Capítulo 5

DESARROLLO DE UN GENERADOR DE INSTRUMENTOS DE EVALUACIÓN CON INTELIGENCIA ARTIFICIAL PARA OPTIMIZAR LA MEDICIÓN DEL RENDIMIENTO ACADÉMICO

Gibrán U. López Coronel
Juan Carlos Guzmán Preciado
Ángel González-Escalante
Josué Raymundo Arce Rodríguez

Universidad Autónoma de Sinaloa
Facultad de Ingeniería Mochis
Preparatoria CU Mochis
Sinaloa, México

https://doi.org/10.36825/SEICIT.2025.03.C05

Resumen

Los instrumentos de evaluación, como rúbricas, listas de cotejo y cuestionarios, son esenciales para medir el rendimiento académico y orientar decisiones pedagógicas. No obstante, su elaboración manual requiere tiempo y conocimientos especializados, lo que limita su frecuencia y calidad. Este estudio tuvo como objetivo diseñar y evaluar un sistema web basado en inteligencia artificial (IA) que automatiza la generación de instrumentos de evaluación para distintos niveles educativos. El prototipo, desarrollado con arquitectura Modelo-Vista-Controlador (MVC), React.js, Python y la API Gemini de Google, procesa datos ingresados por el docente —nivel educativo, criterios, asignatura y contexto— para generar rúbricas y listas de cotejo exportables en formatos estándar. Se adoptó una metodología mixta con la participación de 15 docentes universitarios, quienes utilizaron el sistema y evaluaron su usabilidad mediante la escala SUS, cuestionarios ad hoc y comentarios cualitativos. Los resultados indicaron una puntuación SUS promedio de 82.5, interpretada como "excelente", una adecuación temática del 93% en rúbricas y del 87% en listas de cotejo, así como una reducción promedio del 85% en el tiempo de elaboración frente a métodos manuales. Los comentarios resaltaron la rapidez, pertinencia y facilidad de uso, aunque señalaron la necesidad de guías iniciales y opciones de edición posterior. A pesar de limitantes como que la calidad del instrumento depende directamente de la claridad y precisión de la instrucción dada por el usuario, se concluye que la IA generativa puede actuar como aliada del profesorado, optimizando la creación de instrumentos evaluativos sin comprometer su calidad y ofreciendo un recurso innovador, escalable y adaptable a diversos contextos educativos.

Palabras clave: Inteligencia artificial generativa, React, Google Gemini, Educación, Docente, Aprendizaje, Rúbrica, Lista de cotejo.

Introducción

La evaluación del aprendizaje constituye un pilar fundamental en los procesos educativos, pues a través de ella se obtiene evidencia sobre el grado

de adquisición de conocimientos, destrezas y actitudes por parte de los estudiantes. Tradicionalmente, los docentes han empleado instrumentos como rúbricas, listas de cotejo, exámenes y cuestionarios para medir el desempeño académico; sin embargo, la construcción manual de estos instrumentos demanda una inversión considerable de tiempo y recursos, así como conocimientos específicos en diseño instruccional y criterios de calidad evaluativa (Umaña, 2014). Esta carga de trabajo no siempre resulta compatible con las múltiples responsabilidades que enfrenta el profesorado, lo que puede comprometer la precisión, la coherencia y la objetividad de las evaluaciones aplicadas.

En los últimos años, la inteligencia artificial (IA) ha irrumpido con fuerza en el ámbito educativo, ofreciendo herramientas capaces de automatizar tareas complejas y personalizar experiencias de enseñanza y aprendizaje. Plataformas como Gradescope han incorporado algoritmos para agilizar la corrección de tareas, mientras que otros desarrollos experimentales generan preguntas de opción múltiple a partir de bancos de datos. No obstante, persiste la necesidad de soluciones que permitan crear, de manera dinámica y ajustada al contexto, instrumentos de evaluación completos como rúbricas y listas de cotejo adaptables a diferentes niveles educativos, materias y estilos de aprendizaje, sin renunciar a la flexibilidad y al rigor que exige la práctica docente contemporánea.

El objetivo principal frente a este panorama fue evaluar la efectividad de un sistema web generador de instrumentos de evaluación basado en IA desde su diseño, implementación y validación, cuya misión fuera facilitar el trabajo docente al optimizar los tiempos de creación y mejorar la calidad de las evaluaciones académicas. Dicha herramienta emplearía la IA Gemini de Google para procesar datos proporcionados por el docente como nivel educativo, número de criterios y aspectos, asignatura y contexto temático, y finalmente dar como salida estructuras de rúbricas y listas de cotejo que pueden exportarse en formatos estándar.

Se planteó la hipótesis de que el sistema mejoraría significativamente la eficiencia en la elaboración de instrumentos de evaluación, sin comprometer la calidad percibida por los usuarios docentes. Tras

el desarrollo de un prototipo funcional, se buscó evaluar tanto la usabilidad del sistema como la percepción de precisión y la reducción de tiempo en la creación de instrumentos utilizando para ello encuestas estructuradas basadas en la System Usability Scale (SUS) y comentarios cualitativos proporcionados por un grupo de docentes universitarios que interactuaron con la herramienta. Finalmente, se planteó analizar la aportación del sistema en comparación con métodos tradicionales y plataformas existentes, con el fin de identificar sus principales fortalezas, limitaciones y posibles líneas de mejora para futuras versiones o implementaciones a mayor escala.

La justificación de este proyecto radicó en cubrir la brecha existente entre las necesidades del profesorado y las limitaciones de las herramientas actuales. Al automatizar la construcción de instrumentos de evaluación, se liberan recursos de tiempo y esfuerzo que los docentes podrían destinar a labores de retroalimentación, diseño curricular y atención personalizada al estudiante. Además, la generación sistemática y coherente de rúbricas y listas de cotejo contribuye a elevar la objetividad del proceso evaluativo, al estandarizar criterios y minimizar la subjetividad en la calificación.

Esta propuesta se estructura de la siguiente manera: en la sección de metodología se describe detalladamente el método empleado, incluyendo el diseño del prototipo, las tecnologías utilizadas y el protocolo de evaluación con usuarios; la sección de resultados analiza los datos cuantitativos y cualitativos derivados de la aplicación de la encuesta SUS y el análisis de percepciones docentes; en la sección de discusión se contraponen los hallazgos con otros trabajos relacionados y se reflexiona sobre las implicaciones prácticas y teóricas; finalmente, en la última sección se exponen las conclusiones, las limitaciones del estudio y recomendaciones para futuras líneas de investigación e implementación en entornos de educación superior.

Metodología

La sección metodológica de este estudio adoptó un enfoque de investigación aplicada con metodología mixta, combinando técnicas

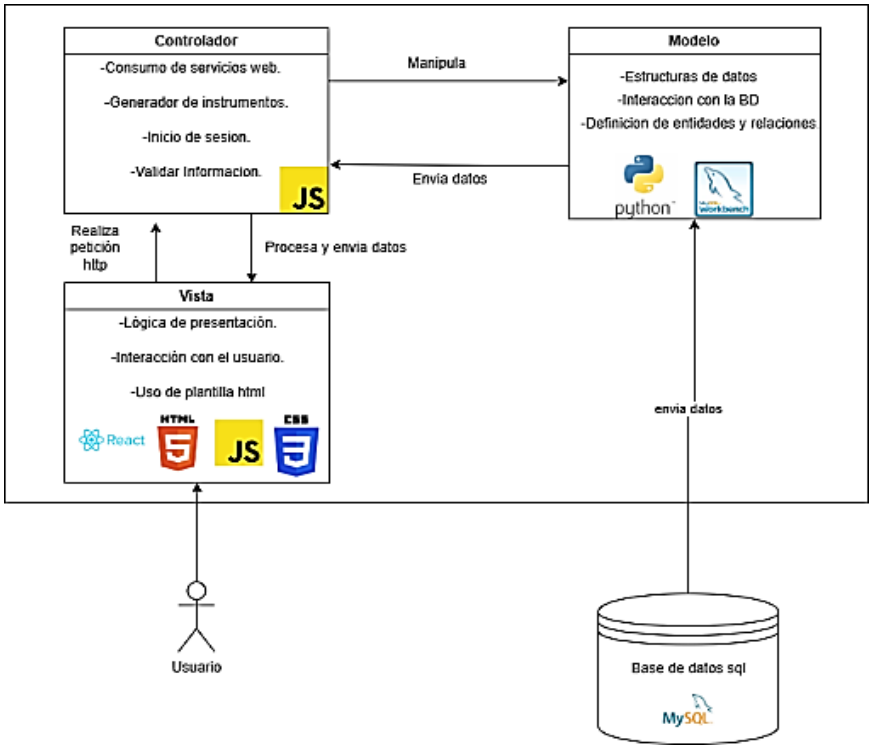
cuantitativas y cualitativas para evaluar tanto la efectividad técnica del prototipo como la experiencia real de los docentes al utilizarlo. La herramienta propuesta emplea la API Gemini de Google para procesar prompts previa incorporación de datos de nivel educativo, número de criterios y aspectos, asignatura y contexto temático proporcionados por el docente, y finalmente dar como salida estructuras de rúbricas y listas de cotejo que pueden exportarse en formatos estándar como CSV y MS Word. En la fase de diseño del sistema se optó por la arquitectura Modelo-Vista-Controlador (MVC), paradigma ampliamente reconocido por facilitar la separación de responsabilidades y mejorar la escalabilidad de aplicaciones web (Romero & González, 2012).

La capa de presentación fue desarrollada en React.js, aprovechando su capacidad para construir interfaces de usuario reactivas y componibles; se empleó Tailwind CSS para asegurar una apariencia limpia y adaptable a distintos dispositivos. El backend se implementó en Python 3.10, lenguaje multiparadigma que proporciona estructuras de datos de alto nivel y una sintaxis legible, lo cual acelera el desarrollo de servicios web (Challenger-Pérez, Díaz-Ricardo & Becerra-García, 2014). Flask actuó como microframework para exponer rutas REST que gestionan la autenticación de usuarios, la generación de prompts y la exportación de resultados. La persistencia de datos se resolvió mediante MySQL 8, servidor de base de datos relacional eficiente y de libre distribución, capaz de manejar las tablas de usuarios, los historiales de generación de instrumentos y los registros de prompts y respuestas (Casillas Santillán, Gibert Ginestà & Pérez Mora, 2014).

Para dotar de inteligencia al sistema se integró la API Gemini de Google en su versión 1.5 Pro, la cual procesa textos de solicitud, también conocidos como prompts que incluyen parámetros como el tipo de instrumento a elaborar (rúbrica o lista de cotejo), el nivel educativo en el que se aplicará, el número de criterios y aspectos, la asignatura y el contexto temático proporcionado por el docente. Esta API retorna una estructura JSON con los títulos de criterios, descriptores de desempeño y niveles de calificación, que posteriormente se parsea analizando sus

componentes y extrayendo datos útiles, ordenándolo en tablas HTML para su visualización directa en la interfaz o para su exportación en archivos CSV y MS Word mediante las librerías nativas de Python (csv y python-docx) (Google, 2024). El flujo completo desde que el usuario envía sus datos hasta que recibe el instrumento generado se ejecuta en contenedores Docker, tal como se observa en la Figura 1, garantizando portabilidad y reproducibilidad en diferentes entornos.

Figura 1
Caso de uso del prototipo



Fuente: Elaboración propia.

La evaluación del prototipo se realizó con 15 profesores de la Facultad de Ingeniería de Software de la Universidad Autónoma de Sinaloa en de Los Mochis, Sinaloa, seleccionados por muestreo de conveniencia. Estos participantes cumplían criterios de experiencia mínima de tres

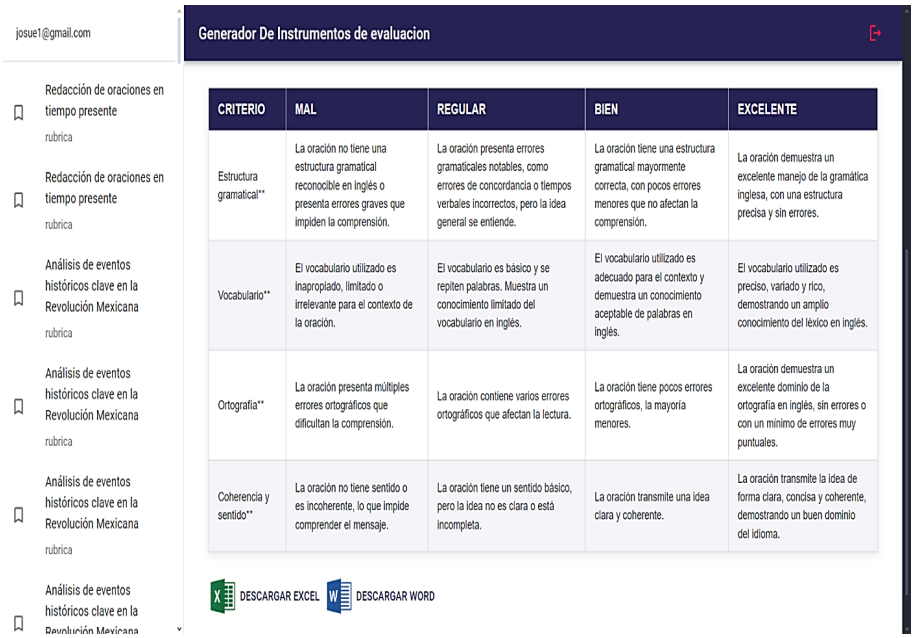
años en docencia universitaria y conocimientos básicos de herramientas ofimáticas y sistemas web. Cada docente participó voluntariamente en una sesión en laboratorio donde se les presentó una demostración del sistema donde ingresaron datos como se muestra en la Figura 2.

Figura 2
Datos para generar una rubrica de cuatro aspectos y cuatro criterios de evaluación

Fuente: Elaboración propia.

Posteriormente se les solicitó generar dos rúbricas y dos listas de cotejo para asignaturas y niveles educativos de su elección obteniendo resultados que son ejemplificados en la Figura 3. Este ejercicio práctico, de aproximadamente cuarenta minutos de duración permitió recoger métricas de tiempo de generación automáticamente desde los logs del servidor, así como garantizar que cada participante experimentara todas las funcionalidades clave: registro, generación de instrumentos y descarga de archivos.

Figura 3
Rúbrica de cuatro aspectos y cuatro criterios



Fuente: Elaboración propia.

Al concluir la fase práctica, los docentes completaron el cuestionario System Usability Scale (SUS), herramienta estandarizada compuesta por diez afirmaciones valoradas en una escala Likert de cinco puntos, que ofrece una medida cuantitativa de la usabilidad del sistema. Aunque el SUS no se fundamenta en un único autor citado en este trabajo, su empleo en estudios previos de tecnologías educativas ha demostrado proporcionar resultados comparables y consistentes en entornos de e-learning (Sandoval Villanueva, 2021; Vergara González & Carrillo Rosúa, 2023). Además, se incluyeron preguntas cerradas ad hoc para evaluar la percepción de precisión, la adecuación de los criterios al contexto temático, la facilidad de exportación y la comparación subjetiva de tiempos frente a métodos manuales. Para completar la recolección de información, se les pidió redactar respuestas abiertas donde describieran fortalezas, debilidades y sugerencias de mejora al sistema.

El análisis cuantitativo contempló el cálculo de medias, medianas y desviaciones estándar de las puntuaciones SUS y de las preguntas Likert del cuestionario ad hoc, así como la distribución de frecuencias para las variables de intención de uso futuro y percepción de reducción de tiempo. Simultáneamente, se compararon los tiempos registrados de generación automática con estimaciones de tiempo manual reportadas por los propios docentes, aunque este último dato se interpretó de manera subjetiva de acuerdo a sus respuestas. Para el análisis cualitativo se llevó a cabo una codificación temática de las respuestas abiertas empleando NVivo. Se identificaron categorías emergentes como "usabilidad", "precisión", "interfaz" y "sugerencias de mejora", logrando un coeficiente kappa de Cohen superior a 0.75, lo que denota fiabilidad inter-evaluador.

La combinación de métodos permitió triangular los hallazgos: mientras que los datos SUS ofrecieron una visión objetiva de la facilidad de uso, las preguntas específicas y los comentarios abiertos aportaron matices sobre la calidad de los instrumentos generados y las áreas de mejora. La transparencia y trazabilidad del proceso se vieron reforzadas por el uso de herramientas de control de versiones (Git) y por el almacenamiento estructurado de todos los registros en la base de datos MySQL. Este diseño metodológico, inspirado en plantillas de documentación de casos de uso y evaluaciones de prototipos en software educativo (Lund et al., 2010), asegura que los resultados sean tanto técnicamente sólidos como relevantes para la práctica docente.

Resultados

La presentación de hallazgos derivados de la implementación experimental del sistema se estructura conforme a los objetivos del estudio, presentando evidencia cuantitativa y cualitativa sobre la usabilidad del sistema, la percepción de precisión de los instrumentos generados y la reducción de tiempo en comparación con métodos tradicionales.

A. Usabilidad del sistema (SUS)

La aplicación del cuestionario SUS (System Usability Scale) arrojó resultados ampliamente favorables. La puntuación promedio obtenida fue de 82.5 puntos sobre 100 (DE = 6.3), ubicándose dentro del rango

interpretado como “excelente” según estándares internacionales de usabilidad. Esta puntuación indica que los participantes encontraron el sistema fácil de aprender, de usar y apropiado para su propósito.

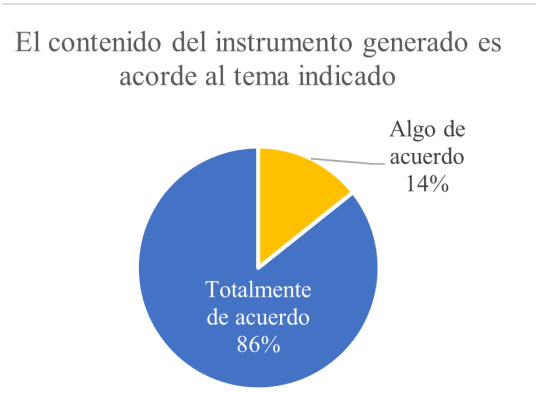
El desglose por ítem muestra que el 93% de los docentes estuvo de acuerdo o muy de acuerdo con la afirmación “Me sentiría cómodo usando este sistema frecuentemente”. Asimismo, el 87% consideró que el sistema fue “fácil de usar”, mientras que el 80% señaló que “la mayoría de las personas podrían aprender a utilizar este sistema rápidamente”. La afirmación “El sistema era innecesariamente complejo” recibió desacuerdo por parte del 93% de los participantes, lo que indica que el sistema fue considerado en general como intuitivo; sin embargo, algunos comentarios sugieren posibles ajustes para mejorar la claridad en los primeros usos.

B. Percepción de precisión y adecuación

Los resultados del cuestionario ad hoc indican que el 85.7% de los docentes consideró que los instrumentos generados fueron coherentes con el tema o contexto introducido, como se muestra en la Figura 4. Un 87% de los participantes afirmó que los criterios y descriptores incluidos en las rúbricas eran adecuados para su asignatura, mientras que el 80% señaló que las listas de cotejo reflejaban de forma pertinente los objetivos de aprendizaje planteados.

En cuanto al nivel educativo, los docentes no reportaron diferencias significativas en la calidad percibida de los instrumentos generados para niveles de primaria, secundaria o educación superior. Esta observación respalda la hipótesis de que el sistema, al estar basado en procesamiento de lenguaje natural con IA, logra una adaptabilidad semántica adecuada a diferentes niveles de complejidad académica.

Figura 4
Opinión del usuario sobre los temas generados en el instrumento



Fuente: Elaboración propia.

C. Reducción del tiempo en la elaboración de instrumentos

Uno de los hallazgos más relevantes del estudio es la reducción del tiempo necesario para crear instrumentos de evaluación. Basándose en los registros del sistema y en las percepciones de los docentes, se observó que la generación automática de una rúbrica o una lista de cotejo tomó entre 1 y 2 minutos por instrumento, incluyendo la visualización y descarga en formato MS Word o CSV. En contraste, los docentes estimaron que la creación manual de una rúbrica de complejidad media les tomaría entre 20 y 30 minutos, dependiendo del número de criterios y del nivel de personalización. En términos porcentuales, se puede afirmar que el sistema permitió una reducción de tiempo del 85% en promedio, lo que representa una ganancia significativa en productividad docente.

D. Comentarios cualitativos de los participantes

El análisis temático de las respuestas abiertas arrojó cinco categorías principales: facilidad de uso, claridad en la interfaz, precisión de los instrumentos, exportación de archivos y sugerencias de mejora. Los comentarios más frecuentes destacaron la rapidez y utilidad del sistema, en frases como: “me sorprendió lo rápido que generó una rúbrica bien estructurada” o “los criterios están bien alineados con los objetivos de la

materia".

Sin embargo, también se identificaron áreas de mejora. Algunos docentes mencionaron que la interfaz inicial podría incluir guías o ejemplos visuales para nuevos usuarios. Asimismo, se reportaron inconsistencias en los nombres de criterios cuando los temas introducidos eran demasiado abstractos o ambiguos, lo que resalta la dependencia del sistema a la claridad de los prompts proporcionados.

La Tabla 1 muestra los resultados generales obtenidos a partir de la aplicación de los instrumentos a los docentes sujetos del estudio.

Tabla 1
Resumen cuantitativo de los principales resultados

Métrica	Resultado	Interpretación
Puntuación media SUS	82.5 (DE=6.3)	Usabilidad excelente
Precisión percibida de rúbricas	93% positiva	Alta adecuación al contexto docente
Adecuación de listas de cotejo	87% positiva	Relevancia para objetivos educativos
Reducción estimada de tiempo de creación	85%	Alta eficiencia
Intención de uso futuro	93% sí	Alta disposición a adoptar la herramienta

Fuente: Elaboración propia.

Discusión

Los resultados del presente estudio validan la hipótesis central de que un sistema web apoyado en inteligencia artificial puede optimizar significativamente la elaboración de instrumentos de evaluación en contextos educativos sin comprometer su calidad. Los datos cuantitativos y cualitativos recopilados refuerzan la idea de que la automatización, cuando es cuidadosamente diseñada, no solo ahorra tiempo, sino que también puede mantener o incluso mejorar la calidad de las evaluaciones diseñadas por docentes.

El puntaje SUS obtenido (82.5) es comparable con plataformas de alta usabilidad como Moodle o Gradescope en su fase de adopción temprana. Moodle, por ejemplo, es una herramienta robusta ampliamente utilizada en educación superior; sin embargo, sus funcionalidades para generar instrumentos de evaluación como rúbricas requieren un proceso manual y no cuentan con automatización inteligente (Amorós Poveda, 2007). En contraste, el presente sistema permite una experiencia más fluida y rápida, aunque con menos opciones de edición manual posterior.

En cuanto a la precisión de los instrumentos generados, se destaca la ventaja de la adaptabilidad semántica de la API Gemini, que interpreta correctamente el tema introducido por el docente y devuelve criterios relevantes. Este hallazgo está en línea con lo reportado por Sandoval Villanueva (2021), quien utilizó IBM Watson para generar preguntas de opción múltiple a partir de temas extraídos de Wikipedia. Aunque ambos sistemas trabajan con procesamiento de lenguaje natural, el enfoque aquí presentado se diferencia por la estructura compleja de salida —rúbricas y listas completas— en lugar de ítems individuales de evaluación.

El estudio también confirma lo observado por Vergara González y Carrillo Rosúa (2023), quienes destacaron que las herramientas basadas en IA como ChatGPT o Perplexity presentan un gran potencial para asistir en tareas docentes, siempre que se empleen con una lógica pedagógica clara. En nuestro caso, el sistema requiere al usuario proporcionar información semiestructurada —nivel educativo, número de criterios, materia y contexto—, lo que permite a la IA generar salidas relevantes pero no arbitrarias. Esto constituye un avance respecto a otros sistemas donde la entrada es excesivamente abierta o desestructurada, lo cual tiende a producir resultados menos útiles.

En términos de aplicación práctica, el sistema representa una herramienta viable para docentes en contextos con alta carga laboral y escasos recursos de apoyo. Al reducir la carga cognitiva asociada a la elaboración de instrumentos, se libera tiempo que puede ser destinado a tareas de planificación, retroalimentación o atención personalizada al alumnado. Este enfoque se alinea con las recomendaciones de Díaz

Rojas y Leyva Sánchez (2013), quienes subrayan la necesidad de mejorar la calidad de los instrumentos de evaluación sin comprometer la carga laboral docente.

No obstante, el sistema también presenta limitaciones importantes. La calidad del instrumento generado depende en gran medida de la claridad y especificidad del prompt ingresado por el docente. En contextos donde los usuarios no están familiarizados con términos técnicos o estructuras evaluativas, la salida puede ser menos precisa. Además, aunque el sistema permite la exportación en MS Word y CSV, actualmente no cuenta con funcionalidades de edición posterior dentro del entorno web, lo que puede limitar su uso si se requieren ajustes finos al resultado generado. Por ello, futuras iteraciones del sistema deberían considerar integrar un editor interno que permita modificar los instrumentos antes de su descarga.

Otra limitación es la dependencia tecnológica: la API de Gemini, si bien poderosa, requiere conexión a internet estable y su versión avanzada puede implicar costos si se desea escalar el sistema a una institución educativa completa. A esto se suma la necesidad de considerar principios éticos en el manejo de los datos ingresados por el usuario, especialmente si estos contienen información sensible sobre estudiantes o contextos particulares (Rouhiainen, 2018).

Desde una perspectiva teórica, el estudio aporta a la discusión sobre la integración de inteligencia artificial en la educación no como sustituto del juicio docente, sino como asistente experto que potencia su toma de decisiones. La automatización de instrumentos de evaluación, en este sentido, debe entenderse como una ampliación de las capacidades del docente, no como una pérdida de control o profesionalidad. Así lo reflejan las respuestas cualitativas obtenidas, donde los participantes valoraron positivamente la posibilidad de ajustar los datos ingresados para obtener mejores resultados.

En comparación con otros sistemas, como COMENIO AI, que permite generar múltiples herramientas didácticas a partir de IA, el presente sistema se enfoca de forma especializada en instrumentos de evaluación

con una interfaz centrada exclusivamente en este fin (Comenio AI, 2024). Esta especialización permite una mayor optimización de los algoritmos para criterios evaluativos, algo que no siempre está presente en plataformas generalistas.

Finalmente, las implicaciones de este trabajo son amplias. El prototipo aquí descrito puede ser integrado, con los debidos ajustes, en plataformas de gestión del aprendizaje (LMS) como Moodle, permitiendo a los docentes importar automáticamente sus instrumentos de evaluación. Además, se abre la puerta a investigaciones futuras sobre la calidad psicométrica de los instrumentos generados, su alineación con marcos de competencias específicas y su aplicación en niveles educativos no universitarios.

Conclusiones

La presente investigación abordó la problemática de la elaboración manual de instrumentos de evaluación en el ámbito educativo, proponiendo una solución tecnológica basada en inteligencia artificial que automatiza la generación de rúbricas y listas de cotejo. A través del desarrollo y la evaluación de un sistema web construido bajo arquitectura MVC y apoyado en la API Gemini de Google, se logró demostrar que es posible diseñar herramientas evaluativas precisas, coherentes y adaptadas a distintos contextos, sin comprometer la calidad pedagógica ni la autonomía del docente.

Entre los hallazgos más relevantes destaca la alta usabilidad del sistema, con una puntuación SUS de 82.5, interpretada como “excelente” según estándares internacionales. Esta valoración se complementa con una percepción positiva generalizada sobre la precisión de los instrumentos generados y una significativa reducción del tiempo requerido para su elaboración —del orden del 85%— en comparación con métodos manuales. Estos resultados permiten afirmar que la solución propuesta cumple de manera efectiva con el objetivo de esta investigación, al demostrar que un sistema inteligente sí puede optimizar procesos evaluativos y ser adoptado de forma favorable por docentes con experiencia en diversos niveles educativos.

Desde una perspectiva más amplia, este trabajo contribuye al campo de la tecnología educativa en varios sentidos. Primero, al ofrecer una aplicación concreta y funcional de la IA generativa en la evaluación educativa, en un momento donde la mayoría de las implementaciones aún se encuentran en etapas exploratorias. Segundo, al enfatizar el valor de las interfaces intuitivas y los flujos de trabajo simplificados como catalizadores de adopción tecnológica, especialmente entre profesionales que no necesariamente poseen formación en ingeniería o informática. Y tercero, al posicionar al docente como agente activo en la interacción con sistemas de IA, resaltando que el valor pedagógico no reside únicamente en la herramienta, sino de cómo esta se integra y aplica de manera pertinente en el contexto educativo.

El sistema, sin embargo, no está exento de limitaciones. La dependencia de prompts bien estructurados sugiere la necesidad de ofrecer asistencia guiada en futuras versiones, mediante validadores de entrada o plantillas personalizables. Asimismo, la falta de edición interna posterior a la generación limita el grado de personalización fina que muchos docentes podrían requerir. Estas debilidades son puntos de partida para líneas futuras de desarrollo, entre las que se incluye la incorporación de editores WYSIWYG, la integración con plataformas LMS como Moodle, y la exploración de modelos multimodales que permitan introducir estímulos visuales como el uso de íconos, emojis o representaciones gráficas, o ejemplos prácticos dentro de los instrumentos.

Desde la óptica investigativa, se sugiere realizar estudios longitudinales sobre el impacto del uso sostenido de la herramienta en la práctica docente y en el rendimiento estudiantil. También es pertinente evaluar la confiabilidad y validez de los instrumentos generados mediante técnicas psicométricas, así como explorar su aplicación en ámbitos de formación técnica, educación básica y educación especial.

En síntesis, el desarrollo presentado en esta investigación constituye un aporte innovador, pertinente y escalable al ecosistema educativo digital, al demostrar que la inteligencia artificial puede actuar como aliada del docente en el diseño de instrumentos evaluativos efectivos,

relevantes y sostenibles.

Referencias

Alonso Martínez, M. (2011). Conocimiento y bases de datos: una propuesta de integración inteligente. Universidad de Cantabria.

Amorós Poveda, L. (2007). MOODLE como recurso didáctico. Universidad Católica del Maule.

Anchundia Medrano, L. A. (2022). Análisis comparativo de tecnologías Front End Angular Js Vs React Js, en el modelo de procesos para el desarrollo de aplicaciones web (Bachelor's thesis, Babahoyo: UTB-FAFI. 2022).

Barker, R. (1994). El modelo entidad-relación CASE* methodtm. Ediciones Díaz de Santos.

Casillas Santillán, L. A., Gibert Ginestà, M., & Pérez Mora, Ó. (2014). Bases de datos en MySQL. Universitat Oberta de Catalunya. https://openaccess.uoc.edu/bitstream/10609/200/5/Bases%20de%20datos_M%C3%B3dulo5_Bases%20de%20datos%20en%20MySQL.pdf

Challenger-Pérez, I., Díaz-Ricardo, Y., & Becerra-García, R. A. (2014). El lenguaje de programación Python. Ciencias Holguín, XX(2), 1-13.

Comenio.ai. (2024, Agosto 30). Conoce a tu asistente docente personal. https://www.comenio.ai

Condor Tinoco, E. E., & Soria Solís, I. (2014). Programación Web con CSS, JavaScript, PHP y AJAX. Quito, Ecuador: Instituto Tecnológico Superior "Telesup".

Díaz Rojas, Pedro Augusto, & Leyva Sánchez, Elizabeth. (2013). Metodología para determinar la calidad de los instrumentos de evaluación. Educación Médica Superior, 27(2), 269-286. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-

21412013000200014&lng=es&tlng=pt.

González, V., & Sosa, K. (2020). Lista de cotejo. Evaluación del y para el aprendizaje: instrumentos y estrategias, 18(3), 89-107.

GradeScope. (2024, Agosto 30). Deliver and Grade Your Assessments Anywhere. https://www.gradescope.com

Hamodi, Carolina, López Pastor, Víctor Manuel, & López Pastor, Ana Teresa. (2015). Medios, técnicas e instrumentos de evaluación formativa y compartida del aprendizaje en educación superior. Perfiles educativos, 37(147), 146-161. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982015000100009&lng=es&tlng=es.

Luna, Ainoa Celaya. (2024). Creación de páginas web: HTML 5. ICB, SL (Interconsulting Bureau SL).

Lund, M. I., Ferrarini Oliver, C., Aballay, L. N., Romagnano, M. G., & Meni, E. (2010). CUPIDo-Plantilla para documentar casos de uso. In V Congreso de Tecnología en Educación y Educación en Tecnología.

Ministerio de Educación Superior (2007). Reglamento de Trabajo Docente y Metodológico. Resolución 210. La Habana: Gaceta Oficial de la República de Cuba.

Pérez, M. H., & Céspedes, L. Á. L. (2021). Definición de un proceso ingenieril para el desarrollo de un chatbot a partir de buenas prácticas establecidas. Revista cubana de transformación digital, 2(3), 90-109.

Picón Jácome, É. (2013). La rúbrica y la justicia en la evaluación. Íkala, revista de lenguaje y cultura, 18(3), 79-94.

Quinaluiza Arias, A. I. (2018). Interfaz de programación de aplicaciones para la generación automática de procedimientos almacenados en Mysql [Tesis de Ingeniería en Sistemas Informáticos y Computacionales]. Universidad Técnica de Ambato.

Romero, Y. F., & González, Y. D. (2012). Patrón modelo-vista-controlador. Revista Telem@ tica, 11(1), 47-57.

Rouhiainen, L. P. (2018). Inteligencia artificial. 101 cosas que debes saber hoy sobre nuestro futuro Madrid, España: Alienta Editorial.

Sánchez Maza, M. Á. (2012). Javascript. Innovación y Cualificación, SL.

Sandoval Villanueva, J. J. (2021). Generación de preguntas y respuestas con información de Wikipedia aplicadas a través de un chatbot. Repositorio Institucional del Tecnológico Nacional de México. https://rinacional.tecnm.mx/jspui/handle/TecNM/4166

Santillán, L. A. C., Ginestà, M. G., & Mora, Ó. P. (2014). Bases de datos en MySQL. Universitat oberta de Catalunya.

Umaña, V. (2014). Evaluación del diseño instruccional en cursos en línea: un enfoque desde ADDIE. Tesis doctoral, Universidad de Costa Rica.https://convite.cenditel.gob.ve/publicaciones/revistaclic/article/download/1129/117

Vergara González, R. M., & Carrillo Rosúa, F. J. (2023). Uso de Inteligencia Artificial para diseñar propuestas didácticas de Física y Química en Educación Secundaria. En REDINE (Ed.). Conference Proceedings CIVINEDU 2023, pp. 125-131. REDINE. https://doi.org/10.58909/ad23314866

Zapata, Carlos Mario, & Garcés, Gilma Liliana. (2008). GENERACIÓN DEL DIAGRAMA DE SECUENCIAS DE UML 2.1.1 DESDE ESQUEMAS PRECONCEPTUALES. Revista EIA, (10), 89-103. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-12372008000200008&lng=en&tlng=es