

EVALUACIÓN DE LA EXPERIENCIA DE USUARIO ANTE INTERFACES WEB DE SOFTWARE DE GESTIÓN A TRAVÉS DEL ANÁLISIS DE EMOCIONES: ESTADO DEL ARTE DE MÉTODOS BASADOS EN INTELIGENCIA ARTIFICIAL

EVALUATING USER EXPERIENCE IN WEB INTERFACES FOR MANAGEMENT SOFTWARE USING EMOTION ANALYSIS: STATE OF THE ART OF ARTIFICIAL INTELLIGENCE-BASED METHODS

Gustavo Poquechoque Foronda
Universidad San Francisco Xavier
poquechoque.gustavo@usfx.bo

Carlos Walter Pacheco Lora
Universidad San Francisco Xavier
pacheco.carlos@usfx.bo

Recibido: 29 Abril 2025 / Revisado: 3 Agosto 2025 / Aceptado: 11 Agosto 2025 / Publicado: 23 Septiembre 2025

Resumen

Debido a la importancia de la experiencia de usuario (UX) como factor crítico en el éxito de productos y servicios digitales, que acentúa la necesidad de incorporar dimensiones emocionales en su evaluación, considerando que hasta ahora la medición de emociones en UX se ha basado en métodos como entrevistas y cuestionarios, que presentan limitaciones en términos de subjetividad y en la captura de emociones subconscientes. En este contexto, la inteligencia artificial (IA) se perfila como una herramienta prometedora para realizar evaluaciones más objetivas mediante el reconocimiento automático de emociones.

Se ha presentado una revisión sistemática de la literatura de los últimos siete años, en la que se exploran métodos basados en IA, donde se destacan desde algoritmos clásicos de machine learning hasta modelos avanzados de aprendizaje profundo. Además, se discute la relevancia de enfoques híbridos que integran múltiples fuentes de datos para obtener una evaluación integral del estado emocional del usuario, se examina los principales conjuntos de datos utilizados en la detección de emociones, paralelamente se identifican y discuten los desafíos éticos y técnicos actuales.

Conclusivamente el estudio evidencia que la aplicación de la IA en la evaluación de UX no solo amplía el horizonte metodológico al incorporar análisis emocionales de manera automatizada, sino que también proporciona una visión holística y detallada de la experiencia del usuario, lo que es fundamental para desarrollar interfaces más intuitivas, personalizadas y satisfactorias, que auguran un futuro donde la evaluación emocional en UX sea más precisa y adaptativa, abriendo nuevas oportunidades de innovación en el campo.

Palabras claves: Inteligencia Artificial, Evaluación de la Experiencia de Usuario, Análisis de Emociones, Aprendizaje Profundo.

Abstract

Given the importance of user experience (UX) as a critical factor in the success of digital products and services, which underscores the need to incorporate emotional dimensions into its evaluation, and considering that until now, the measurement of emotions in UX has relied on methods such as interviews and questionnaires, which present limitations in terms of subjectivity and in capturing subconscious emotions, artificial intelligence (AI) emerges as a promising tool for conducting more objective assessments through automatic emotion recognition.

This paper presents a systematic review of the literature from the last seven years, exploring AI-based methods, ranging from classic machine learning algorithms to advanced deep learning models. Furthermore, it discusses the relevance of hybrid approaches that integrate multiple data sources to obtain a comprehensive assessment of the user's emotional state, examines the main datasets used in emotion detection, and identifies and discusses current ethical and technical challenges.

In conclusion, the study demonstrates that the application of AI in UX evaluation not only expands the methodological horizon by incorporating automated emotional analysis, but also provides a holistic and detailed view of the user experience, which is essential for developing more intuitive, personalized, and satisfying interfaces. This suggests a future where emotional evaluation in UX is more precise and adaptive, opening new opportunities for innovation in the field.

Keywords: Artificial Intelligence, User Experience Evaluation, Emotion Analysis, Deep Learning.

Introducción

La experiencia de usuario (UX) se ha convertido en un diferenciador clave para el éxito de productos y servicios digitales, abarcando aspectos que van más allá de la funcionalidad y usabilidad, tradicionalmente la evaluación de las emociones en UX se ha basado en métodos cualitativos, como entrevistas y cuestionarios, aunque útiles, estos enfoques presentan limitaciones relacionadas con la subjetividad y la capacidad del usuario para expresar con precisión sus sentimientos, además que pueden no capturar emociones subconscientes que afectan el comportamiento del usuario (Hernandez Perez, 2022). En años recientes la inteligencia artificial (IA) ha emergido como una herramienta prometedora para evaluar y mejorar la UX de forma más objetiva y continua (Galindo Monfil et al., 2025), en particular la integración de IA para el reconocimiento automático de emociones permite a las máquinas comprender y responder a las emociones humanas, lo que puede mejorar significativamente la interacción y satisfacción del usuario en diversos contextos (Liu, 2024).

La computación afectiva como disciplina que combina la informática y la psicología, ha desarrollado múltiples métodos de análisis emocional asistidos por IA para inferir el estado emocional del usuario, como el reconocimiento facial, el análisis de voz, el procesamiento de lenguaje natural (PLN) aplicado a textos y la sensorización fisiológica, que consiste en medir indicadores corporales como la frecuencia cardíaca, la conductancia de la piel o la actividad cerebral (EEG) para detectar cambios afectivos involuntarios (Liu, 2024), junto al apoyo de técnicas de machine learning y deep learning ha permitido alcanzar una alta precisión en entornos controlados al reconocer estados emocionales del usuario (Khare et al., 2024a) (Liu, 2024), buscando evaluaciones UX más sensibles y continuas, que puedan lograr métricas operativas

de satisfacción en ámbitos como salud, entretenimiento, educación, entre otros (Razzaq et al., 2023) (Pereira et al., 2024).

La aplicación de IA para la evaluación emocional en UX enfrenta desafíos importantes, uno de ellos son los sesgos algorítmicos, debido a que los modelos de reconocimiento emocional pueden reflejar prejuicios presentes en sus datos de entrenamiento, mostrando diferentes niveles de precisión según la demografía del usuario (como, variaciones por género, edad o cultura) (Verhoef & Fosch-Villaronga, 2023); evidenciando sesgos de género y raza, entre otros, lo cual plantea preocupaciones de equidad en su uso (Plisiecki et al., 2025), el manejo de datos emocionales sensibles también se vincula a aspectos relativos a la privacidad, la recopilación y análisis de expresiones faciales, voces o señales fisiológicas de los usuarios deben realizarse respetando marcos éticos y legales, asegurando que no se vulnere la confidencialidad ni la autonomía del individuo (Liu, 2024).

La literatura reciente presenta una fragmentación disciplinaria; el mapeo sistemático en español de Galindo Monfil et al. muestra un aumento sostenido de estudios sobre emociones y satisfacción del usuario, pero también evidencia que ambas variables se evalúan, en la mayoría de los casos, de forma independiente y con instrumentos heterogéneos (Galindo Monfil et al., 2025). Por su parte, revisiones internacionales centradas en visión por computador destacan la proliferación de propuestas basadas en deep learning, pero subrayan la escasez de análisis comparativos que integren modalidades múltiples y métricas de UX (Pereira et al., 2024). Esta dispersión dificulta a investigadores y profesionales identificar tendencias sólidas, líneas de investigación y prácticas consolidadas, lo que refuerza la necesidad de una revisión de los métodos y sus contextos de aplicación.

Se plantea relevar y ofrecer una visión sintética y crítica de los avances sobre el uso de IA para evaluar la UX a partir del análisis de emociones como alternativa a los cuestionarios tradicionales para la identificación de emociones a partir de distintos recursos, caracterizando los métodos de análisis emocional especificando sus enfoques algorítmicos y métricas, comparando su precisión y aplicabilidad, nivel de efectividad y confianza de resultados logrados en los diferentes trabajos y ámbitos de aplicación, para su posible aplicación en procesos de UX, identificando vacíos y desafíos en relación a con sus sesgos, privacidad de datos, interpretabilidad y estandarización de procedimientos; que pueda servir como una guía para investigadores y profesionales en proyectos de UX impulsados por IA.

Metodología

El artículo adopta una Revisión Sistemática de la Literatura (RSL) como enfoque, adecuado para sintetizar con rigor la evidencia, definiendo los siguientes pasos:

- Identificación de fuentes y búsquedas, cadena principal("user experience evaluation" AND "emotion* analys*" AND AI) y variantes en: IEEE Xplore, ACM Digital Library, Scopus, arXiv. Filtros: 2017 2025, idioma inglés OR español, tipo article / conference paper.
- Definición de preguntas de investigación: Establecer preguntas claras que guían la revisión, como: ¿Qué técnicas de IA se han utilizado para analizar emociones en UX? ¿Cuáles son las ventajas y limitaciones de estos métodos?
- Criterios de inclusión y exclusión: Determinar los criterios para seleccionar estudios relevantes, considerando aspectos como el

período de publicación, idioma, tipo de estudio y pertinencia al tema central.

- **Búsqueda de literatura:** Realizar búsquedas exhaustivas en bases de datos académicos reconocidos (por ejemplo, IEEE Xplore, ACM Digital Library, Scopus) utilizando palabras clave relacionadas, como "experiencia de usuario", "análisis de emociones" e "inteligencia artificial".
- **Selección de estudios:** Aplicar los criterios de inclusión y exclusión para filtrar los estudios obtenidos, asegurando la relevancia y calidad de las fuentes seleccionadas.
- **Extracción y análisis de datos:** Recopilar información clave de los estudios seleccionados, como métodos de IA utilizados, métricas de evaluación, contextos de aplicación y hallazgos principales.
- **Síntesis de resultados:** Analizar y sintetizar los datos extraídos para identificar tendencias, brechas en la investigación y oportunidades para futuros estudios en el campo.

Se oriento el trabajo a partir del establecimiento de las siguientes preguntas: ¿Qué modelos y tecnologías de IA se han utilizado para identificar emociones humanas en diferentes ámbitos y circunstancias similares a las requeridas en los ámbitos de UX? ¿Cuáles son las exigencias o requerimientos técnicos y tecnológicos, las ventajas y limitaciones de estos modelos o tecnologías?

El criterio de inclusión y exclusión de los trabajos revisados consideró: el período de publicación, idioma, tipo de estudio y pertinencia con relación al propósito principal.

Se realizaron búsquedas exhaustivas en las bases de datos de las revistas científicas Arxiv, IEEE Xplore, ACM Digital Library, Scopus entre otros.

El proceso de extracción y análisis de datos, permitió recopilar información clave de los estudios seleccionados, en relación a modelos y tecnologías de IA, principalmente Deep learning, métricas de evaluación, contextos de aplicación y hallazgos principales.

La síntesis de resultados, considero el análisis y síntesis de los datos extraídos, que permitió identificar las tendencias, las limitaciones y desafíos de los trabajos que establecen el estado del arte.

La revisión sistemática de la literatura facilitó la identificación, evaluación y síntesis de investigaciones relevantes, proporcionando una visión integral del estado actual del conocimiento en un área específica:

- Identificar tendencias y avances, junto a la detección de las técnicas de IA más utilizadas en el análisis de emociones aplicadas a la UX.
- Evaluar la eficacia y limitaciones, analizando la efectividad de estos métodos y las posibles áreas de mejora.
- Detectar vacíos en la investigación, señalando áreas que requieren mayor atención o desarrollo futuro.

Resultados

Deep learning aplicado al análisis de emociones a partir de recursos multimodales

Los métodos de IA aplicados al análisis de emociones en UX abarcan desde algoritmos de aprendizaje automático tradicional hasta modelos avanzados de aprendizaje profundo, pero se ha

dado preferencia a técnicas basadas en redes neuronales debido a su alta capacidad (Khare et al., 2024b).

Los modelos de IA aplicados pueden ser clasificados en los siguientes grupos:

- **Clasificadores de machine learning clásicos:** Algoritmos como Support Vector Machines (SVM), k-NN, árboles de decisión, bosques aleatorios o regresión logística se utilizaron ampliamente para clasificar estados emocionales a partir de características extraídas de datos de usuario, por ejemplo, un SVM entrenado sobre rasgos faciales o fisiológicos puede distinguir entre “frustrado” vs “no frustrado”. Estos métodos requieren definir manualmente las características relevantes (p.ej., frecuencia cardíaca promedio, número de clics, palabras positivas/negativas en un comentario).
- **Redes neuronales profundas:** Modelos como las redes neuronales convolucionales (CNN), las redes recurrentes (RNN) y arquitecturas híbridas han ganado protagonismo (Khare et al., 2024b), las CNN son comunes para analizar imágenes y video (p. ej., expresiones faciales), mientras que las RNN o LSTM se usan en secuencias temporales como voz, texto o señales fisiológicas, por ejemplo, las CNN han sido exitosas para reconocer expresiones faciales básicas a partir de cámaras web, y las LSTM para capturar la evolución temporal de señales EEG o el tono de voz.
- **Modelos de IA modernos:** En años recientes se incorporaron técnicas de vanguardia como redes generativas adversarias (GAN) para síntesis de expresiones, modelos de atención y especialmente transformers en el ámbito de PLN. Los modelos transformer (e.g. BERT, GPT) han revolucionado el análisis de

sentimientos en texto, permitiendo detectar emociones con mayor precisión al tener en cuenta el contexto semántico, por ejemplo, BERT y sus variantes pueden clasificar la emoción expresada en reseñas de usuarios o comentarios de redes sociales con resultados superiores al 90% de exactitud en conjuntos de prueba. Asimismo, en visión por computadora se exploran Transformers (Vision Transformers) para reconocimiento facial emocional, estos modelos profundos suelen superar a los algoritmos tradicionales cuando se dispone de grandes cantidades de datos de entrenamiento (Ghatoray & Li, 2025).

- **Enfoques híbridos y de fusión:** Dado que las emociones pueden manifestarse en múltiples canales, se emplean métodos que combinan información de diversas fuentes (análisis multimodal), por ejemplo, sistemas que integran simultáneamente análisis facial vía CNN, tono de voz vía modelos de audio, y análisis de texto de lo que el usuario dice (transcripción) vía modelos de PLN (Ghatoray & Li, 2025), la fusión puede hacerse a nivel de características (uniendo vectores de atributos de cada modalidad) o a nivel de decisión (combinando las salidas de clasificadores independientes), un enfoque de investigación reciente propone asignar pesos dinámicos a cada modalidad según su fiabilidad, usando funciones de mezcla generalizadas, para mejorar la discriminación entre emociones (Khare et al., 2024b).

Los métodos de IA más utilizados abarcan desde técnicas supervisadas clásicas (regresión, SVM, árboles) hasta modelos profundos especializados en modalidades particulares (visión, audio, texto, fisiología), la tendencia clara desde 2018 es el predominio de aprendizaje profundo multimodal, aprovechando arquitecturas neuronales avanzadas para lograr un reconocimiento más robusto de las emociones del usuario (Khare

et al., 2024b) (Ghatoray & Li, 2025). Estos métodos permiten automatizar la detección de estados afectivos relevantes durante pruebas de UX, como detectar frustración, confusión, sorpresa o satisfacción de forma objetiva, algo difícil de lograr solo con métodos tradicionales de encuestas.

Enfoques más efectivos para la medición de emociones en interacción humano-computadora

El estado del arte establece que los enfoques multimodales son los más efectivos para medir emociones en entornos de interacción humano-computadora (HCI), ninguna señal por sí sola capta toda la complejidad emocional; por ello, combinar múltiples fuentes (expresiones faciales, voz, lenguaje verbal, fisiología, comportamientos de interacción) tiende a mejorar la precisión y robustez del análisis (Ghatoray & Li, 2025) (Razzaq et al., 2023).

Los enfoques más empleados son:

Análisis multimodal integrado: Consiste en registrar varias modalidades del usuario en paralelo durante la interacción con el sistema, por ejemplo, usar la cámara para analizar microexpresiones faciales, el micrófono para el tono de voz y el contenido verbal (vía speech-to-text), y posiblemente sensores como eye-tracking o pulseras para ritmo cardíaco. La fusión de estas señales proporciona una “vista completa” de la experiencia emocional del usuario (Ghatoray & Li, 2025).

Estudios recientes demuestran que este enfoque es superior a analizar un solo canal. Razzaq et al., 2023 reportan que un modelo híbrido con fusión de audio, video y texto alcanzó un ~98% de precisión promedio al clasificar emociones básicas (alegría, tristeza, enojo, neutral), superando por margen considerable a modelos

unimodales (Razzaq et al., 2023); esto muestra cómo la combinación de modalidades logra capturar matices emocionales que podrían pasar inadvertidos con un solo tipo de dato.

Medición de respuestas fisiológicas: Las señales fisiológicas (frecuencia cardíaca, actividad electrodermal, respiración, EEG cerebral, dilatación pupilar) son valiosas porque reflejan directamente la activación emocional del sistema nervioso y son difíciles de controlar conscientemente (Khare et al., 2024b), en interacción humano-computadora, el uso de dispositivos como bandas GSR (respuesta galvánica de la piel), wearables (smartwatches con PPG para ritmo cardíaco) o EEG portátiles puede aportar indicadores objetivos de estrés, excitación o carga cognitiva. De hecho, las señales fisiológicas se consideran fuente ampliamente utilizada para identificar emociones debido a que son involuntarias y menos susceptibles a disimulo, por ejemplo, un aumento en la conductancia de la piel y la frecuencia cardíaca durante una tarea puede indicar frustración o ansiedad del usuario, complementando la lectura de su expresión facial, que si bien requieren sensores adicionales, estos datos enriquecen la evaluación emocional, especialmente para emociones internas que no siempre se manifiestan externamente (Khare et al., 2024b).

Evaluación de expresiones faciales y voz: Las expresiones faciales han sido un enfoque clásico y efectivo para detectar emociones en HCI, gracias a avances en visión por computadora, mediante cámaras web comunes permiten capturar las microexpresiones faciales del usuario; algoritmos de reconocimiento facial (Por ejemplo redes neuronales entrenadas en conjuntos de datos como FER-2013 o AffectNet) infieren emociones discretas como alegría, sorpresa, enojo o disgusto con buena precisión en tiempo real (Ghatoray & Li, 2025). Este enfoque es muy útil

para emociones de valencia claramente positiva o negativa (sonrisa = satisfacción), de modo similar, la voz del usuario (tono, ritmo, volumen) se analiza mediante técnicas de procesamiento de audio para detectar estrés o estados afectivos (voz temblorosa, tono agudo en enojo, pausas largas en confusión). La ventaja de estas señales “físicas” (cara y voz) es que se pueden captar de forma no intrusiva durante la interacción y los usuarios suelen aceptarlas mejor que los sensores corporales. Por tanto; combinadas, la expresión facial y la entonación de voz ofrecen un indicador confiable de la reacción emocional momentánea del usuario ante la interfaz.

Instrumentos subjetivos complementarios:

Aunque el presente trabajo está centrado en la IA, es importante mencionar que en evaluación UX a menudo se complementan las mediciones objetivas con autoinformes del usuario (escalas de emoción percibida, encuestas post-tarea como SAM, PANAS, etc.), ya que estudios recientes muestran que el método más utilizado históricamente para medir emociones en HCI ha sido el cuestionario auto-reportado (Galindo Monfil et al., 2025), por su sencillez, sin embargo, estos dependen de la memoria del usuario y pueden interrumpir la tarea; siendo que los enfoques de IA buscan mejorar la efectividad midiendo emociones en tiempo real sin interrumpir al usuario, en la práctica, una estrategia efectiva es la combinación de ambos enfoques, tanto los registros fisiológicos/observacionales durante la interacción, complementados con breves encuestas subjetivas, permite correlacionar las señales medidas con la experiencia auto-reportada para obtener una evaluación más completa de la UX.

En general, los enfoques más efectivos integran múltiples modalidades y métodos, aprovechando la fortaleza de cada uno, la literatura destaca que los sistemas multimodales superan

consistentemente a los unimodales en tasa de acierto y capacidad para reconocer distintas categorías emocionales (Razzaq et al., 2023). Asimismo, combinar medidas objetivas (Expresiones observadas) con las percepciones declaradas por el usuario mejora la validez de la evaluación. Por último, la contextualización es crucial: un enfoque efectivo incorpora información del contexto de la interacción (Qué tarea realiza el usuario, qué estímulos presenta la interfaz) al interpretar las señales emocionales, esto ayuda a distinguir, por ejemplo, si un ceño fruncido indica frustración con el sistema o simplemente concentración. En síntesis, el enfoque multimodal contextual donde varias fuentes de datos se analizan conjuntamente en contexto se considera el paradigma más efectivo actualmente para medir emociones en HCI y UX.

Evolución reciente de la aplicación de ia en el análisis de emociones para ux (2018–2025)

En los últimos siete años ha ocurrido una evolución significativa en cómo se aplica la IA para evaluar emociones en la experiencia de usuario, hacia 2018, muchas investigaciones de UX seguían midiendo emociones principalmente mediante métodos tradicionales (encuestas de satisfacción emocional al final de la sesión, observación manual de gestos), las primeras aplicaciones de IA en este ámbito se centraban en aspectos aislados, por ejemplo, usar reconocimiento facial o análisis de sentimientos de comentarios por separado y a menudo en entornos controlados de laboratorio. Sin embargo, conforme las técnicas de aprendizaje profundo demostraron su eficacia en reconocimiento de emociones en general (imagen, voz, texto), su adopción dentro del campo de UX aumentó rápidamente .

2018-2020, en este periodo inicial se publicaron trabajos pioneros que integraban IA en evaluaciones de UX, comenzaron a explorarse asistentes de voz (ej. Alexa, Siri) evaluados por su capacidad de generar satisfacción emocional, usando cuestionarios de emoción percibida junto con análisis del tono de voz del usuario (Galindo Monfil et al., 2025)

También surgieron estudios sobre el uso de electroencefalografía (EEG) y aprendizaje automático para detectar estados emocionales (como carga cognitiva o interés) durante el uso de sistemas educativos o videojuegos (Fernández-Ordóñez et al., 2019), no obstante, muchas de estas primeras incursiones trataban emociones y satisfacción por separado y de forma experimental. Galindo et al. (2025) hallaron en un mapeo sistemático que hasta 2024 la mayoría de trabajos evaluaban las emociones del usuario y su satisfacción de manera independiente, y que pocas investigaciones integraban ambos aspectos en una evaluación unificada (Galindo Monfil et al., 2025)

2020-2022, estos años vieron un despegue en la cantidad de investigaciones que aplican IA para UX analytics, varias tendencias confluyeron para acelerar la evolución: La madurez de frameworks de deep learning de código abierto (TensorFlow, PyTorch) facilitó a equipos de UX incorporar modelos pre-entrenados de emoción (Con modelos de reconocimiento facial ya entrenados en AffectNet) en sus estudios (Ghatoray & Li, 2025). La pandemia de COVID-19 (2020-2021) impulsó la evaluación remota de UX, generando necesidad de métodos no intrusivos para captar la experiencia del usuario a distancia, esto derivó en trabajos que utilizan la cámara web y micrófono del usuario final para evaluar sus emociones mientras interactúa con un software desde casa, dado que la observación presencial no era posible. Y finalmente aparecieron datasets más grandes y variados de emociones que permitieron entrenar

modelos más generalizables, por ejemplo, en PLN, Google liberó en 2021 GoEmotions (Un amplio conjunto de datos de textos con 27 categorías emocionales), lo cual potenció el desarrollo de análisis de sentimientos más sutiles en reseñas y comentarios de usuarios.

Durante este período, se empieza a apreciar una transición de enfoques unimodales a enfoques multimodales en UX, un estudio de 2021 integró eye-tracking y seguimiento de clics del ratón para inferir la carga emocional durante la navegación web, encontrando correlaciones entre ciertos patrones (movimientos oculares erráticos, clics repetitivos) y frustración del usuario (Ghatoray & Li, 2025). En general, 2020-22 marcó el paso de la teoría a la práctica, con más herramientas accesibles y varios estudios de caso demostrando que la IA podía añadir valor real al proceso de evaluación de experiencia de usuario.

2023-2025, en los años más recientes, la aplicación de IA en análisis emocional para UX ha alcanzado un nuevo nivel de sofisticación y adopción, y se observa un énfasis en la automatización completa del proceso de obtención de insights de UX: por ejemplo, Ghatoray & Li (2025) desarrollan un sistema que analiza automáticamente videos de sesiones de prueba de usuarios, extrayendo emociones faciales, transcribiendo el discurso del usuario a texto (Usando el modelo avanzado Whisper) y luego analizando la carga afectiva del texto con un modelo de emociones entrenado, para finalmente fusionar todos estos datos y generar insights accionables (Ghatoray & Li, 2025)

Otra evolución es la dinámica multimodal adaptativa, en lugar de combinar señales con pesos fijos, se investiga cómo ajustar la contribución de cada modalidad según la situación y la calidad de los datos, Satti et al. (2023) introducen funciones de mezcla generalizadas para asignar pesos dinámicos a

cada modalidad (video, audio, texto) dependiendo de su fiabilidad momento a momento (Por ejemplo, si el usuario deja de hablar, el sistema aumenta el peso de la expresión facial) (Razzaq et al., 2023).

En cuanto a la aceptación industrial, hacia 2025 comienzan a aparecer herramientas comerciales o prototipos avanzados que incorporan IA afectiva en el ciclo de diseño, como UXapp (Cordeiro et al., 2024) propone una plataforma que evalúa productos digitales mediante reconocimiento de emociones del usuario captadas por cámara, generando reportes para los diseñadores (referencia en (Ghatoray & Li, 2025)).

En suma, la evolución 2018–2025 se caracteriza por: Un mayor volumen de investigación, un incremento sostenido de publicaciones que abordan emociones en UX (Galindo Monfil et al., 2025). Progreso tecnológico de métodos simples hacia deep learning multimodal, posibilitando una detección más precisa y en tiempo real. La integración práctica de experimentos aislados hacia herramientas integrales que automatizan la obtención de insights emocionales. Y el reconocimiento del valor de la emoción en UX, pasando de considerarse un “dato complementario” a un componente central para evaluar y optimizar diseños. No obstante, muchos estudios recientes advierten que la aplicación de IA en este campo aún enfrenta retos y está en consolidación, requiriendo validaciones adicionales antes de generalizarse plenamente en entornos industriales y comerciales

Conjuntos de datos utilizados en la evaluación de emociones en ux mediante ia

La disponibilidad de conjuntos de datos etiquetados de emociones ha sido un factor crucial para entrenar y evaluar los modelos de IA en este campo. A continuación, se resumen los

principales datasets empleados en los últimos años, para el reconocimiento de emociones aplicable a UX, tanto multimodales como específicos por tipo de señal, estos conjuntos de datos proveen insumos para entrenar algoritmos que luego se aplican en contextos de UX, o bien sirven como benchmarks para comparar métodos.

- **DEAP (Database for Emotion Analysis using Physiological signals):** Un dataset clásico (Koelstra et al. 2012) ampliamente reutilizado hasta la actualidad, que contiene señales fisiológicas de 32 sujetos (EEG, electrooculograma, GSR, ritmo cardíaco, respiración) mientras ven videos musicales emotivos, con etiquetas de emoción en dimensiones valencia/arousal, es relevante para UX porque proporciona datos biométricos bajo estímulos audiovisuales controlados, útiles para entrenar modelos que luego podrían aplicarse a reacciones de usuario ante interfaces multimedia.
- **FER-2013:** Conjunto de ~35k imágenes de rostros en escala de grises categorizadas en 7 emociones básicas, introducido en una competencia de Kaggle 2013, si bien más pequeño y con datos menos “en el mundo real” que AffectNet, sigue usándose como benchmark ligero. Por ejemplo, Ghatoray & Li (2025) probaron 10 modelos pre-entrenados de reconocimiento facial en FER-2013, AffectNet y CK+ para elegir el más generalizable (Ghatoray & Li, 2025). FER-2013 suele emplearse para pruebas rápidas o entrenar modelos simplificados que se puedan ejecutar en tiempo real durante sesiones de UX.
- **WESAD (Wearable Stress and Affect Dataset):** Publicado en 2018 es un conjunto de datos de sensores wearables enfocado en estrés y afecto, registra señales fisiológicas (cardíacas, GSR, movimientos) de 15 sujetos

mediante un dispositivo tipo banda en el pecho (RespiBAN) y una pulsera, en condiciones de inducción de estrés, relajación y estado neutro, las etiquetas distinguen emociones como estrés vs calma vs diversión. WESAD ha cobrado importancia para entrenar modelos menos intrusivos (solo con *wearables*) que podrían integrarse en evaluaciones de UX móviles o *in-the-wild*. De hecho, estudios recientes privilegian el uso de dispositivos portables para obtener medidas emocionales continuas sin interrumpir la experiencia del usuario.

- **DREAMER:** Dataset de 2018 con 23 sujetos, datos EEG y ECG y etiquetas de valencia, activación y dominancia para estímulos audiovisuales emocionales, se destaca porque muchos estudios reportan resultados muy altos de precisión en él, por ejemplo 100% en clasificación binaria de valencia/activación sirviendo como banco de pruebas para nuevos algoritmos, aparece citado en numerosos trabajos recientes de reconocimiento emocional (Khare et al., 2024b).
- **AMIGOS (Augmented Multimodal Interaction dataset for emotion Recognition):** Publicado en 2017, incluye 40 sujetos con datos multimodales (EEG, ECG, GSR, expresiones faciales en video) mientras ven videos cortos y socializan, con anotaciones de emociones en escalas continuo y discreto, en estudios revisados se menciona que AMIGOS es uno de los conjuntos públicos más usados para emoción multimodal junto con ASCERTAIN (Khare et al., 2024b).
- **ASCERTAIN:** Un dataset de 2018 diseñado para estudiar emociones influenciadas por rasgos de personalidad, registra señales multimodales (EEG, pulsioximetría, expresiones faciales) de 58 participantes expuestos a estímulos audiovisuales, con etiquetas de emoción en dimensiones V/A y también evaluación de rasgos de personalidad. ASCERTAIN ha sido utilizado en al menos 3 estudios recientes según el mapeo realizado (Khare et al., 2024b)
- **AffectNet:** Creado en 2017, es uno de los mayores datasets de expresiones faciales en imágenes estáticas, con ~0.4 millones de fotos de rostros recopiladas de Internet etiquetadas en 11 categorías emocionales + valores de valencia/arousal, es el estándar actual para entrenar y evaluar modelos de reconocimiento facial de emociones en entornos no controlados (Ghatoray & Li, 2025)
- **RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song):** Dataset de 2018 con 24 actores que recitan frases con distintas emociones (calma, alegría, tristeza, enojo, miedo, disgusto, sorpresa) en audio y video, es utilizado para entrenar o evaluar sistemas bimodales de emoción (cara+voz) a pequeña escala. Por ejemplo, proyectos experimentales de UX han usado RAVDESS para probar algoritmos que detecten automáticamente el estado emocional del usuario tanto por su rostro como por su voz en escenarios como videojuegos.
- **Conjuntos de datos de texto (NLP):** En análisis de sentimiento aplicado a UX, se suelen utilizar corpora estándar de opiniones y también datos específicos del dominio, entre los generales se encuentra IMDb reviews (sentimiento de reseñas de películas), Sentiment140 o SemEval (tweets etiquetados en sentimiento o emoción) y el mencionado GoEmotions (Reddit, 27 emociones). Además, existen datasets enfocados en aspectos de usabilidad, por ejemplo corpus de reseñas de aplicaciones móviles donde cada review está etiquetada con facetas de UX (desempeño, UI, etc.) y polaridad (Alonazi, 2023).

• **DUX (Dataset of User eXperience):**

Publicado en 2023, este conjunto de datos fue diseñado específicamente para la intersección de emociones y UX. Reconociendo que “no existía un dataset público para reconocimiento

de emociones a partir solo de interacciones (teclado, ratón, táctil)”, Leppich et al. (2023) crearon DUX para llenar ese vacío (Leppich et al., 2023).

Tabla 1. Conjuntos de datos más utilizados para el análisis automático de emociones en UX.

Conjunto de Datos	Modalidad	Emociones	Uso en investigación UX
FER2013 (Goodfellow et al., 2013)	Imágenes faciales (estáticas)	7 emociones básicas	Referencia estándar para entrenar clasificadores faciales de emociones; ampliamente usado en estudios de UX por su disponibilidad (Santos & Digiampietri, 2024).
AffectNet (Mollahosseini et al., 2019)	Imágenes faciales (in-the-wild)	8 emociones + valores de valencia/arousal	Gran base de datos en condiciones no controladas; permite mejorar la generalización de modelos de expresión facial (Mollahosseini et al., 2019).
IEMOCAP (Busso et al., 2008)	Audio-visual (diálogos actuados)	6 emociones básicas (audio y video)	Muy utilizada para entrenar y evaluar reconocimiento de emociones en la voz y gestos; benchmark en muchos trabajos de HCI (Razzaq et al., 2023).
RAVDESS (Livingstone & Russo, 2018)	Audio-visual (actuaciones)	8 emociones (voz y expresión)	Conjunto controlado de expresiones vocales y faciales, empleado para modelos multimodales en contextos controlados (Razzaq et al., 2023).
EMORepository / Kaggle	Texto (opiniones) o Imágenes	Depende del dataset (sentimiento o emociones básicas)	Colecciones públicas para <i>sentiment analysis</i> (p. ej., reseñas) o retos de emociones faciales (FER, Emotion Kaggle); útiles para adaptar IA a casos de UX textuales o visuales (Santos & Digiampietri, 2024).
Datasets propietarios	Multimodal (video UX, sensores)	Emociones definidas por el estudio (e.g., frustración, engagement)	Muchos investigadores crean sus propios datasets de pruebas de usuario (grabaciones de video, interacciones instrumentadas) para entrenar modelos adaptados a su contexto específico (Santos & Digiampietri, 2024).

Como se observa, FER2013 destaca como uno de los conjuntos más frecuentes para reconocimiento facial en UX (Santos & Digiampietri, 2024), probablemente por su fácil acceso y amplia etiqueta de emociones. Sin embargo, también se han incorporado datasets más ricos, como AffectNet que provee cientos de miles de rostros

en entornos variados con anotaciones continuas de afecto, lo que ha permitido entrenar modelos más generalizables a situaciones del mundo real (Mollahosseini et al., 2019).

En el dominio del audio, corpus como IEMOCAP y RAVDESS son la referencia para evaluar la

capacidad de la IA en detectar emociones a partir de la voz, incluyendo matices como el tono y la cadencia emocional en el habla (Razzaq et al., 2023), para análisis de sentimientos en texto orientado a UX, es común reutilizar conjuntos de reseñas de usuarios o tweets etiquetados emocionalmente (por ejemplo, datasets de competencias SemEval o colecciones de opiniones de productos). Adicionalmente, la tabla refleja que no existe un único dataset “UX-emotions” estándar, sino que a menudo los investigadores generan sus propios datos durante estudios de usuario instrumentados (Santos & Digiampietri, 2024), esto se debe a que las emociones en UX pueden ser muy contextuales; por ende, grabar sesiones de interacción reales (con cámaras, micrófonos, sensores) y luego etiquetar manualmente las emociones experimentadas ofrece datos más específicos para entrenar modelos enfocados en predecir experiencias particulares (como detectar frustración en el uso de cierto software).

En general, los datasets más utilizados reflejan las distintas modalidades, son los de señales fisiológicas (DEAP, MAHNOB, WESAD, ASCERTAIN) alimentan desarrollos en detección de estados internos; los de expresión facial y vocal (AffectNet, FER-2013, IEMOCAP, RAVDESS) potencian sistemas de observación externa no intrusiva; y datasets de texto (GoEmotions, etc.) que permiten extraer sentimientos de las palabras de los usuarios, se muestra preferencia por datasets multimodales públicos que combinen varias entradas (p.ej., audio+video+EEG) para entrenar modelos integrales (Khare et al., 2024b), si bien aún se carece de suficientes datos multimodales “en el mundo real”, una línea emergente es la generación de nuevos conjuntos de datos de emociones específicos de UX, como DUX, que capturen interacciones naturales con sistemas, esta ampliación de datos disponibles seguirá siendo fundamental para avanzar el campo, pues

muchos desafíos actuales (generalización, sesgo) se deben en parte a la limitada diversidad de los datasets empleados.

Métricas para validar la precisión y efectividad de los modelos de ia en evaluación de emociones en ux

Para evaluar el desempeño de los modelos de IA en la tarea de reconocer emociones, se utilizan principalmente las mismas métricas clásicas que en otros problemas de clasificación y predicción, adaptadas al contexto de emociones, entre las métricas cuantitativas más comunes reportadas en la literatura se incluyen:

- **Exactitud (accuracy):** Proporción de aciertos del modelo, es decir, el porcentaje de casos en que la emoción predicha coincidió con la etiqueta real, es una métrica general fácil de interpretar, es así, que una exactitud del 90% indica que en 9 de cada 10 instancias el sistema clasificó correctamente la emoción (Por ejemplo identificó correctamente si el usuario estaba frustrado o no). Muchos trabajos informan esta métrica; de hecho, se cita accuracy en prácticamente todos los estudios de reconocimiento emocional (Khare et al., 2024b). Sin embargo, puede ser engañosa si las clases están desequilibradas (por ejemplo, si el 80% de las muestras son “neutral” el modelo puede lograr 80% de accuracy prediciendo siempre “neutral” sin realmente funcionar para las otras emociones), por ello se complementa con otras medidas.
- **Precisión y Exhaustividad (Precision & Recall):** Estas métricas evalúan el rendimiento por clase (emociones específicas) considerando falsos positivos y falsos negativos. Precisión (también llamada valor predictivo positivo) mide qué proporción de las predicciones de una emoción fueron

correctas; Recall (o sensibilidad) mide qué proporción de las ocurrencias reales de esa emoción fueron detectadas por el modelo. Por ejemplo, en la detección de “frustración” durante una tarea, la precisión indica cuántos de los casos que el modelo marcó como “frustrado” efectivamente lo estaban (evitando alarmas falsas), mientras que la exhaustividad indica cuántos de todos los usuarios frustrados el modelo logró identificar (evitando omisiones) (Khare et al., 2024b).

- **Puntuación F1:** Es la media armónica de precisión y recall, proporcionando un solo indicador por clase que equilibra ambos. La F1-score es útil para comparar modelos cuando hay clases desbalanceadas o cuando se quiere una métrica que penalice tanto los falsos positivos como los falsos negativos; muchos artículos utilizan el promedio F1 macro (promediado sobre todas las clases de emoción) como métrica principal de desempeño del sistema, ya que resume su capacidad global considerando cada clase por igual (Khare et al., 2024b)
- **Matriz de confusión:** Si bien no es un “índice” numérico único, la matriz de confusión es un resultado esencial que muestran las investigaciones, que detalla para cada emoción verdadera, cómo se distribuyen las predicciones del modelo (aciertos y confusiones con otras emociones), esto permite ver patrones. Por ejemplo si la emoción “sorpresa” suele ser confundida con “alegría”, o “neutral” con “aburrimiento”. Es muy útil en UX para entender qué errores comete el modelo, lo que puede guiar mejoras (quizás se necesita más datos o ajustar la definición de ciertas emociones).
- **ROC-AUC (Area Under the ROC Curve):** En escenarios binarios (Por ejemplo, detectar si el usuario está experimentando una emoción

negativa vs no), se utiliza la curva ROC que muestra la tasa de verdaderos positivos vs falsos positivos a distintos umbrales, y su área bajo la curva (AUC) como medida de desempeño global, un AUC de 0.5 indica desempeño aleatorio, 0.9 un modelo muy bueno, en problemas multiemoción, a veces se computa el AUC para cada clase versus el resto.

- **Medidas específicas de regresión:** Cuando las emociones se tratan en dimensiones continuas (valencia, activación) en lugar de categorías discretas, se emplean métricas de regresión como RMSE (root mean square error) o coeficiente de correlación de Pearson entre las predicciones continuas del modelo y los valores auto-reportados. Por ejemplo, si un modelo predice nivel de estrés en escala 1–10, se medirá cuánto difiere de las valoraciones reales del usuario promedio.
- **Índices kappa o alfa:** En algunos trabajos, especialmente aquellos que comparan la IA contra evaluadores humanos, se usan métricas de concordancia como Cohen’s kappa para ver qué tan de acuerdo está el modelo con etiquetas de referencia teniendo en cuenta la concordancia por azar.

En suma, la exactitud global (accuracy) es la métrica más reportada, indicando el porcentaje de aciertos en la detección de la emoción correcta, Razzaq *et al.* (2023) informan que su modelo multimodal H-MMER alcanza un 98.2% de accuracy promedio reconociendo cuatro estados emocionales básicos, superando abordajes previos; No obstante, dado que la distribución de emociones suele ser desigual, también se emplean métricas por clase como la precisión y recuperación (*precision/recall*) por emoción, y el puntaje F1 que equilibra ambas, para asegurar que emociones minoritarias (como “miedo” o “asco”) no queden totalmente opacadas por las

mayoritarias (Santos & Digiampietri, 2024). Es común ver matrices de confusión en los estudios, detallando cuánto se confunden unas emociones con otras y destacando cuáles se reconocen mejor o peor (Razzaq et al., 2023). Por ejemplo, en un experimento de reconocimiento a partir de audio, la emoción *enojo* pudo reconocerse con ~71% de acierto mientras tristeza cayó por debajo de 60%, evidenciando diferencias en detectabilidad según la emoción, además de las métricas puramente algorítmicas, en contexto UX interesa validar cómo esas detecciones reflejan la experiencia del usuario. Por ello, algunos trabajos correlacionan las salidas del modelo con medidas de UX establecidas, un enfoque es comparar la predicción de la IA sobre si un usuario tuvo una UX positiva o negativa con las puntuaciones reales que el usuario dio en cuestionarios de usabilidad/satisfacción. Es así, que Koonsanit & Nishiuchi (2020) extrajeron características emocionales de rostros junto con datos demográficos para predecir el nivel de UX, y luego compararon esa predicción con las respuestas del usuario en un formulario post-uso, obteniendo así una validez concurrente de su método (Santos & Digiampietri, 2024). De manera similar, métricas de correlación (coeficiente de Pearson, etc.) se han utilizado cuando las emociones se miden en dimensiones continuas (valencia/arousal) para ver cuán cerca sigue el modelo las autoevaluaciones emocionales del usuario en escala de agrado, estrés, etc. Concluyentemente, se evidencia que se emplea un conjunto amplio de métricas para validar estos sistemas: desde tasas de clasificación correctas, hasta concordancia con evaluaciones humanas, asegurando tanto la efectividad técnica del modelo como su relevancia práctica para inferir la experiencia del usuario.

Discusión

Principales desafíos y limitaciones en la aplicación de ia para evaluar emociones en ux

Aunque las técnicas de IA para detección de emociones han avanzado notablemente, su aplicación en la evaluación de UX enfrenta desafíos significativos y limitaciones que deben ser considerados:

- **Generalización y robustez limitada:** Un desafío mayor es lograr que los modelos funcionen bien más allá de las condiciones en que fueron entrenados, ya que muchos sistemas de reconocimiento emocional muestran un desempeño excelente en datasets de laboratorio, pero bajan su precisión en escenarios reales de UX donde las condiciones varían (iluminación diferente, ruido de fondo, diversidad de usuarios); La presente investigación ha identificado falta de generalización debido a diferencias en dispositivos de adquisición y duraciones de señales utilizadas (Khare et al., 2024b). Por ejemplo, un modelo entrenado con videos frontales puede fallar si la cámara del usuario está en ángulo lateral; o un modelo de voz entrenado con pocos acentos puede no generalizar a usuarios de otras nacionalidades. Esta sensibilidad reduce la confianza en usar estos modelos directamente en evaluaciones de UX amplias, para lo que se requieren más datos diversos y técnicas como domain adaptation para mejorar la robustez.
- **Falta de explicabilidad y confianza en los resultados:** Las salidas de un modelo de IA emocional a veces contradicen la intuición o la información de otras fuentes, generando desconfianza en stakeholders y expertos humanos (Khare et al., 2024b). Por ejemplo, si un sistema indica “alto nivel de frustración”

pero el usuario en la grabación parece calmado según un observador, se duda del sistema. Por eso se considera que muchos modelos de deep learning son cajas negras, por lo que especialistas en UX son reticentes a basar decisiones de diseño en ellos sin entendimiento del porqué de sus inferencias, eso sin duda es un desafío crítico, ya que se necesitan técnicas de IA explicable (XAI) que permitan interpretar qué indicadores (rasgos faciales, tono, etc.) llevaron a la conclusión de cierta emoción; Es decir, sin explicaciones, es difícil confiar plenamente en la IA, limitando su adopción práctica.

- **Ambigüedad y complejidad de las emociones humanas:** Detectar emociones no es tan sencillo como identificar un color o un objeto; las emociones son estados internos complejos, a veces sutiles, y los observables pueden ser engañosos, porque un mismo gesto (Por ejemplo suspirar) puede significar alivio, aburrimiento o frustración dependiendo del contexto y los modelos actuales aún luchan con contexto, sarcasmo, y múltiples emociones concurrentes (Nandwani & Verma, 2021).
- **Limitaciones en tiempo real e intrusividad:** Algunos métodos de IA aún no son prácticos para uso en tiempo real durante pruebas UX, como procesar señales EEG con modelos complejos puede introducir latencia, dificultando retroalimentación inmediata, también, ciertos sensores (EEG, GSR con electrodos) siguen siendo intrusivos y pueden afectar la naturalidad de la experiencia del usuario durante la evaluación; Si la tecnología de detección interfiere con la UX que se pretende medir, hay un compromiso metodológico, aunque se tiende a métodos más pasivos (cámara, micrófono, *wearables* cómodos), esta sigue siendo una limitación para lograr que la recogida de datos afectivos

no altere el comportamiento normal del usuario (Zhang et al., 2024).

- **Integración con flujos de trabajo de UX:** Desde una perspectiva práctica, otra limitación es cómo integrar estas mediciones en el proceso de evaluación UX existente, ya que los equipos de UX tradicionales quizá no cuenten con expertos en datos o en IA para interpretar resultados, ni con infraestructura para almacenar y procesar grandes volúmenes de video/biometría; Existe una curva de adopción y algunas herramientas comienzan a abstraer esto, pero aún es un cambio de paradigma. Además, se debe demostrar valor añadido claro, porque tras implementar un sistema complejo de medición de emociones los hallazgos no difieren mucho de lo que un buen investigador de UX cualitativo habría notado, puede cuestionarse su costo-beneficio (Ghatoray & Li, 2025).
- **Validación cruzada con métricas UX:** Un desafío final es validar que las métricas afectivas que provee la IA realmente reflejen la calidad de la UX. Por ejemplo, si un modelo indica que los usuarios tuvieron valencia promedio de 0.6 (ligeramente positiva) durante una tarea, ¿cómo se relaciona eso con métricas clásicas como éxito de tarea o puntuación SUS de usabilidad? Aún se está investigando la correlación entre las métricas emocionales y las métricas de UX tradicionales, la falta de correlaciones fuertes podría indicar que estamos midiendo algo diferente o que la IA no capta completamente la experiencia; Un tema abierto y de gran importancia, asegurar que las salidas de IA tengan sentido práctico en la evaluación de UX.

Estos desafíos resaltan que, si bien la IA ofrece herramientas poderosas para evaluar emociones, su aplicación exitosa en UX requiere abordar

cuestiones técnicas (mejorar generalización, explicar modelos), aumentar y diversificar los datos de entrenamiento, y considerar factores humanos (privacidad, interpretación), muchos trabajos de revisión recientes concluyen con recomendaciones de investigación orientadas a superar estos retos, como desarrollar modelos adaptativos, fusionar información de manera más inteligente, generar datasets multimodales más amplios, incorporar XAI y cuantificar incertidumbre en las predicciones (Khare et al., 2024b).

Tendencias emergentes y oportunidades

Mirando hacia el futuro, se vislumbran varias tendencias emergentes que presentan oportunidades de avance, así como nuevos desafíos, en la intersección de IA, emociones y UX, una tendencia clara es profundizar en la integración multimodal y contextual de las evaluaciones, si bien ya se combinan rostro, voz y texto, emergen propuestas para agregar aún más fuentes de datos del usuario, por ejemplo, incorporar sensores fisiológicos (EEG, pulso, dilatación pupilar) para obtener una lectura emocional más profunda que complementa las expresiones externas (Ghatoray & Li, 2025), esto podría brindar oportunidades en aplicaciones donde la UX emocional es crítica (salud digital, entrenamiento, VR), al detectar estados internos como estrés o atención de forma más fiable. De hecho, investigaciones en Brain-Computer Interfaces afectivos sugieren que fusionar señales cerebrales con comportamiento podría mejorar notablemente la detección de emociones sutiles (He et al., 2020). El reto será manejar la complejidad añadida y garantizar la comodidad del usuario al involucrar estos sensores.

Relacionado a lo anterior, se prevé un empuje hacia sistemas de detección emocional en tiempo real totalmente integrados en la interfaz, esto permitiría que las aplicaciones no solo evalúen la

UX *a posteriori*, sino que adapten dinámicamente la experiencia según las emociones del usuario en el momento (Ghatoray & Li, 2025). Por ejemplo, un software tutor podría simplificar las instrucciones si detecta frustración, o un sitio de comercio podría ofrecer asistencia inmediata si percibe confusión o insatisfacción. Lograr esta “UX adaptable emocionalmente” es una oportunidad emocionante para mejorar la personalización y empatía de los sistemas con el usuario, ya existen trabajos exploratorios donde asistentes conversacionales de UX brindan recomendaciones al evaluador basadas en análisis automáticos previos (De Souza Veriscimo et al., 2021), prefigurando una colaboración humano-IA en la evaluación UX, para que esto sea realidad amplia, deberán superarse desafíos de latencia y garantizar que las respuestas del sistema ante las emociones sean apropiadas y beneficiosas.

Otra tendencia emergente es la aplicación de modelos de lenguaje grandes (LLMs) y enfoques de IA generativa para asistir en el análisis de las emociones. Por un lado, modelos como GPT-4 están siendo utilizados para resumir hallazgos de datos de UX, traduciendo los resultados de análisis emocionales en insights narrativos accionables. Un ejemplo reciente integró un asistente tipo chatbot basado en GPT para que, tras recopilar las emociones detectadas en video y texto de una sesión, respondiera preguntas del diseñador de UX en lenguaje natural sobre dónde tuvo problemas el usuario (Ghatoray & Li, 2025), esto facilita que investigadores de UX, que no son expertos en ciencia de datos, aprovechen el poder de IA avanzadas para interpretar patrones emocionales sin perder tiempo en crudos análisis de señal. Además, los LLM ofrecen la oportunidad de analizar grandes colecciones de feedback textual de usuarios (como miles de comentarios) extrayendo temas emocionales recurrentes y correlacionándolos con elementos de la interfaz – tareas que antes hubieran requerido análisis manual intensivo. La tendencia

sugiere que la IA no solo medirá emociones, sino que ayudará a explicar y contextualizar el rol de esas emociones en la experiencia de usuario de manera comprensible para diseñadores y partes interesadas.

En el horizonte también se encuentran innovaciones en técnicas de fusión y aprendizaje, se exploran frameworks híbridos más sofisticados que combinen lo mejor de la fusión temprana y tardía, e incluso enfoques de aprendizaje federado o en el dispositivo para preservar privacidad (de modo que el modelo aprenda de muchos usuarios sin centralizar sus datos emocionales sensibles). Asimismo, hay oportunidades en emplear aprendizaje por refuerzo donde un agente IA modifique elementos de la interfaz y aprenda, mediante retroalimentación emocional del usuario, qué diseños optimizan la experiencia. Este tipo de sistema autónomo podría iterar mejoras de UX de forma personalizada, aunque conlleva desafíos técnicos y éticos.

Por último, surgen consideraciones éticas y de regulación como tendencia obligada, a nivel global, se debate la regulación de los llamados “sistemas de IA de reconocimiento de emociones” dada la sensibilidad de su uso (Por ejemplo, la Unión Europea ha propuesto restringir aplicaciones que puedan violar derechos o resultar en manipulación emocional) (ibdehere, 2023). En el contexto de UX, esto implica que las investigaciones y herramientas deberán adherirse a principios de diseño ético, transparencia en la inferencia emocional y obtención de consentimiento claro de los participantes, que lejos de ser un obstáculo, esto abre la oportunidad de diseñar mejores prácticas y estándares en la industria para el uso responsable de IA afectiva, lo que podría aumentar la confianza en estas tecnologías y facilitar su adopción.

En síntesis, las tendencias apuntan a sistemas de evaluación de UX cada vez más inteligentes,

omnipresentes y centrados en el humano, que aprovechan IA multimodal en tiempo real y análisis avanzado para comprender al usuario a un nivel sin precedentes, las oportunidades para mejorar la personalización, accesibilidad y eficacia de las interfaces mediante la detección emocional son enormes, no obstante, alcanzar esa visión requerirá afrontar retos técnicos (robustez, tiempo real), ampliar repertorios emocionales y navegar cuidadosamente implicaciones éticas, la convergencia de disciplinas desde la ingeniería de datos hasta la psicología y el diseño será crucial en esta próxima etapa.

Conclusiones

La revisión integrativa de trabajos de los últimos siete años muestra que la IA se ha convertido en un aliado prometedor para enriquecer la evaluación de UX mediante el análisis de emociones y sentimientos de los usuarios, se han identificado los métodos más empleados (Visión por computador, audio, PLN, sensores) y constatado que su combinación multimodal potencia la eficacia de la medición emocional en HCI (Razzaq et al., 2023).

La aplicación de IA en este ámbito ha evolucionado rápidamente, pasando de enfoques experimentales a soluciones más maduras que se apoyan en aprendizaje profundo y grandes volúmenes de datos, expandiendo el alcance del análisis emocional a entornos de uso reales y datos no estructurados (Guo et al., 2024) (Alabduljabbar, 2024). Asimismo, se ha destacado los principales recursos (datasets, métricas) que sostienen estas investigaciones, junto con los desafíos vigentes, técnicos, humanos y éticos, que limitan por ahora su máximo potencial (Santos & Digiampietri, 2024), que lejos de ser obstáculos insuperables, estos retos proporcionan una hoja de ruta para futuras investigaciones para mejorar la robustez y rapidez de los modelos, abarcar mayor diversidad

emocional y asegurar equidad y privacidad en las inferencias.

En definitiva, integrar el análisis automatizado de emociones en la evaluación de UX ofrece una visión más rica y holística de la experiencia del usuario, incorporando la dimensión afectiva que tradicionalmente era difícil de captar objetivamente, esto debe permitir a diseñadores e investigadores comprender no solo qué hace el usuario, sino cómo se siente mientras lo hace, lo cual es invaluable para crear productos más satisfactorios, intuitivos y adaptados. Las tendencias emergentes indican que esta línea de trabajo seguirá creciendo, con IA cada vez más sofisticada capaz de empatizar con el usuario y apoyar tanto la evaluación como la optimización continua de las interfaces (Ghatoray & Li, 2025), en un campo tan dinámico, mantener el rigor científico y la centricidad en el ser humano será crucial: los modelos deberán ser validados exhaustivamente y usados como herramientas de apoyo para la toma de decisiones de diseño, y no como sustitutos absolutos de la comprensión humana. Con este equilibrio, la IA aplicada al análisis de emociones en UX se perfila como una pieza clave en la siguiente generación de metodologías de evaluación, brindando oportunidades sin precedente para innovar en la creación de experiencias de usuario más afectivamente conscientes, personalizadas y satisfactorias.

Referencias bibliográficas

- Alabduljabbar, R. (2024). User-centric AI: Evaluating the usability of generative AI applications through user reviews on app stores. *PeerJ Computer Science*, 10, e2421. <https://doi.org/10.7717/peerj-cs.2421>
- Alonazi, M. (2023). Analyzing sentiment in terms of online feedback on top of users' experiences. *International Journal of Advanced Computer Science and Applications*, 14(11). <https://doi.org/10.14569/IJACSA.2023.0141114>
- Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S., & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42(4), 335-359. <https://doi.org/10.1007/s10579-008-9076-6>
- De Souza Veriscimo, E., Bernardes Júnior, J. L., & Digiampietri, L. A. (2021). Facial emotion recognition in UX evaluation: A systematic review. En M. Kurosu (Ed.), *Human-Computer Interaction. Theory, methods and tools* (Vol. 12762, pp. 521-534). Springer International Publishing. https://doi.org/10.1007/978-3-030-78462-1_40
- Fernández-Ordóñez, J. M., Jiménez, L. E. M., Torres-Carrión, P., Barba-Guamán, L., Rodríguez-Morales, G.. (2019). Experiencia afectiva usuario en ambientes con inteligencia artificial, sensores biométricos y/o recursos digitales accesibles: Una revisión sistemática de literatura. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 35, 35-53. <https://doi.org/10.17013/risti.35.35-53>
- Galindo Monfil, A. R., Rojano Cáceres, J. R., & Mezura Godoy, C. (2025). Evaluación de las emociones y la satisfacción del usuario en el contexto de la interacción humano computadora: Un mapeo sistemático. *ReCIBE, Revista electrónica de Computación, Informática, Biomédica y Electrónica*, 14(1), C9-12. <https://doi.org/10.32870/recibe.v14i1.401>
- Ghatoray, S. K., & Li, Y. (2025). *Automated UX Insights from User Research Videos by Integrating Facial Emotion and Text Sentiment* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.2503.22510>

Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., ... Bengio, Y. (2013). *Challenges in representation learning: A report on three machine learning contests* (Versión 1). arXiv. <https://doi.org/10.48550/ARXIV.1307.0414>

Guo, R., Guo, H., Wang, L., Chen, M., Yang, D., & Li, B. (2024). Development and application of emotion recognition technology—A systematic literature review. *BMC Psychology*, 12(1), 95. <https://doi.org/10.1186/s40359-024-01581-4>

He, Z., Li, Z., Yang, F., Wang, L., Li, J., Zhou, C., & Pan, J. (2020). Advances in multimodal emotion recognition based on brain-computer interfaces. *Brain Sciences*, 10(10), 687. <https://doi.org/10.3390/brainsci10100687>

Hernandez Perez, L. A. (2022). Método para la detección de emociones mediante el análisis de expresión facial en videos de evaluación de la experiencia de usuario. <https://rinacional.tecnm.mx/jspui/handle/TecNM/4841>

Ibdehere. (2023, junio 19). Inteligencia artificial y reconocimiento biométrico de emociones: Una valoración a la luz de las enmiendas del Parlamento europeo a la Ley de Inteligencia Artificial. *UNA MIRADA CRÍTICA A LAS RELACIONES LABORALES*. <https://ignasibeltran.com/2023/06/19/inteligencia-artificial-y-reconocimiento-biometrico-de-emociones-una-valoracion-a-la-luz-de-las-enmiendas-del-parlamento-europeo-a-la-ley-de-inteligencia-artificial/>

Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R. (2024). Emotion recognition and artificial intelligence: A systematic review (2014–

2023) and research recommendations. *Information Fusion*, 102, 102019. <https://doi.org/10.1016/j.inffus.2023.102019>

Koelstra, S., Muhl, C., Soleymani, M., Jong-Seok Lee, Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A Database for emotion analysis: Using physiological signals. *IEEE transactions on affective computing*, 3(1), 18-31. <https://doi.org/10.1109/T-AFFC.2011.15>

Leppich, D., Bieber, C., Proschek, K., Harms, P., & Schubert, U. (2023). DUX: A dataset of user interactions and user emotions. *I-Com*, 22(2), 101-123. <https://doi.org/10.1515/icom-2023-0014>

Liu, F. (2024). Artificial intelligence in emotion quantification: A prospective overview. *CAAI Artificial Intelligence Research*, 9150040. <https://doi.org/10.26599/AIR.2024.9150040>

Livingstone, S. R., & Russo, F. A. (2018). The ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE*, 13(5), e0196391. <https://doi.org/10.1371/journal.pone.0196391>

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). AffectNet: A Database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1), 18-31. <https://doi.org/10.1109/TAFFC.2017.2740923>

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 81. <https://doi.org/10.1007/s13278-021-00776-6>

Pereira, R., Mendes, C., Ribeiro, J., Ribeiro, R., Miragaia, R., Rodrigues, N., Costa, N., & Pereira, A. (2024). Systematic review of emotion

detection with computer vision and deep learning. *Sensors*, 24(11), 3484. <https://doi.org/10.3390/s24113484>.

Plisiecki, H., Lenartowicz, P., Flakus, M., & Pokropek, A. (2025). High risk of political bias in black box emotion inference models. *Scientific Reports*, 15(1), 6028. <https://doi.org/10.1038/s41598-025-86766-6>

Razzaq, M. A., Hussain, J., Bang, J., Hua, C.-H., Satti, F. A., Rehman, U. U., Bilal, H. S. M., Kim, S. T., & Lee, S. (2023). A hybrid multimodal emotion recognition framework for UX evaluation using generalized mixture functions. *Sensors*, 23(9), 4373. <https://doi.org/10.3390/s23094373>

Santos, B. L., & Digiampietri, L. A. (2024). User experience evaluation using machine learning and facial expressions: A systematic review. *Anais do XXI Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2024)*, 930-941. <https://doi.org/10.5753/eniac.2024.245150>

Verhoef, T., & Fosch-Villaronga, E. (2023). Towards affective computing that works for everyone. *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1-8. <https://doi.org/10.1109/ACII59096.2023.10388169>

Zhang, Z., Fort, J. M., & Giménez Mateu, L. (2024). Mini review: challenges in EEG emotion recognition. *frontiers in psychology*, 14, 1289816. <https://doi.org/10.3389/fpsyg.2023.1289816>