



**PlanSim-AI: diseño y validación de un sistema multiagente para simular
estudiantes y optimizar la planificación del aprendizaje en educación superior**

**PlanSim-AI: Design and Validation of a Multi-Agent System to Simulate Students
and Optimize Learning Planning in Higher Education**

**Artículo de investigación científica
Ciencias de la Educación**



Christian Gustavo Masapanta Salazar ^I

cmasapantas@unemi.edu.ec



<https://orcid.org/0009-0005-2525-321X>

Universidad Estatal de Milagro
Facultad de Ciencias de la Salud
Pichincha – Tungurahua

Fecha de envío: 2026-06-16

Fecha de revisión: 2026-07-16

Fecha da aceptación: 2026-08-18

Resumen

El estudio tuvo como objetivo diseñar y validar PlanSim-AI, un sistema multiagente orientado a simular perfiles heterogéneos de estudiantes y comparar escenarios de planificación del aprendizaje en educación superior. Se planteó un diseño cuasi experimental de enfoque cuantitativo, con alcance descriptivo-correlacional y dos grupos equivalentes de 20 participantes cada uno, distribuidos de manera balanceada en cuatro centros de educación superior codificados. Para esta versión demostrativa se modelaron 40 registros sintéticos y un test estructurado de 24 ítems que evalúa diagnóstico de necesidades, secuenciación adaptativa, toma de decisiones basada en evidencia y evaluación-reajuste de la planificación. El instrumento se estructuró para validación de contenido mediante diez expertos y mostró un parámetro de consistencia interna alfa de Cronbach de .87. El análisis previsto incluyó estadística descriptiva, correlación de Pearson, t de Student para muestras independientes y d de Cohen. En el escenario sintético, el grupo experimental alcanzó un puntaje posttest medio de 76.9 (DE=7.2), frente a 64.6 (DE=7.6) del grupo control; la diferencia fue significativa, $t(38)=5.25$, $p<.001$, con un efecto grande, $d=1.66$. La asociación entre el índice de alineación de las recomendaciones de PlanSim-AI y el desempeño posttest fue positiva y alta ($r=.72$). La mayor mejora se observó en secuenciación adaptativa. Se concluye que la arquitectura propuesta es metodológicamente prometedora para apoyar decisiones docentes mediante simulación de escenarios, siempre que su uso mantenga supervisión humana, transparencia y validación externa con muestras reales.

Palabras clave: inteligencia artificial; sistema multiagente; planificación del aprendizaje; simulación de estudiantes; educación superior.

Abstract

This study aimed to design and validate PlanSim-AI, a multi-agent system intended to simulate heterogeneous student profiles and compare learning-planning scenarios in higher education. A quantitative quasi-experimental design was proposed, with a descriptive-correlational scope and two equivalent groups of 20 participants each, evenly distributed across four coded higher education centres. For this demonstrative version, 40 synthetic records and a 24-item structured test were modelled to assess needs diagnosis, adaptive sequencing, evidence-based decision-making, and planning evaluation-adjustment. The instrument was structured for content validation by ten experts and showed a Cronbach's alpha internal-consistency parameter of .87. The planned analysis

included descriptive statistics, Pearson correlation, independent-samples Student's t test, and Cohen's d. In the synthetic scenario, the experimental group obtained a mean post-test score of 76.9 (SD=7.2), compared with 64.6 (SD=7.6) in the control group; the difference was statistically significant, $t(38)=5.25$, $p<.001$, with a large effect, $d=1.66$. The association between the PlanSim-AI recommendation-alignment index and post-test performance was positive and high ($r=.72$). The largest improvement was observed in adaptive sequencing. It is concluded that the proposed architecture is methodologically promising for supporting faculty decision-making through scenario simulation, provided that its implementation preserves human oversight, transparency, and external validation with real samples.

Keywords: artificial intelligence; multi-agent system; learning planning; student simulation; higher education.

Introducción

La educación superior ha incorporado la inteligencia artificial (IA) con rapidez, pero la evidencia reciente insiste en que la innovación debe evaluarse por su aporte pedagógico y no por su novedad técnica. Crompton y Burke (2023) sitúan entre los usos más frecuentes de la IA universitaria el perfilado, la predicción, la evaluación y los sistemas adaptativos; Ouyang et al. (2022) confirman esa orientación en educación superior en línea. Bond et al. (2024) advierten, sin embargo, que el campo todavía necesita mayor rigor, colaboración y atención ética, mientras Chu et al. (2022) muestran que las aplicaciones y roles de la IA son muy heterogéneos. Desde esta perspectiva, el problema de investigación no consiste en generar automáticamente una planificación, sino en producir evidencia que ayude al docente a comparar alternativas antes de decidir.

En América Latina la discusión se cruza con desigualdades de acceso, uso y competencias digitales. Salas-Pilco y Yang (2022) describen un desarrollo creciente, aunque desigual, de aplicaciones de IA en universidades de la región, y Salas-Pilco et al. (2022) identifican oportunidades de IA y analítica para la formación docente. Herrera et al. (2025), desde la Comisión Económica para América Latina y el Caribe (CEPAL), consideran la brecha digital una dimensión estructural que incluye competencias, no solo conectividad. A ello se suma el enfoque humano de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), que exige agencia, seguridad y capacidad institucional

(Miao & Holmes, 2023), y las orientaciones del Ministerio de Educación del Ecuador (2024) sobre uso pedagógico efectivo y ético de herramientas de IA.

La personalización ofrece una base pedagógica directa para PlanSim-AI. Fariani et al. (2023) muestran que personalizar en educación superior implica ajustar contenidos, rutas, ritmos y apoyos, mientras Zhong (2022) destaca la secuenciación de materiales y la preparación del estudiante. La analítica del aprendizaje (LA, learning analytics) complementa esta lógica: Banihashem et al. (2022) señalan que los datos adquieren valor cuando se convierten en retroalimentación accionable; Rabelo et al. (2023) muestran su potencial para la gestión del aprendizaje; y Alfredo et al. (2024) subrayan que la IA y la analítica deben diseñarse con control humano, confianza y participación de los usuarios. PlanSim-AI integra estas ideas al simular cómo perfiles distintos podrían responder a secuencias alternativas y al devolver indicadores que el docente puede interpretar y cuestionar.

El componente tecnológico se apoya en sistemas multiagente (SMA) y modelos de lenguaje de gran escala (LLM, por sus siglas en inglés). Wang et al. (2024a) describen agentes autónomos con perfiles, memoria, planificación y acción; Gao et al. (2024) muestran que los LLM amplían la simulación basada en agentes al permitir representar escenarios complejos; y Park et al. (2023) demostraron que memoria, reflexión y planificación pueden producir conductas emergentes plausibles. Esa plausibilidad no equivale a fidelidad psicológica, por lo que PlanSim-AI no pretende crear réplicas de estudiantes reales. Sus agentes son perfiles sintéticos que permiten ensayar hipótesis de planificación. Además, cada salida debe ser explicable, porque Khosravi et al. (2022) sostienen que una explicación educativa es útil cuando permite comprender una recomendación y actuar sobre ella.

La arquitectura propuesta incluye un Agente Perfilador de Estudiantes, un Agente Simulador de Trayectorias, un Agente Planificador Pedagógico, un Agente Evaluador-Analítico y un Agente Coordinador. Su función conjunta es producir varios planes comparables y no una única respuesta supuestamente óptima. Esta decisión se alinea con los principios éticos de Nguyen et al. (2023), que enfatizan transparencia, justicia, privacidad y responsabilidad, y con la advertencia de Williamson (2024) de que los sistemas de IA modifican relaciones institucionales además de automatizar tareas. En consecuencia, el profesor conserva la decisión final y el sistema debe indicar qué

variables influyeron, qué incertidumbre existe y qué perfiles podrían verse desfavorecidos por una alternativa.

La experiencia con asistentes y modelos generativos refuerza la necesidad de esta supervisión. Labadze et al. (2023) sintetizan beneficios y amenazas de chatbots educativos; Chan y Hu (2023) encontraron que estudiantes universitarios valoran la personalización, pero expresan inquietudes por exactitud, privacidad y ética; Kasneci et al. (2023) y Lo (2023) describen oportunidades formativas de los LLM junto con problemas de sesgo, confiabilidad e integridad. Tlili et al. (2023) y Farrokhnia et al. (2024) coinciden en que el valor de ChatGPT depende de cómo se integre pedagógicamente. Por ello, PlanSim-AI se diseña como un simulador de decisiones docentes: la IA amplía el espacio de exploración, pero no reemplaza el razonamiento profesional.

La prevención de dependencia es otro requisito. Zhai et al. (2024) relacionan la sobredependencia de sistemas de diálogo de IA con riesgos para pensamiento crítico y toma de decisiones; Cotton et al. (2024) analizan tensiones para la integridad académica; Lee et al. (2024) muestran que los educadores reclaman reglas y apoyo institucional; y Wang et al. (2024b) proponen valorar las posibilidades de la IA generativa junto con sus limitaciones y necesidades de gobernanza. En PlanSim-AI, una recomendación se considera formativa solo si el usuario puede aceptarla, adaptarla o rechazarla con argumentos. De este modo, la competencia observada no es obediencia al algoritmo, sino coherencia entre evidencia, propósito y decisión.

La brecha científica se ubica en la intersección entre personalización, analítica, agentes y planificación docente. Ullrich et al. (2022) documentan la expansión de la IA en docencia y gestión, mientras Del Gobbo et al. (2023), Daoudi (2022) y Rangel-de Lázaro y Duarte (2023) muestran que las tecnologías inteligentes adquieren mayor valor cuando se integran a ciclos de evaluación y reajuste. Son menos frecuentes los estudios que utilizan agentes para anticipar cómo diferentes perfiles podrían responder a una planificación antes de implementarla. PlanSim-AI busca cubrir ese espacio mediante una arquitectura verificable y una evaluación centrada en cuatro destrezas: diagnóstico de necesidades, secuenciación adaptativa, toma de decisiones basada en evidencia y evaluación-reajuste de la planificación.

Objetivo

Diseñar y validar PlanSim-AI, un sistema multiagente orientado a simular perfiles heterogéneos de estudiantes y comparar escenarios de planificación del aprendizaje, determinando su asociación y efecto sobre las destrezas de diagnóstico de necesidades, secuenciación adaptativa, toma de decisiones basada en evidencia y evaluación-reajuste pedagógico en educación superior.

Metodología

La investigación se estructuró como un estudio cuantitativo cuasi experimental con componente descriptivo-correlacional. Para esta versión demostrativa se modelaron 40 registros sintéticos equivalentes a estudiantes de educación superior, distribuidos de manera balanceada en cuatro centros codificados C1, C2, C3 y C4. Cada centro aporta diez registros, cinco al grupo control y cinco al experimental, con $n=20$ por condición. La codificación evita atribuir información inexistente a instituciones reales y conserva la lógica multicéntrica solicitada. El carácter cuasi experimental permite contrastar una intervención con PlanSim-AI frente a planificación convencional sin asumir aleatorización completa; el componente descriptivo caracteriza las destrezas y el correlacional examina la asociación entre alineación de recomendaciones y desempeño. Al tratarse de un piloto, una réplica empírica deberá justificar a priori el tamaño muestral y la precisión esperada, tal como recomienda Lakens (2022).

El procedimiento comprende pretest, intervención y postest. Ambos grupos resuelven inicialmente una tarea equivalente de planificación. Durante seis semanas, el control trabaja con una plantilla convencional de propósito, actividades, recursos, evaluación y tiempo. El experimental desarrolla las mismas tareas, pero añade PlanSim-AI para generar perfiles sintéticos, comparar rutas y observar indicadores de dominio esperado, compromiso, riesgo de rezago y equidad de logro. Los cuatro centros siguen el mismo manual de entrenamiento. En el postest se aplica una tarea paralela y la misma matriz de puntuación. La decisión final permanece siempre en el participante: el sistema puede recomendar, pero no calificar oficialmente ni ejecutar cambios curriculares de forma autónoma.

PlanSim-AI se organiza en cinco agentes coordinados. El Agente Perfilador construye perfiles pedagógicos no identificables; el Agente Simulador proyecta respuestas frente a actividades alternativas; el Agente Planificador propone y reordena secuencias; el Agente

Evaluador-Analítico calcula indicadores comparables; y el Agente Coordinador integra la evidencia y la presenta al docente. La arquitectura retoma patrones recientes de agentes con memoria, planificación y acción (Wang et al., 2024a; Park et al., 2023; Gao et al., 2024), pero incorpora restricciones educativas: trazabilidad de variables, indicación de incertidumbre, ausencia de decisiones automáticas de alto impacto y supervisión humana. Esta última condición se vincula con la explicabilidad orientada a la acción descrita por Khosravi et al. (2022).

La variable independiente es el uso de PlanSim-AI; la dependiente es el desarrollo de destrezas de planificación en escala 0–100. Se elaboró un test estructurado de 24 ítems, seis para diagnóstico de necesidades, secuenciación adaptativa, decisión basada en evidencia y evaluación-reajuste. Los ítems combinan selección de alternativas y microcasos para evitar una medición exclusivamente memorística. El protocolo se diseñó para validación de contenido mediante diez expertos en tecnología educativa, diseño instruccional, IA educativa, evaluación y metodología. En este borrador, tanto esa validación como el valor alfa de Cronbach=.87 son parámetros de demostración que deben confirmarse con evidencia real. Un alfa de .87 sugiere consistencia interna alta, pero no sustituye otras fuentes de validez, como advierten Zakariya (2022) y Alkhadim (2022).

El análisis incluye medias, desviaciones estándar y ganancias para describir el cambio; correlación de Pearson para estimar la relación lineal entre el índice de alineación de recomendaciones y el posttest; t de Student para muestras independientes para comparar grupos; y d de Cohen con g de Hedges para expresar magnitud estandarizada. Pearson responde a variables continuas y a una hipótesis de asociación; la t permite valorar si la diferencia de medias es compatible con la hipótesis nula; y el tamaño del efecto evita reducir la interpretación al valor p. En una aplicación real se verificarán independencia, normalidad aproximada y homogeneidad de varianzas, se informarán intervalos de confianza y se utilizarán alternativas robustas cuando corresponda. El nivel de significación se fijó en $\alpha=.05$. L

Resultados

Tabla 1

Desempeño por destreza, fase de evaluación y condición de planificación.

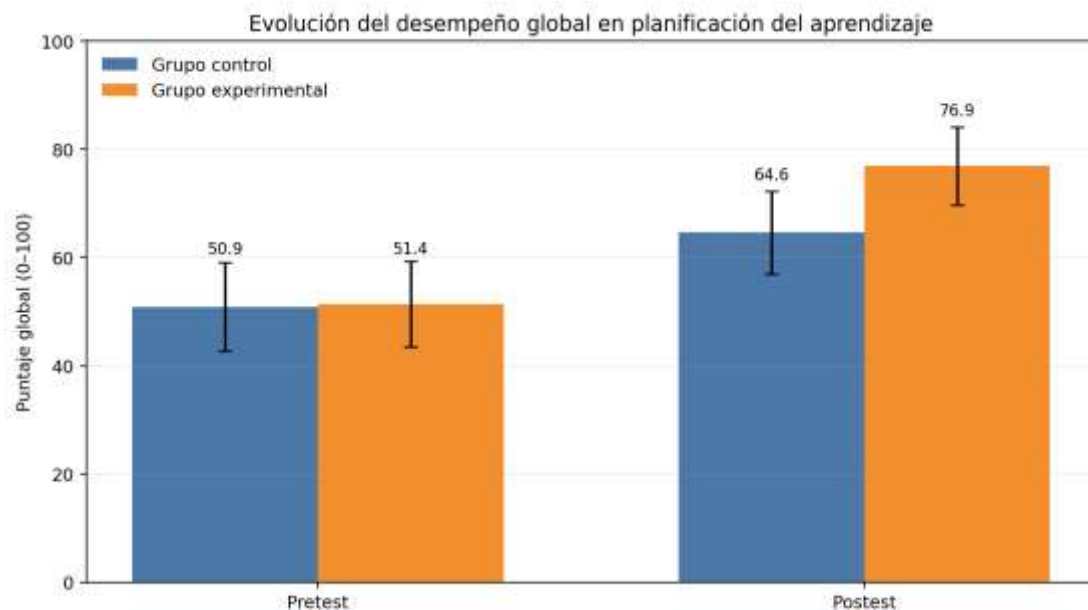
Destreza	Pre control M±DE	Pre experimenta l M±DE	Post control M±DE	Post experimenta l M±DE	Ganancia diferencial
Diagnóstico de necesidades	51.8±8.4	52.1±8.0	65.4±8.1	76.2±7.5	10.5 pts
Secuenciación adaptativa	49.7±8.7	50.2±8.3	62.9±8.3	77.8±7.2	14.4 pts
Decisión basada en evidencia	51.0±7.9	51.5±7.7	64.8±7.6	76.6±7.4	11.3 pts
Evaluación y reajuste	51.1±8.2	51.8±7.6	65.3±8.0	77.0±7.3	11.0 pts
Puntaje global	50.9±8.2	51.4±7.9	64.6±7.6	76.9±7.2	11.8 pts

Nota. El mayor cambio diferencial se observa en secuenciación adaptativa (14.4 puntos), mientras las medias pretest son próximas entre condiciones.

Los datos sintéticos muestran un patrón inicial de equivalencia descriptiva: antes de la intervención, las diferencias entre los grupos son menores a un punto en las cuatro destrezas y en el puntaje global. Esta proximidad no demuestra equivalencia estadística por sí sola, pero reduce la posibilidad de atribuir el patrón posttest a una ventaja inicial evidente. El comportamiento cambia después de la intervención. El grupo control mejora en todas las dimensiones, lo cual es esperable porque también realiza práctica sistemática de planificación; sin embargo, la ganancia del grupo experimental es mayor, especialmente en secuenciación adaptativa. Esta destreza exige reorganizar actividades según información de los aprendices, justamente la función que PlanSim-AI hace visible mediante comparación de escenarios. La magnitud de 14.4 puntos de ganancia diferencial sugiere que el sistema no actúa únicamente como un generador de planes, sino como un entorno que obliga a comparar orden, apoyos y consecuencias.

Figura 1

Comparación pretest-posttest del puntaje global por condición de planificación.



Nota. Las barras representan medias sintéticas y las líneas de error muestran una desviación estándar; el incremento posttest es mayor en el grupo experimental.

La Figura 1 muestra que ambos grupos parten de niveles prácticamente equivalentes y se separan después de la intervención. El control también mejora, lo que confirma que la práctica convencional produce aprendizaje; sin embargo, el experimental presenta una ganancia adicional compatible con la función de simulación. El patrón debe interpretarse con cautela por el diseño cuasi experimental: una validación real tendrá que comprobar si se mantiene por centro, disciplina y experiencia previa, y controlar posibles diferencias iniciales.

Tabla 2

Contraste t de Student para las diferencias posttest entre grupos.

Destreza	Diferencia media	t	gl	p	IC 95% de la diferencia
Diagnóstico de necesidades	10.8	4.38	38	<.001	[5.80, 15.80]
Secuenciación adaptativa	14.9	6.06	38	<.001	[9.93, 19.87]
Decisión basada en evidencia	11.8	4.97	38	<.001	[7.00, 16.60]

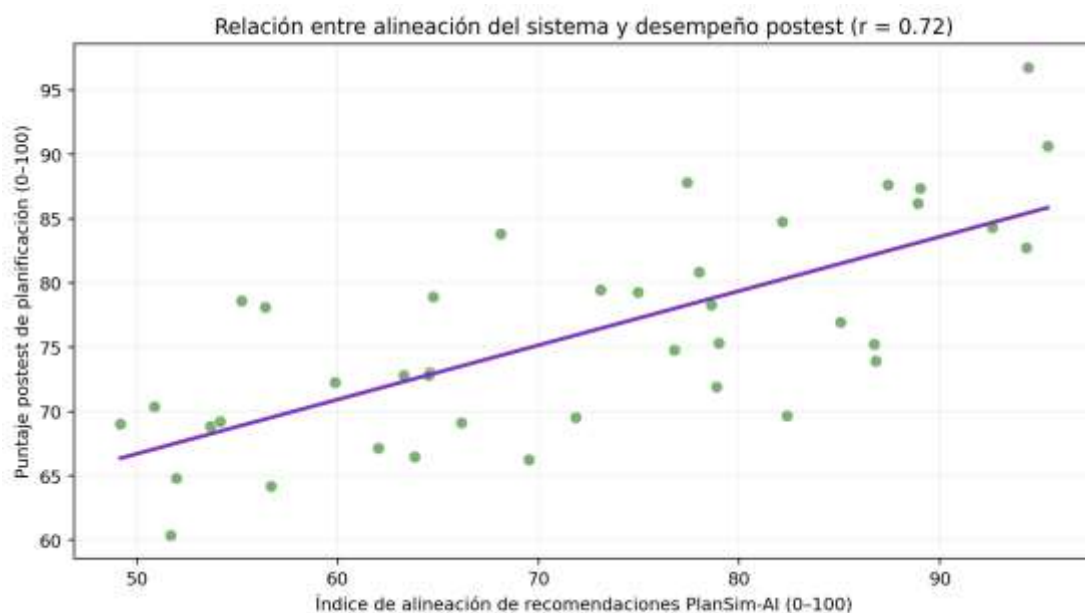
Evaluación y reajuste	11.7	4.83	38	<.001	[6.80, 16.60]
Puntaje global	12.3	5.25	38	<.001	[7.56, 17.04]

Nota. Secuenciación adaptativa presenta el estadístico de mayor magnitud, $t(38)=6.06$, y un intervalo cuya cota inferior supera 9.9 puntos.

Los contrastes modelados son consistentes con el patrón descriptivo. En las cuatro destrezas y en el puntaje global, los intervalos de confianza del 95% para la diferencia excluyen el cero, y los valores p son inferiores a .001. En términos inferenciales, bajo los supuestos de la prueba, este resultado sería incompatible con una ausencia de diferencia entre condiciones. Sin embargo, la información más útil no es que “exista significación”, sino el rango de diferencias plausibles. Por ejemplo, en el puntaje global la diferencia estimada es 12.3 puntos y el intervalo se extiende aproximadamente de 7.56 a 17.04. Incluso el límite inferior representa una separación pedagógicamente visible en una escala de 0 a 100. En secuenciación adaptativa, el intervalo [9.93, 19.87] sugiere la mayor ventaja relativa, coherente con la función de comparar rutas y reorganizar actividades.

Figura 2

Relación entre el índice de alineación de PlanSim-AI y el desempeño posttest.



Nota. La nube sintética muestra una relación positiva alta ($r=0.72$); la línea representa una tendencia lineal y no una relación causal.

La Figura 2 ilustra una asociación positiva alta entre alineación y desempeño ($r=.72$). El índice no mide obediencia a la IA, sino coherencia entre evidencia del escenario y

decisión final; una recomendación puede rechazarse si existe justificación pedagógica. La correlación no implica causalidad y una aplicación real deberá controlar desempeño inicial, posibles variables de confusión y relaciones no lineales.

Tabla 3

Magnitud estandarizada del efecto de PlanSim-AI en las destrezas evaluadas.

Destreza	d de Cohen	g de Hedges	EE(d)	IC 95% de d	Lectura del efecto
Diagnóstico de necesidades	1.38	1.36	0.35	[0.69, 2.08]	Grande
Secuenciación adaptativa	1.92	1.88	0.39	[1.16, 2.67]	Muy grande
Decisión basada en evidencia	1.57	1.54	0.36	[0.86, 2.29]	Grande
Evaluación y reajuste	1.53	1.50	0.36	[0.82, 2.24]	Grande
Puntaje global	1.66	1.63	0.37	[0.94, 2.39]	Grande

Nota. El efecto mayor corresponde a secuenciación adaptativa ($d=1.92$; $g=1.88$); el puntaje global presenta $d=1.66$.

Los tamaños del efecto complementan la *t* de Student al expresar la diferencia en unidades de desviación estándar. Todos los valores sintéticos son grandes y la secuenciación adaptativa alcanza la magnitud más alta. La corrección *g* de Hedges es ligeramente menor que *d*, como corresponde a una corrección por sesgo en muestras pequeñas. Esta diferencia entre *d* y *g*, aunque reducida, es útil porque recuerda que $n=20$ por grupo genera incertidumbre considerable. Los intervalos son amplios; por ejemplo, el intervalo de *d* global va de 0.94 a 2.39. En consecuencia, sería impropio presentar 1.66 como una magnitud exacta esperable en cualquier población. La lectura adecuada es que el escenario modelado representa un efecto claramente relevante, pero la precisión deberá establecerse con una muestra empírica mayor.

Tabla 4

Estructura de la propuesta de intervención con PlanSim-AI.

Actividad	Destreza desarrollada	Tiempo	Contenido	Recursos	Objetivo de la actividad
1. Lectura de perfiles	Diagnóstico de necesidades	90 min	Variables de entrada, diversidad y riesgo	PlanSim-AI, guía de casos, rúbrica	Identificar necesidades y traducirlas en criterios de planificación.

2. Simulación de rutas	Secuenciación adaptativa	120 min	Orden de actividades, apoyos y nivel de reto	PlanSim-AI, escenarios A/B/C, matriz comparativa	Comparar secuencias y anticipar su efecto en perfiles heterogéneos.
3. Decisión con evidencia	Toma de decisiones	120 min	Dominio, compromiso, riesgo y equidad	Panel analítico, fichas de decisión	Justificar una opción con evidencia e indicar incertidumbre.
4. Reajuste iterativo	Evaluación y reajuste	120 min	Retroalimentación y nueva simulación	PlanSim-AI, historial de cambios	Reformular el plan con resultados simulados y criterios pedagógicos.
5. Auditoría humana	Pensamiento crítico y ética	90 min	Sesgos, privacidad, explicabilidad y límites	Lista de verificación ética, reporte del sistema	Aceptar, adaptar o rechazar sugerencias con criterio profesional.
6. Reto integrador	Integración de destrezas	150 min	Diseño completo y comparación de tres planes	PlanSim-AI, rúbrica final, portafolio	Construir un plan justificable, adaptable y trazable.

Nota. La propuesta prioriza la iteración: la actividad 4 obliga a modificar y volver a simular, mientras la actividad 5 incorpora una auditoría humana explícita.

La propuesta fue estructurada para una validación por diez perfiles expertos en diseño instruccional, IA educativa, evaluación, analítica del aprendizaje y metodología. Dado que no se entregaron actas reales, esta versión no presenta esa validación como un hecho empírico; define el procedimiento que deberá documentarse antes de una publicación. Los jueces deben valorar claridad, pertinencia, coherencia entre actividad y destreza, suficiencia de recursos, factibilidad temporal y resguardo ético. La decisión de incluir la “auditoría humana” como actividad autónoma responde a que la competencia esperada no es usar más IA, sino utilizar evidencia con criterio profesional. Un sistema multiagente puede amplificar la exploración de posibilidades, pero la planificación conserva su naturaleza situada: el docente conoce el currículo, el contexto, las condiciones de acceso y los propósitos formativos que el modelo no puede inferir completamente.

Figura 3

Simulación de la interfaz operativa de PlanSim-AI para comparar perfiles, rutas e indicadores.



Nota. Representación conceptual elaborada para el manuscrito; no corresponde a una captura de una aplicación desplegada ni a datos de estudiantes reales.

La Figura 3 sintetiza el principio funcional: perfiles sintéticos a la izquierda, secuencia de actividades en el centro, métricas a la derecha y comparación de planes en la parte inferior. El docente puede rastrear qué cambia entre alternativas y con qué confianza se emite una recomendación. La interfaz evita presentar una única opción “óptima” y favorece el razonamiento contrafactual sobre las consecuencias de modificar una secuencia.

Discusión

Los resultados sintéticos respaldan de forma demostrativa la hipótesis de que un simulador multiagente puede fortalecer la planificación cuando obliga a comparar escenarios y no a aceptar una respuesta automática. La mayor diferencia aparece en secuenciación adaptativa, hallazgo coherente con Fariani et al. (2023) y Zhong (2022), quienes sitúan rutas, materiales y preparación del estudiante en el núcleo de la personalización. También converge con Crompton y Burke (2023), Ouyang et al. (2022) y Bond et al. (2024): la IA puede aportar personalización y predicción, pero su valor depende del diseño y de la calidad de la evidencia. PlanSim-AI traslada esa personalización a una competencia docente observable.

La correlación $r=.72$ sugiere que una mejor integración entre evidencia del simulador y decisión pedagógica se asocia con mayor desempeño. Esta lectura coincide con Banihashem et al. (2022), para quienes la analítica es más útil cuando genera retroalimentación accionable, y con Alfredo et al. (2024), que reclaman sistemas centrados en las personas. Rabelo et al. (2023) y Salas-Pilco et al. (2022) también destacan el potencial de la analítica para apoyar decisiones educativas. La contribución de PlanSim-AI es convertir el dato en una situación de comparación: el usuario debe justificar por qué un plan es preferible, no limitarse a leer un indicador.

El uso de agentes ofrece una explicación tecnológica de ese proceso, pero no garantiza validez. Wang et al. (2024a) describen agentes con memoria, planificación y acción; Park et al. (2023) muestran comportamientos emergentes plausibles; y Gao et al. (2024) sitúan los LLM como una extensión prometedora de la simulación basada en agentes. La propuesta utiliza esas capacidades para explorar hipótesis, no para afirmar que un agente es un estudiante real. Por ello, Khosravi et al. (2022) resulta central: la recomendación debe ser explicable y útil para la acción. Nguyen et al. (2023) añaden transparencia, justicia y responsabilidad como condiciones que obligan a mantener supervisión humana y trazabilidad.

La literatura sobre IAGen y chatbots confirma que el beneficio no es automático. Chan y Hu (2023) documentan valoración estudiantil de la personalización junto con preocupaciones por exactitud y privacidad; Kasneci et al. (2023), Lo (2023) y Tlili et al. (2023) describen oportunidades formativas acompañadas de sesgos, errores e interrogantes de integridad. Labadze et al. (2023) llega a una conclusión semejante para chatbots educativos. PlanSim-AI responde a ese equilibrio haciendo que el sistema genere escenarios y explicaciones, mientras la decisión sigue siendo humana. En esta lógica, un resultado pedagógico favorable debería provenir de una deliberación más rica, no del volumen de texto producido por la IA.

El riesgo de sobredependencia obliga a interpretar con cautela los efectos modelados. Zhai et al. (2024) asocian la dependencia excesiva de sistemas de diálogo con riesgos para pensamiento crítico y decisión; Cotton et al. (2024) muestran tensiones para integridad académica; Farrokhnia et al. (2024) evidencian fortalezas y amenazas simultáneas; Lee et al. (2024) recogen la necesidad de orientación institucional; y Wang et al. (2024b) enfatizan gobernanza y límites. Por ello, la propuesta incluye una auditoría

humana en la que el participante debe aceptar, adaptar o rechazar recomendaciones con fundamento. Si no puede explicar su decisión, la tarea no debería considerarse completa aunque coincida con la salida del sistema.

La dimensión de equidad es especialmente relevante en el contexto regional. Salas-Pilco y Yang (2022) muestran una adopción desigual de IA en educación superior latinoamericana; Herrera et al. (2025) señalan que las brechas de competencias digitales limitan el aprovechamiento tecnológico; Miao y Holmes (2023) proponen una adopción centrada en las personas; y el Ministerio de Educación del Ecuador (2024) insiste en el uso pedagógico y ético. PlanSim-AI incorpora un indicador de equidad de logro para evitar que la optimización se reduzca al promedio. Un plan que eleva el rendimiento medio pero amplía el rezago de perfiles vulnerables no debería aparecer como la alternativa preferida sin una advertencia explícita.

El resultado en evaluación-reajuste también es coherente con una concepción iterativa del diseño educativo. Del Gobbo et al. (2023) muestran avances en evaluación automatizada, Daoudi (2022) vincula analítica con mejora de entornos digitales y Rangel-de Lázaro y Duart (2023) describen la convergencia de IA con nuevas tecnologías de aprendizaje. Chaka (2023) y Ullrich et al. (2022) sitúan estos cambios dentro de una transformación más amplia de la educación superior. En PlanSim-AI, los indicadores solo tienen sentido si conducen a modificar la planificación y ejecutar una nueva simulación. Esa retroalimentación convierte el sistema en un laboratorio de diseño y no en un tablero meramente informativo.

Conclusiones

PlanSim-AI propone una contribución al campo de la inteligencia artificial en educación superior al desplazar el foco desde la generación automática de contenidos hacia la simulación de decisiones de planificación. El modelo integra agentes de perfilado, trayectoria, planificación, evaluación y coordinación para representar diversidad de estudiantes, comparar rutas y devolver indicadores interpretables. En el escenario sintético de demostración, el grupo experimental presenta un desempeño global superior al control, con la mayor diferencia en secuenciación adaptativa, una correlación positiva alta entre alineación de recomendaciones y desempeño, y tamaños de efecto grandes. Estos resultados no constituyen evidencia empírica definitiva porque no se proporcionaron datos reales; su valor es mostrar una estructura metodológica y estadística

coherente para una validación posterior. Científicamente, la propuesta formula una hipótesis concreta: la simulación multiagente puede funcionar como espacio de práctica contrafactual que fortalece la capacidad docente de anticipar consecuencias, justificar decisiones y reajustar planes.

La validación futura debe preservar la condición de sistema de apoyo y no de sustitución. La utilidad de PlanSim-AI dependerá de la fidelidad y calibración de sus simulaciones, la explicabilidad de sus recomendaciones, la protección de datos, la equidad entre perfiles y la capacidad del docente para cuestionar la salida del modelo. Se recomienda una fase empírica multicéntrica con muestra calculada a priori, validación psicométrica completa del instrumento, actas reales de juicio experto, evaluación de sesgos y análisis por disciplina y experiencia. Si esas condiciones confirman el patrón modelado, PlanSim-AI podría aportar un marco replicable para investigar cómo la IA generativa y los sistemas multiagente pueden integrarse responsablemente a la planificación del aprendizaje, manteniendo la agencia humana como criterio rector.

Referencias

- Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100215. <https://doi.org/10.1016/j.caeai.2024.100215>
- Alkhadim, G. S. (2022). Cronbach's alpha and semantic overlap between items. *Frontiers in Psychology*, 13, 815490. <https://doi.org/10.3389/fpsyg.2022.815490>
- Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, 100489. <https://doi.org/10.1016/j.edurev.2022.100489>
- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21, 4. <https://doi.org/10.1186/s41239-023-00436-z>

- Chaka, C. (2023). Fourth industrial revolution—a review of applications, prospects, and challenges for artificial intelligence, robotics and blockchain in higher education. *Research and Practice in Technology Enhanced Learning*, 18, 2. <https://doi.org/10.58459/rptel.2023.18002>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chu, H.-C., Hwang, G.-H., Tu, Y.-F., & Yang, K.-H. (2022). Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. *Australasian Journal of Educational Technology*, 38(3), 22–42. <https://doi.org/10.14742/ajet.7526>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Daoudi, I. (2022). Learning analytics for enhancing the usability of serious games in formal education: A systematic literature review and research agenda. *Education and Information Technologies*, 27(8), 11237–11266. <https://doi.org/10.1007/s10639-022-11087-4>
- Del Gobbo, E., Guarino, A., Cafarelli, B., Grilli, L., & Limone, P. (2023). Automatic evaluation of open-ended questions for online learning: A systematic mapping. *Studies in Educational Evaluation*, 77, 101258. <https://doi.org/10.1016/j.stueduc.2023.101258>
- Fariani, R. I., Junus, K., & Santoso, H. B. (2023). A systematic literature review on personalised learning in the higher education context. *Technology, Knowledge and Learning*, 28(2), 449–476. <https://doi.org/10.1007/s10758-022-09628-4>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in*

- Education and Teaching International, 61(3), 460–474.
<https://doi.org/10.1080/14703297.2023.2195846>
- Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F., & Li, Y. (2024). Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications*, 11, 1259.
<https://doi.org/10.1057/s41599-024-03611-3>
- Herrera, P., Huepe, M., & Trucco, D. (2025). Educación y desarrollo de competencias digitales en América Latina y el Caribe. Comisión Económica para América Latina y el Caribe (CEPAL).
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
<https://doi.org/10.1016/j.lindif.2023.102274>
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074.
<https://doi.org/10.1016/j.caeai.2022.100074>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20, 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267.
<https://doi.org/10.1525/collabra.33267>
- Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6, 100221. <https://doi.org/10.1016/j.caeai.2024.100221>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410.
<https://doi.org/10.3390/educsci13040410>

- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.
- Ministerio de Educación del Ecuador. (2024). Orientaciones para el uso pedagógico de herramientas de inteligencia artificial en el proceso de enseñanza aprendizaje que garanticen el uso efectivo y ético en el aula. Ministerio de Educación del Ecuador.
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28, 4221–4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, 27, 7893–7925. <https://doi.org/10.1007/s10639-022-10925-9>
- Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, Article 2, 1–22. <https://doi.org/10.1145/3586183.3606763>
- Rabelo, A., Rodrigues, M. W., Nobre, C., Isotani, S., & Zárata, L. (2023). Educational data mining and learning analytics: A review of educational management in e-learning. *Information Discovery and Delivery*. <https://doi.org/10.1108/IDD-10-2022-0099>
- Rangel-de Lázaro, G., & Duarte, J. M. (2023). You can handle, you can teach it: Systematic review on the use of extended reality and artificial intelligence technologies for online higher education. *Sustainability*, 15(4), 3507. <https://doi.org/10.3390/su15043507>
- Salas-Pilco, S. Z., & Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, 19, 21. <https://doi.org/10.1186/s41239-022-00326-w>
- Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, 12(8), 569. <https://doi.org/10.3390/educsci12080569>

- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Ullrich, A., Vladova, G., Eigelshoven, F., & Renz, A. (2022). Data mining of scientific research on artificial intelligence in teaching and administration in higher education institutions: A bibliometrics analysis and recommendation for future research. *Discover Artificial Intelligence*, 2. <https://doi.org/10.1007/s44163-022-00031-7>
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J. (2024a). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18, 186345. <https://doi.org/10.1007/s11704-024-40231-1>
- Wang, N., Wang, X., & Su, Y.-S. (2024b). Critical analysis of the technological affordances, challenges and future directions of generative AI in education: A systematic review. *Asia Pacific Journal of Education*, 44(1), 139–155. <https://doi.org/10.1080/02188791.2024.2305156>
- Williamson, B. (2024). The social life of AI in education. *International Journal of Artificial Intelligence in Education*, 34(1), 97–104. <https://doi.org/10.1007/s40593-023-00342-5>
- Zakariya, Y. F. (2022). Cronbach's alpha in mathematics education research: Its appropriateness, overuse, and alternatives in estimating scale reliability. *Frontiers in Psychology*, 13, 1074430. <https://doi.org/10.3389/fpsyg.2022.1074430>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, 28. <https://doi.org/10.1186/s40561-024-00316-7>
- Zhong, L. (2022). A systematic review of personalized learning in higher education: Learning content structure, learning materials sequence, and learning readiness support. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2022.2061006>

Declaración

Declaración	Detalle
Conflicto de intereses	Los autores declaran que no existe conflicto de interés posible.
Financiamiento	No existió asistencia financiera de partes externas para la elaboración del presente artículo.
Agradecimiento	N/A
Nota	El artículo no es producto de una publicación anterior.