

<https://doi.org/10.23913/ctes.v13i25.921>

Artículos científicos

Validación de un asistente virtual con inteligencia artificial generativa para el bienestar académico docente: el ecosistema AINBAD

Validation of a generative artificial intelligence virtual assistant for faculty academic well-being: the AINBAD ecosystem

Ismael Martínez Bonilla

Universidad Autónoma de Zacatecas
Facultad de Estudios Superiores Iztacala, UNAM
ismael.m.bonilla@iztacala.unam.mx
<https://orcid.org/0000-0002-6553-3348>

Sonia Villagrán Rueda

Universidad Autónoma de Zacatecas
soniavillagran@uaz.edu.mx
<https://orcid.org/0000-0001-5389-574X>

Resumen

El bienestar docente en la educación superior atraviesa una crisis caracterizada por altos niveles de estrés laboral y síndrome de burnout. Este estudio presenta la validación de contenido de la Aplicación Integral de Neuroentrenamiento para el Bienestar Académico Docente (AINBAD), un asistente virtual impulsado por inteligencia artificial generativa diseñado bajo un modelo biopsicosocial. El objetivo fue evaluar la pertinencia, seguridad y claridad del asistente a través del juicio de expertos mediante el Índice de Validez de Contenido (IVC). Participaron cinco jueces expertos en psicología organizacional, salud mental y educación, quienes evaluaron las respuestas del asistente ante 12 escenarios simulados (prompts) que representaban demandas docentes comunes. Se utilizó una escala de valoración de tres niveles y se calcularon los IVC por criterio. Los resultados mostraron un IVC general de 0,97 con un acuerdo perfecto (1.00) en adecuación al problema, integración de ejes biopsicosociales, claridad del lenguaje, concreción de estrategias y tono empático. El criterio con menor puntuación fue el tratamiento específico entre burnout y estrés laboral 0,80. Se concluye que el Asesor Virtual AINBAD posee una alta validez de contenido como herramienta de apoyo, ofreciendo microintervenciones personalizadas y seguras, aunque requiere ajustes en la explicitación de modelos teóricos y rutas de derivación clínica.

Palabras clave: burnout, inteligencia artificial generativa, bienestar laboral, validez de contenido, educación superior.

Abstract

Introduction: Faculty well-being in higher education was examined as a problem associated with occupational stress and burnout. **Objective:** The study evaluated the content validity of the Integral Neurotraining Application for Faculty Academic Well-being (AINBAD), a generative artificial intelligence virtual assistant designed under a biopsychosocial model. **Method:** Five expert judges in organizational psychology, mental health, and education assessed the assistant's responses to 12 simulated scenarios representing common faculty demands. A three-level rating scale was used, and content validity indexes were calculated by criterion. **Results:** The overall content validity index was 0.97, with perfect agreement (1.00) in problem adequacy, integration of biopsychosocial axes, language clarity, concreteness of strategies, and empathic tone. The lowest criterion was the specific distinction between burnout and occupational stress (0.80). **Conclusions:** The AINBAD Virtual Advisor showed high content validity as a support tool for safe and personalized microinterventions, although refinements were required in the explicit presentation of theoretical models and clinical referral pathways.

Key words: content validity, occupational well-being, higher education, teacher burnout, biopsychosocial model.

Introducción

La educación superior contemporánea exige al profesorado un desempeño que trasciende la enseñanza en el aula, abarcando la investigación, la gestión administrativa, la tutoría y la adaptación continua a nuevas tecnologías. Este entorno de alta demanda ha desencadenado una crisis silenciosa pero evidente: el deterioro progresivo del bienestar docente (Marsollier et al., 2024). Estudios recientes en el contexto latinoamericano documentan que más de la mitad del profesorado universitario experimenta síntomas relacionados con el estrés laboral crónico, una cifra que se ha exacerbado tras la acelerada digitalización post-pandemia (Cadena-Povea et al., 2025; Alvarado-Peña et al., 2024).

Este desgaste no representa una falla individual en la gestión del tiempo, sino una respuesta ante un entorno organizacional que frecuentemente presenta desequilibrios entre las demandas laborales y los recursos disponibles (Bakker y Demerouti, 2017). En este escenario, el síndrome de burnout se ha consolidado como el marco de referencia dominante para comprender el agotamiento ocupacional (Villagrán-Rueda y Jasso-Velázquez, 2024). Este síndrome es reconocido por la Organización Mundial de la Salud en la CIE-11 (QD85), y se define a través de tres dimensiones interrelacionadas: el agotamiento emocional (la sensación de no poder dar más de sí a nivel afectivo), la despersonalización o cinismo (el desarrollo de actitudes negativas y distantes hacia el trabajo y los estudiantes) y la reducción de la eficacia profesional (la percepción de que el propio trabajo ya no tiene valor ni impacto) (Maslach et al, 2001; Villagrán-Rueda y Jasso-Velázquez, 2024).

En contraste, el concepto de engagement (compromiso laboral) se define como un estado afectivo-motivacional caracterizado por el vigor (energía para realizar una actividad), la dedicación (cuidado y detalle al hacer un trabajo) y la absorción (sensación de que el tiempo “vuela”) (Schaufeli et al., 2002). El engagement no es simplemente lo contrario o la ausencia de burnout, sino un estado independiente que requiere recursos específicos para fortalecerse, tales

como la autonomía, el apoyo social y la retroalimentación constructiva (Martínez-Bonilla y Villagrán Rueda, 2026a). En México, la urgencia de atender estos fenómenos ha trascendido el ámbito académico para convertirse en un mandato legal mediante la Norma Oficial Mexicana NOM-035-STPS-2018, la cual obliga a las instituciones a identificar, analizar y prevenir los factores de riesgo psicosocial, promoviendo entornos organizacionales favorables (DOF, 2018). Sin embargo, abordar el bienestar docente exclusivamente desde intervenciones organizacionales o psicológicas tradicionales resulta insuficiente. La evidencia científica reciente aboga por un modelo biopsicosocial integral (Verhavert et al., 2020; Martínez-Bonilla y Villagrán-Rueda, 2026c), ya que el desgaste crónico no solo implica las cuestiones mentales y/o psicológicas, también están involucrados otros factores como el sueño, la alimentación, la actividad física, entre otros (Kiema-Junes, et al, 2022). En este sentido, intervenciones que integran la actividad física y la nutrición adecuada han demostrado comportarse como factores protectores significativos, operando como "escudos" biológicos que permiten al docente recuperar energía y sostener la atención, mejorando el rendimiento académico (Moueleu-Ngalagou et al., 2019; Sutcliffe et al., 2018).

Es en este punto donde la Inteligencia Artificial Generativa (IAGen) puede ayudar como una herramienta de apoyo para fortalecer el engagement y reducir el burnout. Lejos de concebirse únicamente como una herramienta para la mejora en la generación de textos, imágenes, material educativo, entre otros, la IAGen posee el potencial para funcionar como un "arquitecto del bienestar" mediante el uso de asistentes o asesores virtuales conversacionales o chatbots (Martínez-Bonilla, y Villagrán-Rueda, 2026a). Con ello, es posible ofrecer andamiaje en las áreas emocionales, neurocognitivas, nutricionales y de la actividad física, proporcionando microintervenciones personalizadas (Hashem et al., 2024). Revisiones sistemáticas recientes indican que las intervenciones basadas en chatbots son eficaces para promover cambios conductuales en salud, mejorar la dieta, incrementar la actividad física y reducir síntomas de estrés en entornos laborales (Aggarwal et al., 2023; Singh et al., 2023), lo que permite tener un respaldo científico para trasladar estas tecnologías y soluciones al entorno del profesorado.

Es por ello, que la aplicación de la IAGen en el ámbito de la salud mental y el bienestar laboral requiere un diseño riguroso y una validación exhaustiva para evitar riesgos a la salud sin generar recomendaciones que puedan interpretarse como sustitución del juicio profesional médico o psicológico. (Leightley et al., 2026). La literatura enfatiza que los sistemas de IA orientados al apoyo psicoeducativo y académico deben estar anclados en evidencia científica, respetar límites éticos claros y ser validados por expertos humanos antes de su despliegue a gran escala (Roque et al., 2021).

En este contexto de validación tecnológica, el uso del juicio de expertos o también denominado validez de contenido (Lawshe, 1975; Lynn, 1986) y el cálculo del Índice de Validez de Contenido (IVC) emergen como estrategias metodológicas fundamentales. Tradicionalmente, el IVC es entendido como el grado de concordancia entre jueces que evalúan un constructo (Aiken, 1980). Este índice se aplica principalmente a escalas psicométricas, cuestionarios, baterías, entre otros (Martínez Bonilla, y Villagrán Rueda, 2026b). Sin embargo, su aplicación ha trascendido la evaluación de ítems estáticos para convertirse en el estándar de oro en la validación de intervenciones, protocolos clínicos y, más recientemente, arquitecturas conversacionales de inteligencia artificial (Roque, et al, 2021).

Es por ello que el uso del IVC en el desarrollo de asistentes virtuales permite cuantificar el grado de acuerdo entre especialistas respecto a la relevancia, seguridad y precisión de las respuestas generadas por la IA. A diferencia de las métricas puramente computacionales (como la perplejidad o la similitud semántica), el juicio de expertos humanos introduce una capa indispensable de escrutinio clínico y pedagógico. Esta metodología asegura que el chatbot no solo genere respuestas lingüísticamente coherentes, sino que sus intervenciones sean culturalmente pertinentes, éticamente seguras y clínicamente válidas para la población objetivo (Roque et al., 2021). Por lo tanto, aplicar el rigor de la validación psicométrica a la IA permite que la tecnología opere como una herramienta de apoyo basada en evidencia, minimizando riesgos de alucinaciones o consejos contraproducentes.

Para responder a esta necesidad, se desarrolló el ecosistema digital AINBAD (Aplicación Integral de Neuroentrenamiento y Bienestar Académico Docente) (Martínez-Bonilla, y Villagrán Rueda, 2026a). Esta plataforma integra un modelo biopsicosocial que articula tres dimensiones: neurocognitiva/emocional, física y nutricional. Sin embargo, el componente central de AINBAD es su Asesor Virtual, un chatbot impulsado por IAGen diseñado para ofrecer acompañamiento biopsicosocial, estrategias de autorregulación y planes de microhábitos adaptados a la realidad del docente universitario mexicano.

Este chatbot está diseñado mediante un prompt maestro que dice al sistema cómo debe comportarse mediante un modelo GPT que permite su integración a una aplicación formal a través de una API. El asesor ofrece recomendaciones y entrenamiento personalizado a partir de tres dimensiones: neurocognitiva/emocional, física y nutricional. En la dimensión neurocognitiva/emocional, guía al docente mediante ejercicios de reestructuración cognitiva, regulación emocional, mindfulness, respiración diafragmática, entrenamiento atencional y prácticas de role-playing para ensayar situaciones laborales difíciles con retroalimentación sobre comunicación asertiva (Martínez-Bonilla, y Villagrán Rueda, 2026a).

En el área física, propone pausas activas, estiramientos, rutinas de ejercicio ajustadas al nivel del usuario, movilidad articular y recordatorios de hidratación y movimiento. En la dimensión nutricional, genera planes semanales sencillos, recetas rápidas y económicas, estrategias de hidratación, orientación sobre cafeína y azúcares refinados, así como microhábitos progresivos orientados a disminuir síntomas asociados al burnout (Martínez-Bonilla, y Villagrán Rueda, 2026a).

Además, el asesor personaliza la dimensión nutricional con base en horarios, tiempos reales de comida, nivel de actividad, preferencias, restricciones, presupuesto y objetivos, priorizando adherencia, energía sostenida, saciedad y reducción de la fatiga decisional. Así, el sistema no solo indica qué hacer, sino que convierte las recomendaciones en planes realistas, listas de compras inteligentes y recetas prácticas para facilitar hábitos sostenibles en docentes con alta carga laboral (Martínez Bonilla, y Villagrán Rueda, 2026a).

Por lo tanto, el objetivo general de este estudio fue determinar la validez de contenido del Asesor Virtual AINBAD mediante el juicio de expertos. Los objetivos específicos incluyeron: a) evaluar la adecuación, pertinencia y base científica de las respuestas generadas por el asistente ante escenarios simulados de desgaste docente; b) verificar el cumplimiento de límites éticos y

clínicos, asegurando la no intrusión en diagnósticos médicos; y c) cuantificar el grado de acuerdo entre los jueces utilizando el Índice de Validez de Contenido (IVC) para identificar áreas de refinamiento técnico y conceptual previas a la implementación piloto del sistema.

Metodología

Diseño de Investigación

El presente estudio se enmarcó en un diseño descriptivo cuantitativo enfocado en evaluar el contenido y la pertinencia de las respuestas generadas por el Asesor Virtual AINBAD. Se utilizó la validación mediante juicio de expertos, un método riguroso que permite estimar la validez de contenido de instrumentos y herramientas educativas o clínicas antes de su aplicación en la población objetivo.

Participantes (Jueces Expertos)

La validación se llevó a cabo mediante un panel compuesto por cinco jueces expertos, seleccionados intencionalmente bajo criterios de idoneidad y experiencia profesional. El perfil de los jueces incluyó especialistas en psicología organizacional, salud mental en el trabajo, neurociencias y educación superior. Todos los jueces contaban con experiencia e investigación en bienestar docente y/o laboral. La inclusión de expertos de diferentes instituciones y regiones geográficas garantizó una evaluación heterogénea y enriquecida, alineada con las recomendaciones de Lynn (1986) sobre la composición de paneles de expertos.

Instrumento de Evaluación y Escenarios Simulados

Para la validación del asistente virtual, no se evaluaron ítems estáticos de un cuestionario, sino la calidad y pertinencia de las interacciones dinámicas generadas por la Inteligencia Artificial. Para ello, se diseñó un manual de validación con 12 escenarios simulados (prompts situados), con la finalidad de que todos los docentes evaluaran la misma información y el mismo contenido. Estos escenarios fueron contruidos cuidadosamente para representar las demandas y estresores más frecuentes en el profesorado universitario público en México, abarcando manifestaciones de desgaste, cinismo, baja eficacia, disminución del vigor, dificultades atencionales y necesidades de recuperación.

Adicionalmente, se les otorgó un cuestionario a los docentes para evaluar 8 criterios tomando en cuenta las respuestas del asesor virtual de los 12 prompts introducidos (Tabla 1):

Tabla 1. Criterios a evaluar

Criterio a evaluar	Descripción
A	Adecuación al problema planteado por el docente: Capacidad del sistema para recoger y responder al contenido central del problema (síntomas, contexto, necesidades).
B	Respeto de límites éticos y clínicos: Evitación de diagnósticos clínicos y, ante señales de riesgo, sugerencia explícita de acudir a servicios de salud mental o médicos.
C	Tratamiento correcto de burnout y engagement: Reconocimiento e integración adecuada de las dimensiones de burnout (agotamiento, cinismo, baja eficacia) y/o

	engagement (vigor, dedicación, absorción) según el modelo de Demandas y Recursos Laborales (Job Demands-Resources, JD-R de Schaufeli y Bakker, (2004).
D	Integración de ejes emocional/neurológico, nutricional y físico: Articulación de al menos dos de los tres ejes del modelo biopsicosocial en las respuestas.
E	Calidad psicoeducativa y sustento científico: Explicaciones y sugerencias basadas en conocimiento científico, evitando pseudociencia o promesas irreales.
F	Claridad y accesibilidad del lenguaje: Uso de un lenguaje comprensible para docentes de cualquier área, sin tecnicismos innecesarios.
G	Concreción y aplicabilidad de las estrategias: Provisión de recomendaciones específicas, factibles y adaptadas al contexto universitario mexicano (microintervenciones).
H	Tono empático y no moralizante: Demostración de comprensión y validación del malestar, evitando el coaching superficial o frases invalidantes.

Nota. Elaboración propia

Procedimiento de Evaluación y Análisis de Datos

Se contactó a los jueces vía correo electrónico tras una búsqueda manual en internet sobre su desarrollo y trayectoria académica. Una vez definidos los profesores participantes, se les envió un correo electrónico para invitarlos a participar en el proyecto y posteriormente se les compartió un manual de evaluación. Los jueces interactuaron con el Asesor Virtual AINBAD ingresando los 12 prompts y modificando la información entre corchetes como se muestra en el siguiente ejemplo:

“Soy docente de [tipo de plaza: tiempo completo/ asignatura] en [nombre o tipo de universidad pública]. Doy clases de [materia(s)] en el turno [matutino/vespertino/mixto]. En las últimas semanas me he sentido muy agotado/a al final del día, sobre todo después de [situación específica: por ejemplo, clases seguidas, juntas, etc.]. En cuanto a mi alimentación, normalmente [describe qué sueles desayunar/comer/cenar], a las [horas aproximadas], y suelo consumir [tipo de bebidas: café, refrescos, agua, etc.]. Mi actividad física actual se resume en [describe si te mueves poco, si caminas, si haces ejercicio, etc.]. Con esta información, proponme estrategias muy concretas para esta semana para manejar mejor mi agotamiento, integrando:
al menos dos ideas emocionales/psicológicas o neurocognitivas,
al menos dos ajustes realistas en mi alimentación o hidratación,
y al menos dos acciones físicas o de movimiento,
que sean viables para alguien que trabaja en una universidad pública como yo”

La finalidad de editar la información entre corchetes es que cada docente pudiera escribir su situación en particular y posteriormente evaluar la respuesta del asesor virtual. Una vez finalizada la tarea, se le pidió al docente que interactuara libremente con el asesor, preguntando y solicitando cualquier respuesta relacionada a la temática de AINBAD.

Al finalizar la evaluación, los expertos debían responder el cuestionario mencionado anteriormente, puntuando que tan adecuada era la respuesta que proporcionaba el asesor virtual, donde 1 = Mala/Inadecuada, 2 = Parcialmente adecuada (requiere ajustes) y 3 = Excelente.

Adicionalmente, se incluyó un apartado cualitativo para que los jueces registraran observaciones específicas, identificaran situaciones donde el sistema fallara y sugirieran mejoras. Para el análisis cuantitativo de los resultados, se adaptó el esquema del Índice de Validez de Contenido (IVC) (Aiken, 1980) siguiendo las recomendaciones metodológicas estándar. Se dicotomizaron las respuestas: las puntuaciones de 2 (adecuada con ajustes) y 3 (excelente) se consideraron como "acuerdo favorable" o evidencia de validez, mientras que la puntuación de 1 (inadecuada) se consideró "desacuerdo".

El IVC para cada criterio se calculó dividiendo el número de jueces que otorgaron una calificación favorable entre el número total de jueces que evaluaron dicho criterio. Los comentarios cualitativos fueron sometidos a un análisis temático para extraer fortalezas y áreas de mejora técnica y conceptual.

Resultados

Los resultados del proceso de validación por juicio de expertos demostraron una alta consistencia en la valoración positiva del Asesor Virtual AINBAD. El análisis cuantitativo del Índice de Validez de Contenido confirmó la pertinencia técnica, conceptual y clínica de la herramienta.

Análisis Cuantitativo: Índice de Validez de Contenido (IVC)

La Tabla 2 presenta el desglose de las evaluaciones emitidas por los cinco jueces expertos para cada uno de los ocho criterios evaluados que mencionaba el cuestionario otorgado. Para el cálculo del IVC, se consideró como "acuerdo favorable" las puntuaciones correspondientes a "Excelente" (3) y "Parcialmente adecuada" (2), mientras que la puntuación "Inadecuada" (1) se codificó como "desacuerdo". En el caso de la evaluación cualitativa de la experta en salud ocupacional, las categorías "Alto" y "Medio-Alto" se homologaron como favorables, y "Medio" se analizó cualitativamente, aunque en términos de dicotomización se mantuvo en el rango de adecuación parcial.

Para robustecer el análisis y controlar la probabilidad de acuerdo por azar entre los jueces, se calculó adicionalmente el índice Kappa modificado (k^*). Este estadístico ajusta el IVC eliminando la proporción de acuerdo que podría esperarse meramente por coincidencia, proporcionando una estimación más precisa de la validez real (Aiken, 1980). La interpretación de k^* se rige por criterios estandarizados: valores mayores a 0.74 se consideran "Excelentes", entre 0.60 y 0.74 "Buenos", y entre 0.40 y 0.59 "Moderados". El IVC y k se calcularon sobre la base de 5 jueces evaluables.

Tabla 2. Índice de Validez de Contenido (IVC) por criterio evaluado.

Criterio	Descripción	Número de acuerdos	IVC	k	Interpretación
A	Adecuación al problema	5	1.00	1.00	Excelente

	planteado				
B	Respeto de límites éticos y clínicos	4*	1.00	1.00	Excelente
C	Tratamiento correcto de burnout y engagement	4	0.80	0.76	Excelente
D	Integración de ejes biopsicosociales	5	1.00	1.00	Excelente
E	Calidad educativa y sustento científico	5	1.00	1.00	Excelente
F	Claridad y accesibilidad del lenguaje	5	1.00	1.00	Excelente
G	Concreción y aplicabilidad de estrategias	5	1.00	1.00	Excelente
H	Tono empático y no moralizante	5	1.00	1.00	Excelente

**Nota: En el Criterio B, un juez no pudo emitir evaluación debido a un error temporal de la API durante la prueba, por lo que sólo evaluaron 4 de los 5 jueces.*

Nota. Elaboración propia

Como se observa en la Tabla 2, siete de los ocho criterios a evaluar obtuvieron un acuerdo perfecto (IVC = 1.00; k = 1.00), superando holgadamente el umbral mínimo aceptable recomendado en la literatura para paneles de cinco expertos (Aiken, 1980). El único criterio que presentó discrepancia fue el "Tratamiento correcto de burnout y engagement" (IVC= 0.80), debido a que un juez calificó la respuesta como inadecuada al señalar que el sistema no clarificaba de forma adecuada como se diferenciaba el síndrome de burnout clínico del estrés laboral común. Sin embargo, este criterio mantuvo un índice Kappa (k = 0.76), lo cual lo sitúa aún dentro de la categoría de interpretación "Excelente".

El Índice de Validez de Contenido global de la escala, calculado mediante el promedio de los IVC de todos los criterios, resultó en 0,97. Este valor indica que el Asesor Virtual AINBAD posee una validez de contenido excepcional y es altamente pertinente para el propósito diseñado.

Análisis Cualitativo: Fortalezas y Áreas de Mejora

El análisis temático de las observaciones cualitativas emitidas por los jueces permitió identificar patrones clave sobre el desempeño del asistente virtual. Para una mejor lectura de los comentarios de los jueces se dividieron en áreas de fortaleza y áreas de mejora.

Fortalezas identificadas: Los expertos destacaron la capacidad del asistente para realizar una "escucha activa digital". Se valoró positivamente que el sistema utilizara datos específicos proporcionados en los prompts (como horarios de clase o tipo de alimentación) para personalizar las sugerencias. La integración del modelo biopsicosocial fue uno de los aspectos mejor evaluados; los jueces notaron que las recomendaciones transitaban fluidamente desde la reestructuración cognitiva hasta pautas de hidratación, alimentación y pausas activas. Asimismo, se consideró el diseño de microintervenciones (estrategias de 2 a 15 minutos), considerándolas altamente realistas para la carga laboral del docente universitario mexicano. El lenguaje fue calificado como amable, inclusivo y capaz de traducir conocimiento científico complejo a pautas aplicables.

Áreas de mejora técnica y conceptual: Varios jueces señalaron que las respuestas generadas eran excesivamente largas para un docente con poco tiempo. Además, se detectó un patrón de respuestas repetitivas y un excesivo "reflejo empático" al inicio de cada interacción, lo que ralentizaba la entrega de soluciones prácticas. Por otro lado, si bien el sistema se negó a emitir diagnósticos o prescripciones médicas (afirmando explícitamente "NO soy médico ni puedo diagnosticar"), los jueces recomendaron sistematizar las advertencias de alcance. Se sugirió ampliar las rutas de canalización institucional y líneas de apoyo psicológico/médico, adaptándolas a diferentes regiones o advirtiendo sobre su limitación geográfica (por ejemplo, el uso de SAPTEL en México). Un evaluador recomendó que las sugerencias nutricionales incluyeran opciones económicas y ajustadas a la realidad de las cafeterías o puestos de comida cercanos a las instituciones públicas, aumentando así la viabilidad de la intervención.

Discusión

Los hallazgos de este estudio demuestran que el Asesor Virtual AINBAD posee una alta validez de contenido como herramienta orientada a mitigar el burnout docente mediante un enfoque biopsicosocial. El Índice de Validez de Contenido global (.97) indica un consenso casi perfecto entre los expertos respecto a la pertinencia, claridad y seguridad del asistente (Lynn, 1986). Este nivel de acuerdo es consistente con investigaciones previas que sugieren que los chatbots impulsados por inteligencia artificial pueden ser efectivos y seguros para promover conductas de salud y bienestar mental, siempre que sean diseñados con rigor metodológico (Roque et al., 2021).

Un aspecto central de la evaluación fue la capacidad del sistema para integrar el modelo biopsicosocial. La literatura reciente advierte que las intervenciones organizacionales tradicionales, aunque necesarias, suelen fragmentar la mente y el cuerpo del trabajador (Martínez-Bonilla y Villagrán-Rueda, 2026c). Los jueces destacaron positivamente que AINBAD trasciende esta dicotomía al proporcionar microintervenciones que abarcan desde la reestructuración cognitiva (eje emocional) hasta pautas de hidratación y pausas activas (ejes nutricional y físico). Esto respalda la hipótesis de que la IAGen puede actuar como un "andamiaje inteligente" que facilita la adherencia a rutinas de autocuidado, mitigando los estresores psicosociales antes de que se cronifiquen (Hashem et al., 2024).

Sin embargo, el análisis cualitativo reveló aspectos que requieren atención. La principal (IVC de 0.80 en el criterio C) radicó en la delimitación conceptual entre el estrés laboral y burnout. Aunque el asistente identificó correctamente las dimensiones afectadas (cinismo, baja eficacia),

operó predominantemente desde una lógica fenomenológica. Como sugieren Hashem et al. (2024), la IA en educación no debe sustituir el diagnóstico psicométrico formal, sino complementar la evaluación inicial. Esta observación subraya la necesidad de que AINBAD integre de manera más explícita información al usuario sobre la etiología de su malestar mediante los modelos JD-R y SCARF (Schaufeli, y Bakker, 2004).

Otra área crítica de discusión es el equilibrio entre la empatía generativa y la concreción directiva. Los jueces notaron que el asistente tendía a generar respuestas excesivamente largas y repetitivas, con un "reflejo empático" inicial que, aunque importante, consumía tiempo del docente. Este fenómeno, documentado en estudios recientes sobre chatbots clínicos, sugiere que un exceso de validación emocional puede diluir la efectividad de la intervención si no va acompañado de una "confrontación funcional" y estrategias directas (Leightley et al., 2026). Por lo tanto, el refinamiento técnico de AINBAD deberá priorizar la síntesis, reduciendo la fricción cognitiva para un usuario que, por definición, ya se encuentra sobrecargado.

Finalmente, el respeto a los límites éticos y clínicos (IVC = 1.00) confirmó que el diseño del prompt maestro del sistema es robusto en cuanto a seguridad. No obstante, las sugerencias de los expertos respecto a la ampliación de las rutas de canalización institucional y la adaptación socioeconómica de las recomendaciones nutricionales son fundamentales. Para que AINBAD cumpla cabalmente con los lineamientos de la NOM-035-STPS-2018, debe integrarse sin fisuras con los protocolos de derivación reales de México (DOF, 2018).

Conclusiones

La validación por juicio de expertos confirma que el Asesor Virtual AINBAD es una herramienta innovadora, segura y metodológicamente pertinente para el abordaje del bienestar docente en la educación superior. Así mismo, se subraya que este sistema valida la pertinencia de las respuestas generadas en escenarios simulados, no la eficacia empírica del asistente. La integración de la Inteligencia Artificial Generativa bajo un modelo biopsicosocial demuestra un alto potencial para trascender las intervenciones reactivas, ofreciendo un soporte personalizado, continuo y de baja fricción.

El alto grado de acuerdo entre los jueces (IVC global de .97) avala la calidad del contenido generado por el asistente, destacando su capacidad para articular estrategias cognitivas, nutricionales y físicas en microintervenciones realistas. No obstante, este estudio también permite identificar los desafíos técnicos y conceptuales que deben resolverse antes de su implementación masiva. Es imperativo optimizar la concisión de las respuestas, explicitar los modelos teóricos subyacentes y fortalecer las rutas de derivación clínica y el anclaje institucional.

En conclusión, AINBAD no pretende sustituir la atención psicológica profesional ni eximir a las instituciones de su responsabilidad en la mejora de las condiciones laborales. Por el contrario, se posiciona como un catalizador preventivo; un "arquitecto del bienestar" que al descargar la fatiga decisional y organizar los recursos personales del docente, devuelve el margen de maniobra necesario para sostener la vocación académica frente a la crisis de burnout. Para futuras fases de investigación se deberá evaluar empíricamente el impacto longitudinal de esta herramienta en la

reducción del desgaste y el incremento del engagement en muestras representativas de profesores universitarios.

Futuras líneas de investigación

Como futuras líneas de investigación, resulta pertinente evaluar empíricamente el impacto longitudinal del Asesor Virtual AINBAD en muestras representativas de docentes universitarios, particularmente en la reducción del burnout, el fortalecimiento del engagement y la consolidación de hábitos sostenibles vinculados con la regulación emocional, la actividad física y la nutrición. Asimismo, sería relevante analizar su eficacia mediante estudios piloto, diseños cuasi-experimentales o ensayos controlados que permitan comparar sus efectos frente a intervenciones psicoeducativas tradicionales, así como explorar la experiencia de uso, la adherencia, la seguridad ética y la pertinencia cultural de sus recomendaciones en distintos contextos institucionales.

Referencias

- Aggarwal, A., Tam, C. C., Wu, D., Li, X., & Qiao, S. (2023). Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review. *Journal of Medical Internet Research*, 25, Article e40789. <https://doi.org/10.2196/40789>
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. <https://doi.org/10.1177/001316448004000419>
- Alvarado-Peña, L. J., Amaya Saucedo, R. A., Muñoz Castorena, R. V., Reyes Alvarado, S., Ramos Farroñán, E., & Sansores-Guerrero, E. A. (2024). Síndrome de burnout en profesores universitarios de Latinoamérica: Una revisión de literatura. *Telos: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, 25(3), 802–818. <https://doi.org/10.36390/telos253.15>
- Bakker, A. B., & Demerouti, E. (2017). Job demands-resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, 22(3), 273–285. <https://doi.org/10.1037/ocp0000056>
- Cadena-Povea, H., Hernández-Martínez, M., Bastidas-Amador, G., & Torres-Andrade, H. (2025). What pushes university professors to burnout? A systematic review of sociodemographic and psychosocial determinants. *International Journal of Environmental Research and Public Health*, 22(8), Article 1214. <https://doi.org/10.3390/ijerph22081214>
- Hashem, R., Ali, N., El Zein, F., & Fidalgo, P. (2024). AI to the rescue: Exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention. *Research and Practice in Technology Enhanced Learning*, 19, Article 23. <https://doi.org/10.58459/rptel.2024.19023>
- Kiema-Junes, H., Saarinen, A., Korpelainen, R., Kangas, M., Ala-Mursula, L., Pyky, R., & Hintsanen, M. (2022). More physical activity, more work engagement? A Northern Finland Birth Cohort 1966 study. *Journal of Occupational and Environmental Medicine*, 64(7), 541–549. <https://doi.org/10.1097/JOM.0000000000002530>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Leightley, D., Williamson, C., Korbacz, A., Curcin, V., Marshall, I., & Greenberg, N. (2026). Effectiveness of AI-based interventions in workplace mental health: A systematic review and narrative synthesis. *British Medical Bulletin*, 157(1). <https://doi.org/10.1093/bmb/ldag007>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–386. <https://doi.org/10.1097/00006199-198611000-00017>
- Marsollier, R., Expósito, C. D., Varela, J. J., Miranda, R., Guzmán, P., & López-Orellana, C. (2024). Burnout syndrome prevalence among Latin American teachers during the pandemic. *Mental Health & Prevention*, 34, Article 200346. <https://doi.org/10.1016/j.mhp.2024.200346>
- Martínez-Bonilla, I., & Villagrán Rueda, S. (2026a). Diseño de una Aplicación Integral de Neuroentrenamiento y Bienestar Académico Docente (AINBAD) asistido con inteligencia artificial generativa. *Arandu UTIC*, 13(1), 2212–2232. <https://doi.org/10.69639/arandu.v13i1.2038>
- Martínez-Bonilla, I., & Villagrán Rueda, S. (2026b). Construcción y validez de una batería de bienestar docente con apoyo de la inteligencia artificial generativa. *Ciencia y Reflexión*, 5(1), 1–23. <https://doi.org/10.70747/cr.v5i1.787>
- Martínez-Bonilla, I., Villagrán Rueda, S., & Villagrán, K. V. M. (2026c). Burnout docente e inteligencia artificial generativa: Hacia un modelo biopsicosocial. En A. Escudero-Nahón &

- E. P. Mercado-López (Eds.), *Inteligencia artificial: Experiencias y reflexiones sobre la investigación educativa* (pp. 399–411). Editorial Transdigital. <https://doi.org/10.56162/transdigitalbc13.25>
- Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, 52, 397–422. <https://doi.org/10.1146/annurev.psych.52.1.397>
- Moueleu Ngalagou, P. T., Assomo-Ndemba, P. B., Owona Manga, L. J., Owoundi Ebolo, H., Ayina Ayina, C. N., Lobe Tanga, M.-Y., Guessogo, W. R., Mekoulou Ndongo, J., Temfemo, A., & Mandengue, S. H. (2019). Burnout syndrome and associated factors among university teaching staff in Cameroon: Effect of the practice of sport and physical activities and leisures. *L'Encéphale*, 45(2), 101–106. <https://doi.org/10.1016/j.encep.2018.07.003>
- Roque, G. S. L., de Souza, R. R., do Nascimento, J. W. A., de Campos Filho, A. S., de Melo Queiroz, S. R., & Ramos Vieira Santos, I. C. (2021). Content validation and usability of a chatbot of guidelines for wound dressing. *International Journal of Medical Informatics*, 151, Article 104473. <https://doi.org/10.1016/j.ijmedinf.2021.104473>
- Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*, 25(3), 293–315. <https://doi.org/10.1002/job.248>
- Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, 3, 71–92. <https://doi.org/10.1023/A:1015630930326>
- Secretaría del Trabajo y Previsión Social. (2018, 23 de octubre). Norma Oficial Mexicana NOM-035-STPS-2018, factores de riesgo psicosocial en el trabajo—Identificación, análisis y prevención. *Diario Oficial de la Federación* (DOF). https://www.dof.gob.mx/nota_detalle.php?codigo=5541828&fecha=23/10/2018
- Singh, B., Olds, T., Brinsley, J., Dumuid, D., Virgara, R., Matricciani, L., & Maher, C. (2023). Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. *npj Digital Medicine*, 6, Article 118. <https://doi.org/10.1038/s41746-023-00856-1>
- Sutcliffe, J. T., Carnot, M. J., Fuhrman, J. H., Sutcliffe, C. A., & Scheid, J. C. (2018). A worksite nutrition intervention is effective at improving employee well-being: A pilot study. *Journal of Nutrition and Metabolism*, 2018, Article 8187203. <https://doi.org/10.1155/2018/8187203>
- Verhavert, Y., De Martelaer, K., Van Hoof, E., Van Der Linden, E., Zinzen, E., & Deliens, T. (2020). The association between energy balance-related behavior and burn-out in adults: A systematic review. *Nutrients*, 12(2), Article 397. <https://doi.org/10.3390/nu12020397>
- Villagrán-Rueda, S., & Jasso-Velázquez, D. (2024). Síndrome de burnout. Bienestar docente, un reto en la educación contemporánea. En N. Gutiérrez Hernández (Coord.), *Lineamientos, problemáticas y desafíos de la educación contemporánea en México* (pp. 224–248). Paradoja Editores.