empírico: aplicación a sistemas con distintas modalidades y estudios con personas usuarias. El tercero es longitudinal: seguimiento de aprendizaje, mantenimiento, abandono y evolución de accesibilidad. Esta secuencia permitiría convertir el marco inicial en una herramienta con evidencia acumulada de utilidad y confiabilidad.

7.3 Recomendaciones de reporte para estudios futuros

Los estudios futuros sobre sistemas musicales accesibles pueden mejorar su reutilización si reportan de forma explícita la población, las condiciones funcionales relevantes, la tarea musical, el entorno, la modalidad de interacción y los apoyos requeridos. También conviene describir el componente de IA, su función dentro del sistema y las alternativas disponibles cuando la automatización falla. Esta información permite distinguir el aporte del diseño de interfaz, del procesamiento musical y del componente inteligente.

La evaluación debería incluir, cuando corresponda, logro de tarea, tiempo, errores, latencia, carga, necesidad de asistencia, satisfacción y evidencia cualitativa sobre barreras. Ninguna métrica aislada representa autonomía. Por ejemplo, reducir tiempo puede acompañarse de menor control, y una alta satisfacción inicial puede no sostenerse después de semanas de uso. Por ello, se recomienda combinar resultados técnicos, experiencia de uso y consecuencias funcionales.

También es importante documentar resultados negativos y condiciones en las que el sistema no funciona. La publicación selectiva de experiencias positivas dificulta comprender riesgos y limita la comparación. Reportar fallos, abandono, calibración, incompatibilidades y necesidad de soporte aporta información esencial para la dimensión de sostenibilidad. En sistemas con IA, deben registrarse incertidumbre, comportamiento inesperado, mecanismos de corrección y decisiones sobre supervisión humana.

Finalmente, los artículos deberían diferenciar claramente entre intención de accesibilidad, evidencia de accesibilidad y resultado funcional. Una solución puede haber sido diseñada con principios inclusivos sin demostrar todavía que reduce barreras. De forma semejante, completar una tarea en una prueba controlada no implica participación sostenida en contexto real. Esta distinción permite formular conclusiones proporcionadas a la evidencia y evita

presentar prototipos iniciales como soluciones plenamente validadas.

7.4 Utilidad esperada en contextos académicos y profesionales

En el ámbito académico, el marco puede apoyar revisiones de literatura, evaluación de prototipos y diseño de protocolos de prueba. Sus dimensiones sirven como guía para identificar qué evidencia falta y qué preguntas deberían incorporarse en nuevos estudios. En cursos o proyectos de ingeniería de software, la matriz puede utilizarse para relacionar requisitos de calidad con necesidades de accesibilidad y resultados de uso.

En contextos profesionales, el instrumento puede apoyar decisiones preliminares de adquisición, diseño o mejora. Una organización podría comparar alternativas y reconocer dependencias ocultas antes de invertir en una tecnología. No obstante, la puntuación documental no sustituye pruebas con la población objetivo. Su función es ordenar evidencia, transparentar incertidumbre y orientar una evaluación más completa, no emitir una certificación automática.

Esta utilidad potencial depende de conservar la trazabilidad. Cada resultado debe mantener vínculo con la fuente, el criterio y la justificación. Una futura herramienta digital debería permitir adjuntar evidencia, registrar versiones y comparar perfiles sin ocultar las diferencias contextuales. Así, el marco podría evolucionar como repositorio de aprendizaje acumulado y no únicamente como formulario de calificación.

8. Conclusiones

Fue posible diseñar un marco evaluativo estructurado y trazable para sistemas de IA musical orientados a la accesibilidad y autonomía de personas con discapacidad motora. La revisión sistemática permitió clasificar soluciones heterogéneas y evidenció que los estudios existentes no comparten una base integral de evaluación.

ISO/IEC 25010:2023 permitió representar calidad del software mediante criterios observables; ISO/IEC 25059:2023 y NIST AI RMF incorporaron transparencia, incertidumbre y riesgo; WCAG 2.2 fortaleció la accesibilidad digital; y CIF/ICF vinculó el análisis con actividad, participación y autonomía. La integración produjo seis dimensiones y una escala común que distingue cumplimiento evidenciado, evidencia parcial, evidencia insuficiente, ausencia de reporte y no aplicabilidad.

La aplicación ilustrativa mostró capacidad inicial para organizar evidencia, reconocer fortalezas y vacíos y construir perfiles por dimensión. El hallazgo más relevante es que la presencia de IA no garantiza accesibilidad ni autonomía: su aporte debe demostrarse mediante tareas musicales significativas, comportamiento confiable, control informado, sostenibilidad y resultados funcionales. El marco constituye una base para validación y refinamiento posterior, no un instrumento definitivamente validado.

El artículo demuestra que la contribución principal no consiste en incorporar más IA, sino en mejorar la forma en que se juzga su aporte. La accesibilidad debe verificarse en tareas musicales reales; la autonomía requiere control, comprensión y menor dependencia; y la sostenibilidad exige considerar mantenimiento, compatibilidad y continuidad. Bajo esta perspectiva, una alternativa sin IA puede ser preferible cuando ofrece resultados funcionales equivalentes con menor complejidad, mientras que la IA se justifica cuando produce beneficios verificables que no serían alcanzables de otra forma.

El marco constituye una base inicial y no una certificación definitiva. Su valor reside en hacer visibles preguntas que la literatura responde de manera desigual y en obligar a vincular cada juicio con evidencia. La validación futura con expertos, evaluadores independientes y personas con discapacidad motora permitirá refinar indicadores, comprobar consistencia y determinar su utilidad para investigación, diseño y selección de tecnologías musicales accesibles.

Referencias

[1] ISO/IEC, ISO/IEC 25010:2023, Systems and software quality models, 2023.

[2] ISO/IEC, ISO/IEC 25059:2023, Quality model for AI systems, 2023.

[3] W3C, Web Content Accessibility Guidelines (WCAG) 2.2, 2023.

[4] World Health Organization, International Classification of Functioning, Disability and Health (ICF), 2001.

[5] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, 2023.

[6] M. L. Tielman et al., "Explainable AI for all—A roadmap for inclusive XAI for people with cognitive disabilities," Technology in Society, vol. 78, 2024.

[7] N. Tsakalakis et al., "A typology of explanations to support explainability-by-design," ACM Journal on Responsible Computing, vol. 2, no. 1, 2025.

[8] J. Biolchini, P. G. Mian, A. C. C. Natali, and G. H. Travassos, Systematic Review in Software Engineering, COPPE/UFRJ/PESC, 2005.

[9] B. A. Iványi et al., "DuoRhythmo: Design and remote user experience evaluation of a collaborative accessible digital musical interface for people with ALS," Proc. CHI, 2023.

[10] A. Ilsar, G. Kenning, S. Trolland, and C. Frame, "Inclusive improvisation: Exploring the line between listening and playing music," ACM Transactions on Accessible Computing, vol. 15, no. 2, 2022.

[11] T. Colafiglio et al., "Neural musical instruments through brain-computer interface and biofeedback," Proc. UMAP Adjunct, pp. 489–494, 2025.

[12] A. Lucas et al., "The qualities of convivial tools and their relevance to music software accessibility," ACM Transactions on Accessible Computing, vol. 18, no. 3, 2025.

[13] C. Macrae, "Managing risk and resilience in autonomous and intelligent systems," Risk Analysis, 2024.

[14] L. Farah et al., "Assessment of performance, interpretability, and explainability in AI-based health technologies," Mayo Clinic Proceedings: Digital Health, vol. 1, no. 2, pp. 178–194, 2023.

[15] V. Faschi et al., "An accessible software interface for collaborative music performance," Proc. Sound and Music Computing Conference, 2024.