



Impacto de los Agentes de Inteligencia Artificial en las Aplicaciones del Siglo XXI y su Influencia en el Desempeño y la Utilidad

Impact of Artificial Intelligence Agents on 21st Century Applications and Their Influence on Performance and Utility

Dalton Andrés Pazmiño Pérez¹, Christian Miguel Muñoz Carrasco²

¹Universidad Bolivariana del Ecuador, Guayaquil, Ecuador, daltonpazmino@ube.edu.ec

²Universidad Bolivariana del Ecuador, Guayaquil, Ecuador, christianmunoz@ube.edu.ec

Resumen.

El presente artículo científico analiza el impacto de los agentes de inteligencia artificial (IA) en las aplicaciones del siglo XXI, así como su influencia directa en el desempeño y la utilidad de los sistemas digitales en distintos contextos organizacionales y sociales. A partir de una revisión teórica y un enfoque analítico, se examina cómo la incorporación de agentes inteligentes ha transformado los procesos de automatización, toma de decisiones, eficiencia operativa y experiencia del usuario. Los resultados evidencian que la implementación adecuada de agentes de IA contribuye significativamente a la optimización del rendimiento, la reducción de errores humanos y la generación de valor estratégico, siempre que su adopción esté alineada con criterios éticos, técnicos y normativos.

Palabras clave: inteligencia artificial, agentes inteligentes, desempeño, utilidad, aplicaciones digitales.

Abstract.

This scientific article analyzes the impact of artificial intelligence (AI) agents on 21st-century applications and their direct influence on system performance and utility across organizational and social contexts. Through a theoretical review and analytical approach, it examines how intelligent agents have transformed automation processes, decision-making, operational efficiency, and user experience. The findings show that proper implementation of AI agents significantly contributes to performance optimization, reduction of human error, and strategic value creation, provided their adoption aligns with ethical, technical, and regulatory criteria.

Keywords: artificial intelligence, intelligent agents, performance, utility, digital applications.

1. Introducción

La consolidación de los agentes de inteligencia artificial en las aplicaciones del siglo XXI constituye uno de los fenómenos tecnológicos más influyentes y, al mismo tiempo, más complejos de la transformación digital contemporánea. En un entorno caracterizado por la automatización intensiva, la



toma de decisiones basada en datos y la creciente interdependencia entre sistemas socio-técnicos, los agentes de IA han pasado de ser herramientas experimentales a convertirse en componentes estratégicos en sectores clave como la industria, la salud, la educación, la seguridad, las finanzas y la administración pública. Su capacidad para percibir el entorno, razonar, aprender y actuar de manera autónoma ha redefinido no solo los modelos de operación de las organizaciones, sino también las expectativas de eficiencia, calidad y valor por parte de los usuarios finales. En este contexto, analizar el impacto real de estos agentes sobre el desempeño y la utilidad de las aplicaciones digitales resulta una cuestión de alta relevancia científica, económica e institucional.

Desde una perspectiva ideal, la incorporación de agentes de inteligencia artificial en las aplicaciones modernas debería traducirse en sistemas más eficientes, precisos, adaptativos y centrados en el usuario. Bajo este supuesto, la automatización inteligente permitiría optimizar procesos, reducir errores humanos, mejorar la calidad de las decisiones y maximizar la utilidad percibida de los servicios tecnológicos. Sin embargo, la realidad evidencia una brecha significativa entre este escenario deseado y los resultados obtenidos en múltiples contextos de implementación. En numerosos casos, la adopción de agentes de IA no genera mejoras proporcionales en el desempeño de las aplicaciones, o incluso introduce nuevas fuentes de complejidad, opacidad y dependencia tecnológica. Esta divergencia plantea un problema de investigación relevante: ¿en qué medida los agentes de inteligencia artificial influyen efectivamente en el desempeño y la utilidad de las aplicaciones del siglo XXI, y bajo qué condiciones dicha influencia se materializa de forma positiva?

Diversos estudios han abordado parcialmente esta problemática desde enfoques técnicos, computacionales y de ingeniería de software, centrándose en el desarrollo de algoritmos, arquitecturas inteligentes o métricas de eficiencia específicas. Investigaciones previas han demostrado avances significativos en áreas como el aprendizaje automático, los sistemas multiagente y la inteligencia artificial distribuida; no obstante, gran parte de estos trabajos se ha concentrado en el rendimiento algorítmico aislado, sin integrar de manera suficiente variables organizacionales, contextuales y de utilidad práctica. Asimismo, otros enfoques han evaluado la aceptación de la IA desde la perspectiva del usuario, pero sin establecer una relación estructurada entre el tipo de agente implementado, el desempeño del sistema y el valor funcional generado. Como resultado, persiste una fragmentación conceptual que limita la comprensión integral del impacto real de los agentes de IA en aplicaciones contemporáneas.

Las consecuencias de esta falta de comprensión son significativas y multidimensionales. A nivel organizacional, una implementación inadecuada de agentes de inteligencia artificial puede derivar en inversiones tecnológicas poco eficientes, resistencia al cambio, dependencia excesiva de sistemas opacos y deterioro de la confianza en las soluciones digitales. En el ámbito social y económico, la baja utilidad o el desempeño subóptimo de aplicaciones basadas en IA puede profundizar brechas digitales, afectar la calidad de los servicios públicos y privados, y generar percepciones negativas sobre la inteligencia artificial como herramienta de progreso. Estas implicaciones refuerzan la necesidad de estudios que trasciendan la visión puramente técnica y aborden el fenómeno desde un enfoque analítico y crítico.

En este sentido, se identifica un vacío de conocimiento relacionado con la ausencia de modelos integrales que permitan evaluar, de manera sistemática, la relación entre los agentes de inteligencia artificial, el desempeño de las aplicaciones y la utilidad generada en contextos reales de uso. Falta profundizar en enfoques que consideren múltiples criterios de evaluación, así como la gestión de la incertidumbre, la ambigüedad y la heterogeneidad de los entornos digitales actuales. Este estudio se propone contribuir a dicho vacío mediante un análisis conceptual que articula el comportamiento de los agentes inteligentes con indicadores de desempeño y utilidad, incorporando una visión holística que supera los enfoques tradicionales centrados exclusivamente en la eficiencia técnica.

En síntesis, apoyándose en aportes teóricos clásicos y contemporáneos sobre inteligencia artificial y agentes inteligentes, esta investigación se diferencia de los estudios previos al integrar de forma

explícita el análisis del desempeño y la utilidad como dimensiones interdependientes. El trabajo introduce un marco conceptual que permite comprender cómo y por qué los agentes de inteligencia artificial influyen en las aplicaciones del siglo XXI, ofreciendo una base analítica sólida para futuras investigaciones y para la toma de decisiones estratégicas en entornos académicos, organizacionales e institucionales.

2. Planteamiento del problema

A pesar de la adopción creciente de agentes de inteligencia artificial en aplicaciones contemporáneas, no siempre se evidencia una mejora proporcional en el desempeño o en la utilidad percibida por los usuarios finales. En muchos casos, la falta de una adecuada integración, capacitación del talento humano o evaluación de impacto limita los beneficios esperados. Esto plantea la necesidad de analizar de manera sistemática cómo y en qué medida los agentes de IA influyen en el rendimiento y la utilidad real de las aplicaciones del siglo XXI.

3. Objetivos del estudio

3.1 Objetivo general

El propósito central de esta investigación es analizar de manera integral el impacto de los agentes de inteligencia artificial en las aplicaciones del siglo XXI y examinar cómo su incorporación influye en el desempeño operativo y en la utilidad percibida de dichos sistemas en contextos organizacionales y sociales. A diferencia de enfoques centrados exclusivamente en la eficiencia técnica o en el rendimiento algorítmico, este estudio busca articular una visión analítica que considere simultáneamente dimensiones tecnológicas, funcionales y de valor para el usuario.

3.2 Objetivos Específicos

1. **Examinar la degradación por envejecimiento** en las respuestas de los agentes de IA, utilizando el modelo AAS para entender cómo la precisión de los sistemas se transforma tras interacciones repetidas, más allá del tiempo cronológico.
2. **Integrar la incertidumbre y la contradicción** en la evaluación de la IA mediante el uso de **conjuntos neutrosóficos**, permitiéndonos capturar esos «matices grises» que la lógica binaria tradicional suele ignorar al enfrentarse a información incompleta.
3. **Valorar el giro hacia una «IA centrada en los datos» (Data-Centric AI)**, analizando cómo la curaduría inteligente de la información puede ser un antídoto contra la **deuda técnica** y la inestabilidad que afecta a los modelos actuales.
4. **Analizar la utilidad real de los modelos de lenguaje (LLM)** en entornos profesionales como la ingeniería y el desarrollo de software, observando cómo la **ingeniería de instrucciones (prompt engineering)** potencia la productividad y la calidad del código generado.
5. **Proponer un marco evaluativo estable y humano** que permita comparar el desempeño de la IA de forma justa entre distintas arquitecturas, asegurando que las métricas finales sean **normalizadas y resistentes** a las perturbaciones del entorno operativo

3.3 Justificación del estudio



La relevancia de este estudio radica en la creciente dependencia de los agentes de inteligencia artificial como elementos estructurales de las aplicaciones digitales actuales. Si bien la literatura reconoce su potencial transformador, persiste una brecha entre la adopción acelerada de estas tecnologías y la comprensión profunda de sus efectos reales sobre el desempeño y la utilidad de los sistemas en escenarios de uso concreto. Desde el punto de vista académico, la investigación aporta un enfoque integrador que conecta la teoría de agentes inteligentes con marcos de evaluación del desempeño y la utilidad, contribuyendo a superar la fragmentación conceptual observada en estudios previos.

Desde una perspectiva práctica, los resultados del estudio ofrecen insumos relevantes para la toma de decisiones estratégicas en organizaciones públicas y privadas que implementan soluciones basadas en inteligencia artificial. Comprender cómo y bajo qué condiciones los agentes inteligentes generan valor permite reducir riesgos asociados a inversiones tecnológicas, mejorar la alineación entre objetivos organizacionales y capacidades técnicas, y fortalecer la confianza de los usuarios en las aplicaciones basadas en IA. En este sentido, la investigación se orienta a ofrecer una respuesta analítica a un problema real y recurrente en los procesos de transformación digital contemporáneos.

3.4 Enfoque del estudio desde el modelo CARS

Siguiendo el modelo CARS (Create A Research Space), este trabajo establece, en primer lugar, el territorio de investigación al reconocer el impacto creciente de los agentes de inteligencia artificial en las aplicaciones del siglo XXI y su papel central en la automatización, la toma de decisiones y la optimización de procesos en múltiples sectores. La literatura reciente coincide en señalar que estos agentes constituyen un eje estratégico de la innovación tecnológica, con implicaciones económicas, sociales e institucionales de amplio alcance [1].

En segundo lugar, se identifica el nicho de investigación al evidenciar que, a pesar del volumen de estudios existentes, aún resulta limitada la comprensión integrada de cómo los agentes de IA influyen simultáneamente en el desempeño de las aplicaciones y en su utilidad práctica. La mayoría de los trabajos se concentra en métricas técnicas aisladas o en análisis de aceptación tecnológica, sin articular ambas dimensiones en un marco analítico coherente.

Finalmente, este estudio ocupa dicho nicho al proponer un análisis conceptual que vincula explícitamente los agentes de inteligencia artificial con indicadores de desempeño y utilidad, ofreciendo una perspectiva holística que amplía y complementa los enfoques previos. De este modo, la investigación no solo contribuye al debate académico, sino que también establece una base teórica y analítica para el desarrollo de aplicaciones de inteligencia artificial más efectivas, útiles y alineadas con las necesidades reales de los usuarios.

4. Revisión de la literatura

4.1 Inicio general y contextualización

La expansión de los agentes de inteligencia artificial en las aplicaciones del siglo XXI ha redefinido los criterios tradicionales con los que se evalúa el desempeño y la utilidad de los sistemas digitales. Más allá de su capacidad computacional, estos agentes operan hoy como entidades socio-técnicas integradas en procesos organizacionales complejos, donde interactúan de forma continua con usuarios humanos, datos heterogéneos y entornos cambiantes. Desde una perspectiva científica, este fenómeno plantea el desafío de comprender no solo cómo funcionan los agentes de IA, sino cómo evolucionan, se degradan o se estabilizan a lo largo del tiempo y de las interacciones. En el plano práctico, la relevancia es aún mayor: organizaciones públicas y privadas dependen crecientemente de estos sistemas para tareas críticas, lo que convierte al desempeño sostenido y a la utilidad real en dimensiones estratégicas del valor tecnológico [1].

4.2 Síntesis crítica de la literatura relevante



La literatura clásica sobre agentes inteligentes, representada por trabajos fundacionales como los de [1], establece las bases conceptuales sobre autonomía, percepción, razonamiento y acción. Estos estudios, de carácter principalmente teórico, han contribuido a definir arquitecturas y propiedades formales de los agentes, pero presentan limitaciones al abordar su comportamiento dinámico en escenarios prolongados de uso. Investigaciones más recientes han comenzado a cuestionar esta visión estática, incorporando enfoques orientados al desempeño sostenido y a la interacción continua con entornos reales.

En este contexto, [1] introducen el modelo AAS (Agent Aging and Saturation) con el objetivo de analizar cómo los agentes de IA experimentan transformaciones funcionales tras múltiples ciclos de interacción. Mediante simulaciones controladas, los autores demuestran que la precisión de las respuestas no se mantiene constante, sino que puede degradarse o saturarse como resultado de ajustes acumulativos y retroalimentación reiterada. Si bien el estudio aporta un marco innovador para comprender el envejecimiento funcional, su principal limitación radica en la falta de validación en entornos organizacionales reales, donde las variables contextuales son más complejas.

Por otra parte, la evaluación del desempeño de los agentes de IA ha sido tradicionalmente abordada mediante métricas deterministas, lo que ha motivado críticas desde corrientes alternativas de pensamiento. Smarandache [11] propone la teoría de los conjuntos neutrosóficos como un modelo capaz de representar simultáneamente verdad, falsedad e indeterminación, superando las restricciones de la lógica binaria y difusa. Estudios aplicados, como los de Salmerón y Smarandache [9], utilizan metodologías neutrosóficas para evaluar sistemas de decisión inteligente bajo condiciones de información incompleta, evidenciando una mayor capacidad explicativa frente a enfoques clásicos. No obstante, estos trabajos suelen centrarse en sistemas de apoyo a la decisión, dejando abierta la pregunta sobre su aplicabilidad directa a agentes de IA de propósito general.

El giro hacia la Inteligencia Artificial centrada en los datos (Data-Centric AI) constituye otro eje relevante del estado del arte. Ng [1] plantea que la mayor parte de los fallos de desempeño en sistemas de IA no proviene de los algoritmos, sino de la calidad deficiente de los datos. Este enfoque ha sido respaldado empíricamente por estudios como los de Polyzotis et al. [2], quienes analizan cómo errores de etiquetado, sesgos y falta de gobernanza de datos generan deuda técnica y degradación progresiva del rendimiento. Si bien estos trabajos aportan evidencia sólida sobre la centralidad de los datos, tienden a subestimar la interacción entre la curaduría de la información y el envejecimiento funcional de los agentes.

En el ámbito de los modelos de lenguaje de gran escala (LLM), [12] examinan su impacto transversal en múltiples dominios profesionales, destacando tanto su potencial como sus riesgos. Investigaciones posteriores, como las de [3], analizan específicamente el rol de la ingeniería de instrucciones (prompt engineering) en el desarrollo de software, mostrando incrementos medibles en productividad y calidad del código. Sin embargo, estos estudios reconocen limitaciones importantes, entre ellas la dificultad para evaluar la utilidad real de los LLM más allá de métricas de corto plazo y benchmarks controlados.

4.3 Aporte de la literatura al problema de investigación

En conjunto, los estudios revisados contribuyen a la comprensión del problema central al ofrecer perspectivas parciales sobre el desempeño y la utilidad de los agentes de IA. El modelo AAS aporta una visión dinámica del envejecimiento funcional; la neutrosofía introduce herramientas conceptuales para gestionar la incertidumbre; el enfoque Data-Centric AI redefine las prioridades en el diseño de sistemas; y la literatura sobre LLM evidencia el valor práctico de la IA cuando se integra en flujos de trabajo humanos. No obstante, estos aportes permanecen en gran medida desconectados entre sí.

4.4 Comparación de resultados y vacíos de conocimiento



La comparación crítica de los resultados revela patrones claros y contradicciones persistentes. Mientras algunos estudios reportan mejoras sostenidas en el desempeño mediante ajustes algorítmicos, otros evidencian degradación progresiva debido a la acumulación de interacciones y a la calidad deficiente de los datos. Asimismo, existe consenso en la necesidad de métricas más ricas, pero desacuerdo sobre cómo operacionalizarlas. El principal vacío identificado es la ausencia de un marco evaluativo que integre envejecimiento funcional, incertidumbre neutrosófica, centralidad de los datos y utilidad práctica en un mismo modelo analítico.

4.5 Evaluación crítica del estado del arte

Desde una perspectiva crítica, puede afirmarse que la literatura actual no responde de manera integral a los objetivos del campo. Predomina una fragmentación metodológica y conceptual que dificulta la comparación justa entre arquitecturas de IA y limita la evaluación de su utilidad real en entornos perturbados. Aspectos como la estabilidad del desempeño, la resistencia a la incertidumbre y el valor humano generado permanecen insuficientemente abordados.

4.6 Relación con los objetivos del estudio

Este estudio se posiciona explícitamente para llenar la brecha identificada. Al vincular el análisis del envejecimiento funcional de los agentes (modelo AAS), la evaluación bajo incertidumbre mediante conjuntos neutrosóficos, el paradigma Data-Centric AI y la utilidad real de los LLM en contextos profesionales, la investigación propone un marco evaluativo estable, humano y normalizado. De este modo, la revisión de la literatura no solo fundamenta las preguntas de investigación, sino que justifica la necesidad de un enfoque integrador capaz de avanzar el estado del arte en la evaluación del impacto de los agentes de inteligencia artificial en las aplicaciones del siglo XXI.

5. Materiales y Métodos

Diseño del estudio

El estudio se desarrolló bajo un enfoque aplicado, descriptivo y analítico, sustentado en la utilización combinada de los métodos de decisión multicriterio Analytic Hierarchy Process (AHP) y Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). Este diseño metodológico se aplicó en el contexto del sector tecnológico y de ingeniería de software, específicamente en entornos organizacionales donde los agentes de inteligencia artificial y los modelos de lenguaje de gran escala son utilizados como apoyo a procesos de desarrollo, análisis y toma de decisiones. El periodo de ejecución del estudio correspondió a los meses de enero a marzo de 2025.

La selección del diseño AHP–TOPSIS respondió a la necesidad de evaluar el desempeño y la utilidad de la inteligencia artificial desde una perspectiva estructurada, transparente y comparativa. AHP permitió descomponer el problema complejo en criterios jerárquicos y asignar pesos relativos de manera consistente a partir del juicio experto, mientras que TOPSIS facilitó la comparación objetiva de alternativas mediante su proximidad a una solución ideal positiva y su alejamiento de una solución ideal negativa. Esta combinación metodológica ha sido ampliamente reconocida por su robustez para la evaluación de sistemas complejos bajo múltiples criterios, especialmente cuando intervienen factores cualitativos y cuantitativos simultáneamente (Saaty, 2008; Hwang & Yoon, 1981).

Nivel teórico

El estudio se sustentó en un nivel teórico orientado a la integración de la teoría de la decisión multicriterio con enfoques contemporáneos de evaluación de sistemas de inteligencia artificial. Se adoptaron como marcos conceptuales el Proceso Analítico Jerárquico (AHP) propuesto por Saaty y el método TOPSIS desarrollado por Hwang y Yoon, junto con aportes recientes sobre envejecimiento funcional de agentes inteligentes (AAS), inteligencia artificial centrada en los datos (Data-Centric AI)



y lógica neutrosófica. Este nivel permitió definir los criterios de evaluación y estructurar conceptualmente las relaciones jerárquicas necesarias para el modelamiento del problema de decisión.

Nivel empírico

Desde el nivel empírico, la investigación se desarrolló como un estudio aplicado, descriptivo y analítico, orientado a evaluar la utilidad de agentes de inteligencia artificial en contextos profesionales asociados a procesos de transformación digital. El trabajo empírico se llevó a cabo durante el período comprendido entre marzo y junio de 2025, considerando escenarios reales de uso y toma de decisiones tecnológicas. La información empírica se obtuvo mediante el juicio experto, lo que permitió capturar conocimiento especializado que difícilmente puede ser representado por datos puramente observacionales.

Población y muestra

La población estuvo conformada por profesionales con experiencia en ingeniería, desarrollo de software, analítica de datos y evaluación de sistemas de inteligencia artificial. La muestra se seleccionó mediante muestreo no probabilístico por juicio y estuvo integrada por informantes clave que cumplían criterios de inclusión tales como experiencia profesional mínima de cinco años, participación en proyectos de IA y conocimiento de metodologías de evaluación multicriterio. Este enfoque fue coherente con la naturaleza del método AHP, que prioriza la calidad del juicio experto sobre el tamaño muestral.

Nivel matemático-estadístico

El nivel matemático-estadístico comprendió la formalización y el procesamiento de la información mediante los métodos AHP y TOPSIS. En el AHP se construyeron matrices de comparación por pares, se calculó el vector propio principal y se verificó la consistencia de los juicios mediante el índice de consistencia, asegurando valores inferiores a 0.10. Posteriormente, el método TOPSIS permitió la construcción de la matriz de decisión, su normalización, la determinación de las soluciones ideal positiva y negativa, el cálculo de las distancias euclidianas y la obtención del índice de cercanía relativa (C_i), con el fin de establecer el ranking final de alternativas.

Método neutrosófico

De forma complementaria, se incorporó un enfoque neutrosófico para modelar explícitamente la incertidumbre, la indeterminación y la contradicción inherentes a la evaluación de sistemas de IA. Cada criterio fue interpretado a partir de tres componentes: grado de verdad (T), grado de indeterminación (I) y grado de falsedad (F). Este enfoque permitió capturar matices evaluativos que la lógica binaria tradicional no logra representar, fortaleciendo la estabilidad conceptual y la robustez interpretativa del modelo AHP–TOPSIS.

Participantes o informantes

La población del estudio estuvo conformada por 15 expertos seleccionados mediante un muestreo no probabilístico por juicio. Los participantes contaban con experiencia profesional comprobada en áreas relacionadas con inteligencia artificial, ingeniería de software, análisis de sistemas y gestión tecnológica. Los criterios de inclusión consideraron un mínimo de cinco años de experiencia laboral en el área, participación directa en proyectos que involucraran agentes de IA o modelos de lenguaje, y conocimiento previo en procesos de evaluación tecnológica o toma de decisiones técnicas. Este perfil permitió garantizar la validez del juicio experto empleado en la ponderación de los criterios y en la valoración de las alternativas analizadas.

Instrumentos y materiales



Para la aplicación del método AHP se utilizó la herramienta AHP Calculator desarrollada por Business Performance Management Singapore (BPMSG), disponible en <https://bpmsg.com/ahp-calculator/>. Esta herramienta fue seleccionada por su accesibilidad, facilidad de uso y precisión en el cálculo automático de matrices de comparación por pares, vectores propios y ratios de consistencia. Para la aplicación del método TOPSIS se empleó la plataforma Decision Radar, accesible en <https://www.decision-radar.com/topsis/>, la cual permitió construir matrices de decisión, normalizar los datos y calcular de forma automatizada las distancias a las soluciones ideales.

Ambas herramientas facilitaron la visualización gráfica de los resultados intermedios y finales, lo que contribuyó a la trazabilidad del proceso metodológico. Como respaldo del análisis, se generaron imágenes y registros digitales correspondientes a la matriz AHP de ponderación de criterios y al ranking final obtenido mediante TOPSIS, los cuales fueron conservados como evidencias del procedimiento aplicado.

Procedimiento del estudio

El procedimiento metodológico se estructuró en cuatro etapas claramente diferenciadas. En la primera etapa se realizó la identificación y validación de los criterios de evaluación, a partir de la revisión de la literatura especializada y de sesiones de consenso con los expertos participantes. Esta etapa permitió definir un conjunto de criterios coherentes con los objetivos del estudio y relevantes para evaluar el desempeño y la utilidad de la inteligencia artificial en entornos profesionales.

En la segunda etapa se aplicó el método AHP para la determinación de los pesos relativos de cada criterio. Los expertos realizaron comparaciones por pares utilizando la escala fundamental propuesta por Saaty, lo que permitió construir la matriz de comparación. Posteriormente, se calcularon los vectores de prioridad y se verificó la consistencia de los juicios mediante el índice de consistencia (CR), asegurando que este fuera inferior a 0,10, conforme a los estándares metodológicos aceptados.

La tercera etapa correspondió a la aplicación del método TOPSIS. En esta fase se construyó la matriz de decisión a partir de las valoraciones de las alternativas en función de los criterios definidos. La matriz fue normalizada y ponderada utilizando los pesos obtenidos mediante AHP. A continuación, se determinaron las soluciones ideal positiva y negativa, se calcularon las distancias euclidianas a dichas soluciones y se obtuvo el índice de cercanía relativa (C_i) para cada alternativa, siguiendo el procedimiento clásico descrito por Hwang y Yoon.

Finalmente, en la cuarta etapa se realizó la integración y validación del proceso AHP–TOPSIS. Esta etapa permitió verificar la coherencia interna de los resultados, asegurar la trazabilidad de las decisiones metodológicas y garantizar la replicabilidad del estudio en contextos similares. La integración de ambos métodos fortaleció la solidez del análisis multicriterio aplicado.

Medidas y variables de salida

Las principales variables de salida del estudio fueron los pesos relativos de los criterios obtenidos mediante AHP y el ranking final de las alternativas generado por TOPSIS. Los criterios evaluados incluyeron: precisión de las respuestas, estabilidad del desempeño, manejo de la incertidumbre, calidad de los datos, utilidad percibida en entornos profesionales y adaptabilidad del sistema. Estas variables permitieron capturar de manera estructurada las dimensiones clave del desempeño y la utilidad de la inteligencia artificial.

Análisis de datos

El análisis de los datos se realizó de forma automatizada mediante las herramientas AHP Calculator y Decision Radar, complementándose con la integración y sistematización de la información en hojas de cálculo de Microsoft Excel. Se verificó la consistencia de los juicios expertos, la correcta aplicación de



las fórmulas multicriterio y la coherencia de los resultados intermedios. Todo el proceso se desarrolló bajo principios de rigor metodológico, validez interna y transparencia analítica, garantizando que las decisiones tomadas pudieran ser auditadas y replicadas en futuras investigaciones.

6. Resultados

Los resultados obtenidos mediante la aplicación integrada de los métodos AHP y TOPSIS permitieron analizar de manera estructurada y comparativa el impacto de distintos criterios asociados al desempeño y la utilidad de los agentes de inteligencia artificial en aplicaciones contemporáneas. En coherencia con los objetivos planteados, el análisis se centró en evaluar la degradación por envejecimiento funcional de los agentes, la gestión de la indeterminación, la calidad de los datos bajo el enfoque Data-Centric AI, la eficacia en contextos de ingeniería y la estabilidad de los marcos evaluativos.

En primer lugar, los resultados del AHP mostraron una jerarquización claramente asimétrica de los criterios evaluados. El criterio "Degradación por Envejecimiento (AAS)" alcanzó un peso relativo de 64.4 %, posicionándose de forma dominante frente al resto. Este hallazgo sugiere que, desde la perspectiva de los expertos, la pérdida progresiva de precisión y coherencia en las respuestas de los agentes de IA constituye el principal factor que condiciona su utilidad real, incluso por encima de aspectos tradicionalmente priorizados como la calidad de los datos o la estabilidad del sistema. Este resultado se alinea con la teoría del envejecimiento funcional de sistemas adaptativos, la cual postula que la exposición reiterada a interacciones complejas puede inducir desviaciones acumulativas en los modelos de decisión, independientemente del tiempo cronológico transcurrido [2, 1].

El segundo criterio en importancia fue la "Gestión de la Indeterminación", con un peso de 16.4 %. Este resultado refuerza los planteamientos de la teoría neutrosófica, según la cual los sistemas inteligentes operan de forma más realista cuando incorporan grados simultáneos de verdad, falsedad e indeterminación, especialmente en entornos con información incompleta o contradictoria [11]. La distancia significativa entre este criterio y el primero sugiere que, si bien la incertidumbre es un componente crítico, su impacto es percibido como subordinado al deterioro funcional acumulativo de los agentes.

Los criterios "Calidad de Datos (Data-Centric AI)", "Eficacia en Ingeniería" y "Estabilidad y Normalización" obtuvieron pesos de 9.5 %, 6.4 % y 3.3 %, respectivamente. Estos valores indican que, aunque relevantes, estos factores fueron considerados de menor influencia relativa en la evaluación global. Este resultado contrasta parcialmente con la literatura reciente que enfatiza el giro hacia una IA centrada en los datos como principal estrategia para mitigar fallos de generalización y deuda técnica [1, 2]. La diferencia observada puede explicarse por el enfoque específico del estudio, centrado en la degradación operativa y en la estabilidad evaluativa a largo plazo, más que en el rendimiento inicial de los modelos.

Desde la perspectiva de TOPSIS, la matriz de decisión normalizada y el cálculo de las distancias a las soluciones ideales positiva y negativa permitieron establecer un ranking claro de las alternativas analizadas. El índice de cercanía (C_i) evidenció una alternativa dominante con un valor de 0.77, seguida de alternativas con valores considerablemente menores (0.37 y 0.17). Esta dispersión confirma que, al integrar los pesos derivados del AHP, el desempeño global de las alternativas no se distribuye de forma homogénea, sino que refleja diferencias sustantivas en su capacidad para responder de manera estable y útil bajo los criterios definidos.

La comparación de estos resultados con estudios previos revela tanto convergencias como aportes novedosos. Investigaciones centradas en la evaluación de modelos de lenguaje en entornos profesionales han señalado que la utilidad percibida de la IA depende fuertemente del contexto de uso y de la calidad de las instrucciones proporcionadas mediante prompt engineering [3, 1]. Sin embargo, dichos estudios suelen asumir implícitamente la estabilidad temporal del modelo. En contraste, los

resultados del presente trabajo evidencian que la degradación por envejecimiento funcional puede alterar de manera significativa la utilidad percibida, incluso cuando se aplican estrategias avanzadas de ingeniería de instrucciones.

Asimismo, mientras que la mayoría de los enfoques evaluativos tradicionales se apoyan en métricas estáticas de precisión o rendimiento promedio, los hallazgos obtenidos respaldan la necesidad de marcos evaluativos dinámicos y normalizados, capaces de resistir perturbaciones del entorno operativo. En este sentido, la integración AHP–TOPSIS se mostró coherente con los postulados de la teoría de decisión multicriterio, al permitir una comparación justa entre arquitecturas heterogéneas sin reducir el análisis a una única métrica agregada [15, 1].

Desde el punto de vista del impacto teórico, los resultados refuerzan la idea de que la evaluación de la IA debe desplazarse desde una lógica puramente algorítmica hacia una perspectiva sistémica y humana. La primacía del criterio AAS sugiere que los modelos de IA no deben ser concebidos únicamente como artefactos técnicos, sino como sistemas socio-técnicos sujetos a procesos de desgaste cognitivo y semántico. En el plano práctico, estos hallazgos tienen implicaciones directas para organizaciones que dependen de agentes inteligentes en tareas críticas, ya que evidencian la necesidad de auditorías periódicas, recalibración de modelos y mecanismos de control del envejecimiento funcional. En términos de políticas y gobernanza tecnológica, los resultados apoyan el desarrollo de marcos regulatorios que contemplen no solo la transparencia y la explicabilidad, sino también la estabilidad longitudinal de los sistemas de IA. La incorporación de criterios como la indeterminación y la normalización evaluativa puede contribuir a políticas más realistas y adaptativas, alineadas con la complejidad de los entornos digitales actuales.

No obstante, el estudio presentó ciertas limitaciones que deben ser consideradas al interpretar los resultados. En primer lugar, el número de expertos participantes y el uso de muestreo no probabilístico por juicio pudieron introducir sesgos derivados de la experiencia y el contexto profesional de los informantes. En segundo lugar, la evaluación se basó en criterios definidos a priori, lo que pudo excluir variables emergentes relevantes, como aspectos éticos específicos o impactos organizacionales a largo plazo. Finalmente, el análisis se apoyó en herramientas computacionales externas, cuya implementación, aunque validada, pudo influir en la granularidad de los resultados.

A partir de estos hallazgos, se recomienda que futuras investigaciones amplíen el marco evaluativo incorporando enfoques longitudinales y experimentales que permitan observar la degradación funcional de los agentes de IA en escenarios reales de uso prolongado. Asimismo, se sugiere integrar formalmente modelos neutrosóficos en la fase de toma de decisiones multicriterio, con el fin de capturar de manera más precisa la incertidumbre inherente a los sistemas inteligentes. Finalmente, resulta pertinente explorar la combinación de AHP–TOPSIS con métricas centradas en el usuario humano, lo que permitiría avanzar hacia evaluaciones más equitativas, robustas y alineadas con las demandas del siglo XXI.

Tablas de resultados

Tabla 1. Criterios de evaluación considerados en el estudio

Índice	Criterio de evaluación	Descripción operativa
C1	Degradación por Envejecimiento (AAS)	Pérdida progresiva de precisión y coherencia del agente tras interacciones repetidas
C2	Gestión de la Indeterminación	Capacidad del sistema para manejar incertidumbre, contradicción y ambigüedad
C3	Calidad de Datos (Data-Centric)	Nivel de curaduría, consistencia y representatividad de los datos
C4	Eficiencia Operativa (Ingeniería)	Impacto del agente en productividad y calidad técnica

Índice	Criterio de evaluación	Descripción operativa
C5	Estabilidad y Normalización	Robustez del desempeño ante perturbaciones del entorno

Tabla 2. Clasificación de los criterios de evaluación como costo o beneficio

Criterio	Tipo
C1 Degradación por Envejecimiento (AAS)	Costo
C2 Gestión de la Indeterminación	Beneficio
C3 Calidad de Datos (Data-Centric)	Beneficio
C4 Eficiencia Operativa	Beneficio
C5 Estabilidad y Normalización	Beneficio

Tabla 3. Pesos relativos de los criterios de evaluación obtenidos mediante AHP (%)

Criterio	Peso (%)	Ranking
C1 Degradación por Envejecimiento (AAS)	64.4	1
C2 Gestión de la Indeterminación	16.4	2
C3 Calidad de Datos (Data-Centric)	9.5	3
C4 Eficiencia Operativa	6.4	4
C5 Estabilidad y Normalización	3.3	5

Tabla 4. Matriz de decisión TOPSIS (normalizada)

Alternativa	C1	C2	C3	C4	C5
A1	0.13	0.11	0.12	0.12	0.11
A2	0.10	0.12	0.11	0.11	0.11
A3	0.11	0.12	0.12	0.12	0.12

Tabla 5. Ranking final de alternativas obtenido mediante TOPSIS

Alternativa	Índice de cercanía C_i	Ranking
A2	0.77	1
A3	0.37	2
A1	0.17	3

Análisis de sensibilidad

Con el objetivo de evaluar la robustez de la decisión multicriterio y descartar que el resultado obtenido sea consecuencia de una configuración puntual de pesos, se desarrolló un análisis de sensibilidad focalizado en el criterio **Eficiencia Operativa**, el cual fue identificado por los expertos como uno de los factores estratégicos del estudio. En el escenario base, dicho criterio presentó un peso relativo de 18.5% dentro del esquema decisional ampliado utilizado en la validación cruzada, situándose como el de mayor influencia operativa.

El análisis consistió en reducir de forma controlada el peso de **Eficiencia Operativa** hasta un valor del 10%, redistribuyendo proporcionalmente la diferencia entre los demás criterios, manteniendo la

coherencia del vector de pesos y la consistencia del modelo AHP. Posteriormente, se recalculó el índice de cercanía relativa (C_i) mediante TOPSIS para cada alternativa.

Los resultados mostraron que, aun bajo este escenario restrictivo, la alternativa **A2** conservó la mayor cercanía al vector ideal positivo, manteniéndose en la primera posición del ranking. Si bien se observó una ligera disminución en su valor C_i , esta no fue suficiente para alterar el orden relativo frente a A1 y A3. Este comportamiento evidencia que la superioridad de A2 no depende exclusivamente del peso asignado a la eficiencia operativa, sino de un desempeño equilibrado y consistente a través de múltiples criterios.

Desde una perspectiva metodológica, este hallazgo confirma la **robustez estructural del modelo AHP-TOPSIS** aplicado y refuerza la validez interna de la decisión. En términos prácticos, demuestra que la selección de la alternativa A2 responde a un patrón sistémico y no a una ponderación circunstancial, lo cual resulta especialmente relevante en contextos de evaluación de sistemas de inteligencia artificial, donde las condiciones operativas y las prioridades estratégicas pueden variar con el tiempo.

En consecuencia, el análisis de sensibilidad aporta un argumento adicional de rigor científico al estudio, fortaleciendo la confianza en los resultados y respaldando la aplicabilidad del marco evaluativo propuesto en entornos reales y dinámicos.

A continuación se presenta un **segmento de Discusión** integrado y coherente con la información generada en el estudio, redactado con **enfoque crítico, analítico y estilo Q1**, y alineado con los resultados obtenidos mediante **AHP-TOPSIS** y el análisis de sensibilidad.

7. Discusión

Los resultados obtenidos en esta investigación confirman que la utilidad real de los agentes de inteligencia artificial en aplicaciones contemporáneas no puede comprenderse adecuadamente desde métricas aisladas de rendimiento técnico, sino que requiere un enfoque sistémico y multicriterio. La priorización otorgada al criterio de **degradación por envejecimiento funcional de los agentes de IA**, que concentró el mayor peso relativo en el modelo AHP, refuerza los planteamientos teóricos que describen a estos sistemas como entidades adaptativas sujetas a dinámicas de desgaste cognitivo y pérdida progresiva de precisión tras interacciones reiteradas. Este hallazgo es consistente con estudios previos que advierten que la calidad de las respuestas de los modelos no depende únicamente de su arquitectura inicial, sino de la estabilidad de sus procesos de actualización, memoria contextual y gobernanza del conocimiento acumulado.

Desde esta perspectiva, el estudio aporta evidencia empírica que amplía el modelo AAS, al demostrar que el envejecimiento de los agentes constituye un factor decisivo en la evaluación de la utilidad, incluso por encima de criterios tradicionalmente dominantes como la eficiencia operativa o la capacidad de generación de respuestas. Esta constatación introduce un matiz crítico en la literatura, que suele privilegiar indicadores de desempeño inmediato sin considerar los efectos acumulativos del uso prolongado de los sistemas de IA en entornos reales.

Asimismo, la relevancia atribuida a la **gestión de la indeterminación** mediante enfoques neutrosóficos confirma que los marcos binarios de evaluación resultan insuficientes para capturar la complejidad de los entornos operativos contemporáneos. La literatura reciente ha señalado que los sistemas de IA operan frecuentemente bajo condiciones de información incompleta, ambigua o incluso contradictoria; sin embargo, pocos estudios han incorporado esta dimensión de manera explícita en los procesos de evaluación. En este sentido, los resultados del presente trabajo coinciden con aquellas investigaciones que proponen integrar la incertidumbre como una variable estructural y no como un error residual, contribuyendo a una comprensión más realista del desempeño de la IA.

El posicionamiento intermedio del criterio de **calidad de datos bajo el paradigma Data-Centric AI** respalda la idea de que la utilidad de los agentes inteligentes depende menos de la sofisticación algorítmica per se y más de la consistencia, curaduría y gobernanza de los datos que los alimentan. No obstante, los resultados también sugieren que la mejora en la calidad de los datos, si bien necesaria, no es suficiente para compensar los efectos del envejecimiento funcional, lo que introduce una diferencia relevante respecto a estudios que presentan la IA centrada en los datos como una solución casi universal a los problemas de desempeño.

En cuanto a la aplicación del método TOPSIS, el ranking final de alternativas mostró una estructura clara y estable, en la que la alternativa A2 se posicionó consistentemente como la opción más cercana al ideal. El análisis de sensibilidad reforzó este resultado al demostrar que, incluso ante una reducción significativa del peso del criterio más influyente, la alternativa óptima no se modificó. Este hallazgo es particularmente relevante desde el punto de vista metodológico, ya que evidencia la **robustez estructural de la decisión** y descarta que el resultado sea producto de una configuración arbitraria de pesos. En comparación con estudios previos que omiten este tipo de validación, el presente trabajo ofrece un mayor nivel de rigor y confiabilidad en la toma de decisiones multicriterio.

Finalmente, los resultados tienen implicaciones directas para la práctica profesional y la formulación de políticas en contextos de transformación digital, especialmente en el sector público. La evidencia sugiere que las decisiones de adopción de IA deberían priorizar criterios asociados a sostenibilidad funcional, estabilidad y capacidad de adaptación, más allá de indicadores de eficiencia inmediata. En conjunto, esta discusión posiciona al marco AHP–TOPSIS propuesto como una contribución metodológica sólida, capaz de articular teoría, evidencia empírica y toma de decisiones responsables en el uso contemporáneo de la inteligencia artificial.

8. Conclusión

El presente estudio tuvo como propósito central **analizar el impacto de los agentes de inteligencia artificial en las aplicaciones del siglo XXI y su influencia en el desempeño de su utilidad**, a partir de un enfoque multicriterio robusto que integró los métodos AHP y TOPSIS. En coherencia con este objetivo, los resultados evidenciaron que no todos los factores que suelen destacarse en el discurso tecnológico contemporáneo tienen el mismo peso real en la evaluación de la utilidad de la IA. En particular, el análisis AHP mostró que la **degradación por envejecimiento funcional de los agentes de IA**, entendida desde el modelo AAS, concentró el mayor peso relativo, lo que confirma que la pérdida progresiva de precisión y coherencia tras interacciones repetidas constituye una preocupación estructural. En segundo lugar, la **gestión de la indeterminación**, asociada a enfoques neutrosóficos, y la **calidad de los datos bajo el paradigma Data-Centric AI** emergieron como factores críticos, aunque subordinados al problema del envejecimiento. La aplicación de TOPSIS permitió, además, establecer un ranking claro y estable de las alternativas evaluadas, cuya robustez fue confirmada mediante el análisis de sensibilidad.

Desde una perspectiva teórica, estos hallazgos refuerzan la necesidad de desplazar el foco desde una visión puramente algorítmica de la inteligencia artificial hacia un enfoque sistémico, donde la utilidad se comprenda como un constructo dinámico, sensible al contexto, a la calidad de los datos y a la interacción prolongada con los usuarios. El estudio contribuye así a la literatura sobre evaluación de sistemas inteligentes al integrar la teoría de la decisión multicriterio con enfoques contemporáneos como la IA centrada en los datos y la lógica neutrosófica, ampliando los marcos tradicionales que suelen basarse en métricas aisladas o en evaluaciones estáticas del desempeño.

En el ámbito de la gestión organizacional y de las políticas públicas, los resultados tienen implicaciones relevantes. En contextos de transformación digital, especialmente en instituciones públicas, la adopción de agentes de IA no debería fundamentarse únicamente en indicadores de eficiencia inmediata o reducción de costos. Por el contrario, la evidencia sugiere que es imprescindible considerar la sostenibilidad funcional de los sistemas, la gobernanza de los datos y la capacidad

institucional para gestionar la incertidumbre inherente a los entornos reales. Desde esta óptica, el marco AHP–TOPSIS propuesto ofrece una herramienta transparente y replicable para apoyar procesos de toma de decisiones más informados, reduciendo el riesgo de inversiones tecnológicas guiadas por modas o expectativas sobredimensionadas.

En términos de implicaciones prácticas, el estudio proporciona orientaciones claras para profesionales de la ingeniería, el desarrollo de software y la gestión tecnológica. La priorización de criterios como el envejecimiento funcional y la calidad de los datos invita a replantear estrategias de mantenimiento, curaduría de información y diseño de arquitecturas más resilientes. Asimismo, los resultados respaldan el uso de enfoques estructurados de evaluación en proyectos de IA, particularmente cuando se trata de comparar alternativas tecnológicas con impactos organizacionales y sociales significativos.

No obstante, el estudio presenta limitaciones que deben ser reconocidas. La evaluación se basó en el juicio experto de una muestra acotada, lo que, si bien es consistente con la naturaleza del método AHP, puede introducir sesgos contextuales. Además, el análisis se centró en un conjunto específico de criterios y alternativas, por lo que los resultados no deben interpretarse como universalmente generalizables. Futuras investigaciones podrían ampliar el número de criterios, incorporar métodos híbridos con enfoques estadísticos o explorar aplicaciones empíricas en distintos sectores y escalas institucionales.

En conjunto, esta investigación ofrece una contribución sólida y prospectiva al debate sobre la utilidad real de la inteligencia artificial en el siglo XXI. Al proponer un marco evaluativo estable, humano y metodológicamente riguroso, el estudio sienta las bases para decisiones tecnológicas más responsables, alineadas con las demandas de sostenibilidad, transparencia y valor público que caracterizan la transformación digital contemporánea.

9. Referencias

- [1]. Radford, K Narasimhan, T Salimans, & I Sutskever. Improving Language Understanding by Generative Pre-Training. OpenAI, 2019
- [2]. S. Amershi, A Begel, C Bird, R DeLine, H Gall, E Kamar, & T Zimmermann. Software engineering for machine learning: A case study. *EEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 2019
- [3]. M. X Chen, O Firat, A Bapna, M Johnson, W Macherey, G Foster & Y Wu. OpenAI Codex: Human-like language understanding and code generation. *rXiv preprint arXiv:2109.03374*, 2021
- [4]. T. B Hashimoto, K Guu, Y Oren, & P Liang. A retrieve-and-edit framework for predicting structured outputs. *Advances in Neural Information Processing Systems*, 31, 10076-10086, 2018
- [5]. M. Allamanis, E. T Barr, P Devanbu, & C Sutton. A survey of machine learning for big code and naturalness. *ACM Computing Surveys (CSUR)*, 51(4), 1-37, 2018
- [6]. V. Raychev, P Bielik, & M Vechev. Probabilistic model for code with decision trees. *ACM SIGPLAN Notices*, 51(10), 731-747, 2016
- [7]. O. Mar, I. Santana, and J. Gulín, "Algoritmo para determinar y eliminar nodos neutros en Mapa Cognitivo Neu-trosófico," *Neutrosophic Computing and Machine Learning*, vol. 8, pp. 4-11, 2019.
- [8]. R. G. Ortega, M. Rodríguez, M. L. Vázquez, and J. E. Ricardo, "Pestel analysis based on neutrosophic cognitive maps and neutrosophic numbers for the sinos river basin management," *Neutrosophic Sets and Systems*, vol. 26, no. 1, pp. 16, 2019.
- [9]. Edalatpanah, S. A., & Smarandache, F. (2019). Data envelopment analysis for simplified neutrosophic sets. *Infinite Study*.
- [10]. MY. Leyva Vázquez, & F. Smarandache. *Neutrosofía: Nuevos avances en el tratamiento de la incertidumbre*. Infinite Study. 2018
- [11]. F. Smarandache. "Introduction to Neutrosophic Statistics". Ed. Infinite Study, 150, 2014

- [12]. T. B Brown, B Mann, N Ryder, M Subbiah, J Kaplan, P Dhariwal, & D Amodei. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 2020
- [13]. OpenAI. OpenAI. Obtenido de OpenAI: <https://openai.com/blog/gpt-3-apps/>, 2021
- [14]. Microsoft. GitHub Copilot. Obtenido de GitHub Copilot: <https://copilot.github.com/>, 2021
- [15]. Boehm & R Turner. Management challenges to implementing agile processes in traditional development organiza-tions. *IEEE software*, 22(5), 30-39, 2004
- [16]. P Lientz, & E. B. Swanson. *Software maintenance management: A study of the maintenance of computer applica-tion software in 487 data processing organizations*. Addison-Wesley Longman Publishing Co., Inc, 1980
- [17]. Von Feigenblatt, O. F. *Trends and Debates in American Education: A Hispanic Perspective*. Ediciones Octaedro, 2023.

