




# Analíticas del aprendizaje basadas en datos e inteligencia artificial en la educación superior: una revisión sistemática

Laura Icela González-Pérez , Francisco José García-Peñalvo , and Amadeo José Argüelles-Cruz 

**Abstract**— *La integración responsable de la inteligencia artificial en la educación (IAED) ofrece una oportunidad estratégica para alinear los entornos formativos con los principios de la Sociedad 5.0, potenciando la sinergia humano-tecnología en favor de una educación de calidad y del bienestar social. Este estudio presenta una revisión sistemática de 36 artículos arbitrados (2021–2025), centrados en aplicaciones educativas que emplean analíticas de aprendizaje (LA) con enfoques data-driven e integran modelos de Machine Learning (ML) como parte de su evidencia empírica. En cada estudio se identificaron tres elementos clave: el contexto de aplicación de la IAED, el enfoque data-driven adoptado y el modelo de ML utilizado. Los hallazgos revelan una desconexión entre los modelos de IA empleados y los datos educativos, los cuales, en muchos casos, se reducen a logs de acceso o calificaciones capturadas manualmente, que no permiten medir procesos cognitivos de manera profunda. Esta limitación compromete tanto la capacidad de los modelos ML para entrenarse de manera efectiva como su utilidad para ofrecer intervenciones pedagógicas útiles, como pueden ser rutas de aprendizaje personalizadas, retroalimentación en tiempo real, detección temprana de dificultades y seguimiento y visualización. Otro hallazgo relevante es la ausencia de marcos psicopedagógicos integrados a estándares de calidad y de gobernanza de datos, indispensables para avanzar hacia enfoques prescriptivos y éticos, coherentes con las metas de aprendizaje. Se recomienda que los líderes educativos promuevan aplicaciones de IAED sustentadas en marcos de gestión de datos y ética, asegurando métricas válidas y confiables que impulsen una educación más equitativa e inclusiva.*

**Palabras clave:** analítica del aprendizaje; aprendizaje automático; data-driven; aplicaciones; sistemas; educación superior; inteligencia artificial; Educación 4.0; innovación educativa; sociedad 5.0

## I. INTRODUCCIÓN

La educación basada en enfoques de datos (*data-driven*) está en aumento y aun persiste la ausencia de marcos psicopedagógicos y éticos que orienten la integración de la IA y consoliden una arquitectura de datos capaz de potenciar el desarrollo metacognitivo y el rendimiento académico. Rens & Hilbig [1] destacan la complejidad de integrar las analíticas del aprendizaje (LA, por sus siglas en inglés) y la IA para diseñar trayectorias adaptativas de enseñanza y aprendizaje. La minería de datos educativos (EDM) contribuye al desarrollo de técnicas que permiten identificar patrones de comportamiento estudiantil a partir de distintos tipos de datos: estructurados, semiestructurados y no estructurados [2], [3]. La LA y la EDM se han centrado en la medición y el análisis de datos estudiantiles, basados en registros como el tiempo dedicado y el acceso a materiales en sistemas de gestión del aprendizaje (LMS, por sus siglas en inglés, Learning Management System), con el propósito de contrastar hipótesis sobre hábitos de estudio, motivación y compromiso, con el fin de identificar patrones de comportamiento de aprendizaje y, de este modo, optimizar los procesos de aprendizaje [4], [5], [6], y [7]. Para diseñar un marco educativo basado en datos, son necesarios objetivos claros, datos de calidad y trazables procedentes de entornos virtuales, así como estructuras de gobernanza que permitan a la IA apoyar el aprendizaje de manera efectiva.

El modelado de procesos cognitivos [8] respaldado por IA y *Machine Learning* (ML), posibilita el aprendizaje personalizado [9], facilita la integración de fuentes heterogéneas de datos y el desarrollo de ecosistemas de aprendizaje escalables y adaptativos [10]. En paralelo, las herramientas de IA generativa en contextos académicos se adoptan cada vez con mayor frecuencia como apoyo para la atención, soporte e incluso formación personalizada [11]. Para integrar la IA en los sistemas educativos de aprendizaje de las Instituciones de Educación Superior (IES), es necesario promover la alfabetización en gobernanza de datos, ética y principios deontológicos que aseguren que los modelos de ML se entrenan con datos válidos, confiables, minimizando riesgos de sesgo y sustentados en métricas rigurosas de desempeño algorítmico.

De acuerdo con lo anterior, el objetivo de esta revisión es responder a las siguientes preguntas de investigación: (PI1): ¿Cómo ha evolucionado la utilización de métodos conceptuales, empíricos y/o mixtos en los estudios publicados entre 2021 y 2025 relacionados con enfoques basados en datos (*data-driven*) e integran modelos de IA con ML? (PI2) ¿Qué enfoques *data-driven* se utilizan en los estudios y a la vez emplean modelos ML? (PI3): ¿Qué aplicaciones de IAED se han implementado en las IES? (PI4): ¿Qué modelos de IA con ML han sido desarrollados?

## II. MARCO TEÓRICO

### A. Analítica de aprendizaje (LA) y Minería de Datos Educativos (EDM).

Para lograr una gobernanza efectiva de datos y la IA, las IES deben alinear su estrategia institucional con los estándares de gobernanza ética, y garantizar al mismo tiempo la identificación y mitigación de los riesgos asociados a su uso. La definición de LA se describe como la medición, recopilación, análisis y reporte de datos de los estudiantes para mejorar su aprendizaje y sus entornos [12], [13]. Los marcos de madurez digital y los estándares de calidad de los datos podrían ser las guías necesarias para sistematizar y enriquecer de manera estructurada los procesos de enseñanza-aprendizaje al integrarlos con modelos de IA, lo que ofrecería predicciones de los resultados académicos más libres de sesgos e inclusivas [14], [15]. El modelo CRISP-DM (*Cross-Industry Standard Process for Data Mining*, por sus siglas en inglés) es una metodología validada en múltiples sectores y aplicada en minería y ciencia de datos. Se concibe como un ciclo iterativo que abarca desde la comprensión del negocio y sus datos hasta la preparación, el modelado, la evaluación y el despliegue de las arquitecturas resultantes [16], [17]. Situar un marco de gobernanza en datos en el núcleo de un ecosistema de aprendizaje digital, permitirá articular todas las plataformas tecnológicas y favorecer la interoperabilidad entre diversos sistemas [18]. La OCDE insta a las IES a que validen, mediante evidencia científica, los beneficios y limitaciones de los datos utilizados y que busquen aprovechar cada vez más infraestructuras de macrodatos contextualizadas a las necesidades educativas [19]. Para que la

IA ofrezca funcionalidades libres de sesgos, es fundamental apoyarse tanto en modelos psicopedagógicos e instrumentos o escalas psicopedagógicas, como en las métricas de validación algorítmica del entrenamiento de modelos ML, y con ello contar métricas integrales que demuestren el impacto positivo de su uso.

### B. IAED y aplicaciones en los procesos de aprendizaje

La identificación de roles y contextos de docentes, estudiantes y líderes académicos permite diseñar marcos de competencias en IA y gobernanza de datos que impulsen modelos educativos sostenibles e innovadores sustentados en enfoques data-driven. El diseño instruccional que incorpora herramientas tecnológicas emergentes debe integrar la medición de procesos cognitivos con el fin de aprovechar la IA en la creación de estrategias de aprendizaje adaptativas con funcionalidades de autorregulación y métricas de desempeño [20]. En este marco, el *Cognitive Diagnostic Modeling* (CDM) ofrece una perspectiva psicométrica para identificar con precisión el dominio de habilidades específicas [21], mientras que el *Cognitive Task Analysis* (CTA) facilita la representación de estructuras complejas de conocimiento. Integrados en plataformas potenciadas por IA, estos enfoques pueden fortalecer la personalización del aprendizaje y generar retroalimentación inteligente, incrementando la eficacia de los procesos educativos [22]. in embargo, la ausencia de marcos psicopedagógicos sólidamente fundamentados en datos sobre el aprendizaje limita la vinculación entre las interacciones observadas y los resultados esperados; de ahí la necesidad de avanzar hacia modelos integradores que articulen teoría, evidencia empírica y prácticas educativas sostenibles [23].

Estas estrategias resultan esenciales para rediseñar entornos virtuales de aprendizaje orientados al desarrollo y escalamiento de competencias. Al integrar la influencia de factores socioeconómicos, culturales y ambientales, es posible diseñar itinerarios formativos que contribuyan a superar las brechas existentes y a fortalecer la equidad en los procesos educativos [24], [25], y potenciar el desarrollo de habilidades de orden superior, como el pensamiento complejo, el razonamiento crítico, la creatividad, la solución de problemas y el liderazgo, entre otras [26], [27]. Entrenar modelos ML con datos educativos fiables y bajo protocolos éticos rigurosos permite comprender con mayor precisión los patrones de comportamiento estudiantil, detectar áreas de mejora e incorporar prácticas innovadoras que fortalezcan procesos mentales complejos como el pensamiento crítico, la creatividad y la resolución de problemas [28]. El impacto pedagógico del uso de tecnologías como la IA solo será sostenible si se respalda con una gobernanza institucional que legitime el uso de los datos. La combinación de LA o EDM con el modelado conductual perfecciona la comprensión de los perfiles y comportamientos de los estudiantes [29], por ello se vuelve más necesario un marco de gobernanza para un uso ético y transparente de los datos. En esta línea, resulta fundamental determinar el tipo de interacción o aplicación educativa que se analiza. Algunos autores, [30] agrupan las aplicaciones de IAED en cuatro categorías: (1) aprendizaje, (2) enseñanza, (3) evaluación y (4) gestión administrativa. Otros estudios las clasifican en funciones más específicas, como el perfilado y

predicción del desempeño de los estudiantes [31], las herramientas de evaluación [32], los sistemas adaptativos y personalizados [33], [34] los sistemas de tutoría inteligente [35], y las de generación de contenido automático [36].

### C. Enfoques data-driven

El enfoque *data-driven* está transformando la manera en que las organizaciones evalúan sus procesos y fortalecen los mecanismos que les permiten generar y mantener ventajas competitivas. Los autores [37] sostienen que implementar un modelo *data-driven* requiere que las organizaciones: (1) transformen los datos recopilados en información significativa y conocimiento aplicable; (2) desarrollen una infraestructura tecnológica robusta junto con un ecosistema digital articulado; (3) desplieguen estrategias de analíticas e integración de datos para optimizar recursos; y (4) fortalezcan las capacidades de los actores clave. Algunos estudios enfatizan que la toma de decisiones basadas en datos permite integrar y analizar la información para obtener resultados más precisos. Este enfoque se basa en la recopilación, el análisis y la interpretación de grandes volúmenes de datos [38] [39], se ha consolidado como un pilar de la competitividad empresarial y la innovación, adquiriendo una relevancia cada vez mayor [40]. De manera similar, las IES están llamadas a transformarse gracias a su capacidad para convertir los datos en información procesable que permita cuantificar, apoyar y gestionar el aprendizaje, lo que exige el desarrollo de nuevas arquitecturas de datos. Como se sintetiza en la Tabla 1, los enfoques *data-driven* se clasifican en cuatro categorías principales: (1) diagnóstica, (2) descriptiva, (3) predictiva y (4) prescriptiva [41] [42].

TABLA I.  
CLASIFICACIÓN DE ENFOQUES DATA-DRIVEN  
BASADO EN [41], [42].

Métodos	Objetivo del análisis	Técnica
Enfoque diagnóstico	Utiliza los datos para identificar las causas subyacentes de tendencias y correlaciones. En este sentido, la regresión diagnóstica se orienta a responder la pregunta: «¿Por qué ocurrió esto?»	Prueba de hipótesis, correlación frente a causalidad y análisis de regresión
Enfoque descriptivo	Ofrece información sobre el pasado, lo que resulta útil para comunicar los cambios a lo largo del tiempo y utilizar las tendencias como base para estudios posteriores. Su propósito central es responder a la pregunta: «¿Qué ha ocurrido?».	Estadística descriptiva y técnicas de minería de datos.
Enfoque predictivo	Emplea datos históricos y patrones de tendencia para estimar escenarios futuros, orientándose a responder la pregunta: «¿Qué podría ocurrir?».	Técnicas de minería de datos, regresión y ML.
Enfoque prescriptivo	Utiliza los datos para cuantificar los posibles efectos de las decisiones futuras y ofrece recomendaciones para responder a la pregunta: «¿Qué deberíamos hacer a continuación?»	Reglas de negocio, diseño experimental, modelado computacional y ML, incluyendo Procesamiento del Lenguaje Natural (PLN)

Un enfoque data-driven, sustentado en información estructurada y en la aplicación de diversos métodos de análisis, permite examinar con mayor objetividad los comportamientos y actitudes de los usuarios, sobre todo en las plataformas de gestión del aprendizaje, y aporta pruebas muy valiosas para la

toma de decisiones y el diseño de estrategias pedagógicas. En este contexto, el ML adquiere un papel destacado al emplear enfoques *data-driven* para descubrir, predecir y clasificar patrones de comportamiento con el propósito de orientar la mejora continua de los procesos educativos [43]. El ML supervisado consiste en utilizar datos etiquetados con categorías predefinidas para entrenar y validar modelos predictivos o de clasificación. Entre los algoritmos más empleados se encuentran los árboles de decisión, los bosques aleatorios, el clasificador *Naïve Bayes* y las redes neuronales artificiales [44]. Sin embargo, estas técnicas plantean consideraciones éticas, especialmente en lo que respecta a la necesidad de validar las métricas de rendimiento y de verificar tanto las arquitecturas de datos como los procesos de etiquetado, con el fin de garantizar su fiabilidad y transparencia. Los sesgos presentes en instrumentos insuficientemente validados pueden amplificarse en modelos de IA, lo que resalta la importancia de validación rigurosa alineada con la gobernanza de datos. En esta línea, los principios *QuantCrit* examinan críticamente las prácticas predominantes en la investigación cuantitativa, como se destaca en [45] al enfatizar: (1) la centralidad del racismo, (2) la falta de neutralidad de los datos numéricos, (3) la naturaleza socialmente construida de categorías como género y raza, (4) la naturaleza interpretativa de los datos como un constructo producido socialmente y (5) el imperativo de un compromiso con la justicia social y la equidad. Una comprensión sólida de los fenómenos educativos, por tanto, requiere combinar la analítica cuantitativa con la indagación cualitativa. Asimismo, los constructos psicológicos, como la motivación, la personalidad o las creencias epistémicas, que a menudo se miden con métodos centrados en la persona, requieren pruebas estrictas de validez y fiabilidad, ya que, sin dicho escrutinio, son vulnerables desde el punto de vista metodológico y ético [46]. En el caso de los modelos de ML, incluida la IA generativa, resulta fundamental contar con mecanismos de detección de errores, sesgos y alucinaciones, ya que estas limitaciones pueden comprometer la fiabilidad y validez de las métricas, la seguridad y la confianza en su aplicación educativa.

#### D. Modelos de Inteligencia Artificial

La IA tiene el potencial de transformar los procesos educativos, siempre y cuando se establezcan arquitecturas sólidas de gobernanza de datos que garanticen la interoperabilidad de los sistemas y respalden la adopción de medidas basadas en pruebas. La IA se puede definir como un sistema informático capaz de adaptarse, aprender, sintetizar información, corregir errores y aprovechar datos heterogéneos para resolver tareas complejas [47]. Diversos autores [48] identifican tecnologías clave de la IA: (1) ML [49], (2) Procesamiento de lenguaje natural (PLN) [50], (3) Visión de computadora [51], y (4) Robotica [52]. El ML es un campo de estudio que utiliza algoritmos computacionales para transformar datos empíricos en modelos utilizables [53], abarca cuatro paradigmas principales: (a) supervisado, (b) no supervisado, (c) semisupervisado y (d) aprendizaje por refuerzo [54], [55] (Fig. 1).

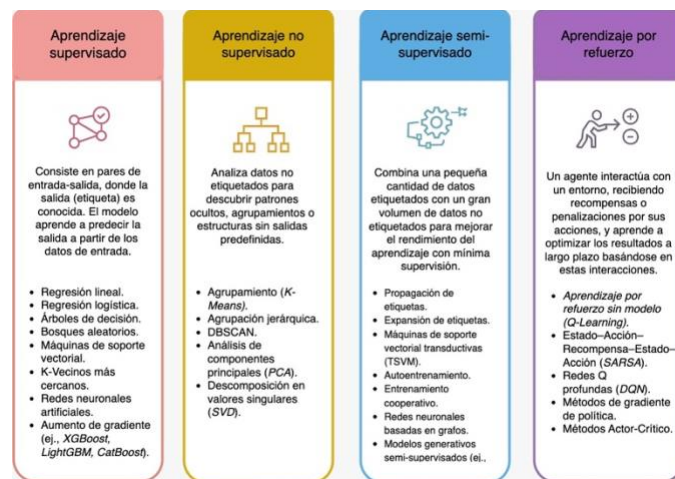


Fig. 1. Modelos ML [54] [55]

La implementación de IA en las IES puede implicar el procesamiento de datos sensibles de los estudiantes sin contar con mecanismos de seguridad plenamente consolidados, lo que incrementa el riesgo de un uso inadecuado de información privada y de una detección deficiente de indicadores relacionados con la salud mental, el bienestar emocional y los patrones de aprendizaje. La IA ha evolucionado desde la programación de tareas específicas y poco estructuradas, carentes de conexión con enfoques o modelos interoperables, hacia métodos sustentados en marcos formales. Entre estos últimos destacan los problemas de satisfacción de restricciones (*Constraint Satisfaction Problems, CSP*), las redes bayesianas, los procesos de decisión de Markov parcialmente observables y los juegos generalizados (*general game playing*), todos ellos abordados mediante algoritmos independientes del dominio [56]. Este paradigma respalda avances significativos en los sistemas expertos, los agentes inteligentes, bajo modelos predictivos y los entornos de aprendizaje personalizados [57], [58]. Los ML enriquecen aún más el análisis al revelar patrones latentes y generar conocimientos aplicables [59], lo que refuerza la necesidad de una gobernanza de la IA en educación que asegure la protección integral de los derechos humanos fundamentales de los involucrados.

Algunos autores destacan que la aplicación de modelos de ML en la educación ha puesto de manifiesto principalmente tres desafíos: (1) integrar datos heterogéneos para identificar a los estudiantes en riesgo, (2) garantizar una evaluación rigurosa de la validez metodológica y (3) establecer directrices fiables para el desarrollo y la evaluación de los modelos [60]. Ahmad et al. [61] enumera las aplicaciones clave basadas en datos —predicción del desempeño, aprendizaje personalizado y análisis de sentimientos— y enfatiza la necesidad de frenar los sesgos algorítmicos y preservar la privacidad de los estudiantes. El despliegue más amplio de sistemas de ML también ha suscitado preocupaciones sobre posibles resultados discriminatorios [62]. En consecuencia, las plataformas inteligentes deben basarse en constructos conductuales respaldados por pruebas [63] y deben diseñarse para garantizar la equidad, la transparencia, la rendición de cuentas y una gestión responsable de los datos, en consonancia con los objetivos instruccionales.

### III. METODOLOGÍA

En esta revisión sistemática de la literatura (RSL) se examina cómo los estudios conceptuales, empíricos y de métodos mixtos, publicados entre 2021 y 2025, emplean LA y EDM junto con enfoques *data-driven* y ML en aplicaciones educativas. Se catalogan los modelos de IA, se identifican las tendencias metodológicas y se evalúa la profundidad de la integración de ML e IA en procesos educativos. La revisión siguió el protocolo de tres etapas de Kitchenham —planificación, ejecución e informe— para garantizar el rigor metodológico [64]. Además, se siguieron las directrices para revisiones sistemáticas de literatura (RSL) propuestas por Brereton et al. [65] las cuales incluyen protocolos de búsqueda predefinidos, pruebas piloto, criterios claros de inclusión y exclusión, así como una documentación transparente de las estrategias de búsqueda y de los procedimientos de selección de estudios.

#### A. Fase de planeación

Con el fin de respaldar un análisis riguroso de la literatura, identificar vacíos de investigación y formular preguntas pertinentes, se delimitó el alcance de la revisión. Por tanto, se seleccionaron estudios que implementaron soluciones de IAED —plataformas, sistemas, interfaces o tecnologías relacionadas— basadas en enfoques *data-driven* que utilizan modelos ML en entornos educativos orientados por datos. Las preguntas de investigación que guiaron el estudio se resumen en la Tabla II.

TABLA II.  
PREGUNTAS DE INVESTIGACIÓN

Preguntas de investigación	Posibles respuestas
PI1. ¿Cómo ha evolucionado el uso de métodos empíricos y/o mixtos en los estudios publicados entre 2021 y 2025 relacionados con los enfoques basados en datos ( <i>data-driven</i> ) y ML en educación?	Empírico, Conceptual, Mixto (2021, 2022, 2023, 2024 y 2025)
PI2. ¿Qué enfoques <i>data-driven</i> se utilizan en los estudios e integran modelos de IA con ML?	Descriptivo, Diagnóstico, Predictivo, Prescriptivo
PI3. ¿Qué aplicaciones de IA en educación se han implementado en las IES?	Aprendizaje, Enseñanza, Evaluación, Administración o Transversal
PI4. ¿Qué modelos de IA con ML se han desarrollado?	Aprendizaje supervisado, aprendizaje no supervisado, Semisupervisado, aprendizaje por refuerzo

El protocolo del SLR se basa explícitamente en el marco PICOC [66]:

#### i. Marco de referencia PICOC

- Población: estudiantes, profesores y gestores académicos de las IES.
- Intervención: identificación de estudios que apliquen metodologías de LA o EDM basadas en datos que incorporen ML para mejorar los resultados de aprendizaje.
- Comparación: no aplicable, ya que la revisión no incluye intervenciones comparativas.
- Resultado: análisis de cómo los estudios empíricos integran el análisis de datos basado en ML dentro de IAED, destacando el rigor metodológico, la implementación del modelo de IA y los resultados basados en evidencia.
- Contexto: Uso educativo de IAED y LA en entornos de IES entre 2021 y 2025.

#### ii. Criterios de inclusión y exclusión

Criterios de inclusión

- Publicado entre enero de 2021 y el 27 de febrero de 2025.
  - Tipo de documento: Artículos.
  - Idioma: Inglés.
  - Estudios que aplican modelos de ML.
- Criterios de exclusión:

- Se trata de artículos para los que no se disponía de acceso al texto completo a pesar de los esfuerzos razonables de recuperación.
- Estudios publicados en idiomas distintos del inglés.
- Estudios que no aplican modelos de ML.

#### iii. Estrategia de búsqueda

El proceso de búsqueda en las bases de datos Scopus y Web of Science (WoS) comenzó con la definición de las palabras clave que se utilizarían en sus motores de búsqueda. Para seleccionar los artículos que conforman la base de datos, se establecieron criterios de inclusión y exclusión de acuerdo con el marco PICC y las siguientes palabras clave:

*"artificial intelligence" OR "smart" OR "intelligent" OR "machine learning") AND "learning analytics" OR "educational data mining" ) AND "platform" OR "system" OR "Interface" OR "Technology" OR "framework" OR "environment" OR "process") AND ("data driven")"*.

La búsqueda se realizó el 27 de febrero de 2025. El período 2021-2025 se eligió para reflejar el reciente auge en la adopción de tecnologías de aprendizaje basadas en datos y modelos de IA. Este marco temporal coincide con el contexto pospandémico, que ha impulsado la transformación digital de las IES.

#### B. Fase de revisión

La revisión sistemática siguió las pautas PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, por sus siglas en inglés), como se ilustra en la Figura 2 [67], [68]. Esta metodología ampliamente reconocida abarca cuatro etapas clave: identificación, selección, evaluación de elegibilidad basada en criterios de exclusión e inclusión final.

Los datos se sistematizaron utilizando Excel y Tableau para generar tabulaciones cruzadas y representaciones visuales. Inicialmente, Excel se utilizó para fusionar los conjuntos de datos y eliminar las entradas duplicadas. Después, se creó una nueva hoja de trabajo para clasificar los estudios, y se asignó a cada uno un código de identificación único basado en el orden alfabético según el título del artículo. A continuación, se crearon las siguientes columnas para categorizar los aspectos clave de cada estudio:

- Tipo de método de investigación: conceptual, empírico o mixto.
- Tipo de enfoque basado en datos: descriptivos, diagnósticos, predictivos o prescriptivos.
- Enfoque LA o EDM
- Modelo de ML: aprendizaje supervisado, aprendizaje no supervisado, semisupervisado, aprendizaje por refuerzo.
- Tipo de aplicaciones de IA en educación: aprendizaje, enseñanza, evaluación, procesos administrativos o aplicaciones transversales.
- Descripción del estudio

- Algoritmos utilizados
- Observaciones

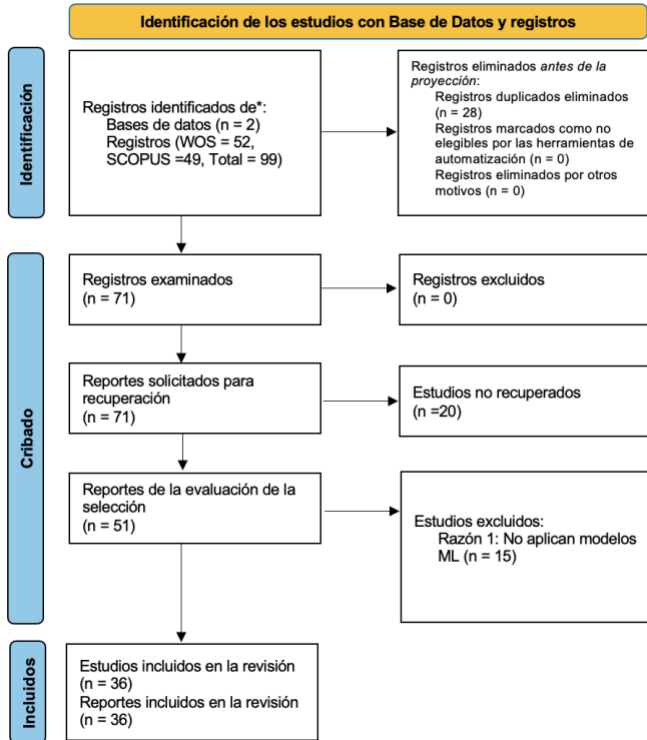


Fig. 2. Diagrama de flujo de PRISMA 2020

### C. Fase de reporte

El tratamiento de los datos se ajustó a las normas éticas y se basó en las pruebas recopiladas. De acuerdo con las recomendaciones de Traxler, se analizaron las observaciones de la plataforma con respecto a las preguntas de investigación y las categorías de codificación del estudio [69]. Posteriormente, se excluyeron 15 estudios tras comprobar mediante un análisis más detallado que no empleaban modelos de ML. El conjunto de datos se encuentra disponible en Zenodo en <https://zenodo.org/records/15802477>.

## IV. RESULTADOS

La Tabla III orienta sobre los 36 estudios revisados por pares publicados entre 2021 y 2025 que aplican métodos de datos impulsados por ML a la educación habilitada por IA. Enumerados alfabéticamente por título e indexados con identificadores consecutivos Id, estos trabajos documentan colectivamente los avances en curso en la transformación digital de la enseñanza y el aprendizaje. El identificador es el número de identificación del artículo y se utilizará con este fin en las siguientes secciones.

TABLA III.  
LISTA DE ESTUDIOS INCLUIDOS

Id	Cite	ID	Cite	ID	Cite
S2	[70]	S20	[71]	S33	[72]
S3	[73]	S21	[74]	S34	[75]
S4	[76]	S22	[77]	S35	[78]
S5	[79]	S23	[80]	S37	[81]

S7	[82]	S24	[83]	S38	[84]
S8	[85]	S25	[86]	S40	[87]
S9	[88]	S26	[89]	S41	[90]
S11	[91]	S27	[92]	S43	[93]
S12	[94]	S28	[95]	S44	[96]
S13	[97]	S29	[98]	S46	[99]
S14	[100]	S30	[101]	S49	[102]
S17	[103]	S31	[104]	S50	[105]

La cobertura geográfica se determinó en función de la afiliación del autor correspondiente, lo que reveló una fuerte presencia de universidades de regiones de Europa, Asia y Norteamérica. Europa representó 14 estudios, entre ellos: España (S17, S44), Suecia (S25, S31), Turquía (S21, S41), Italia (S5, S20), Austria (S7), Croacia (S49), Francia (S3), Alemania (S13), Noruega (S33) y Portugal (S37). Asia contribuyó con 12 estudios, con Japón (S12, S38, S43), China (S2, S40), Indonesia (S9, S50), Bangladesh (S11), India (S28), Malasia (S4), Viet Nam (S24) y Arabia Saudita (S23). Las Américas incluyeron cuatro estudios: EE.UU. (S26, S29, S30) y Canadá (S22). Oceanía estuvo representada por cuatro estudios de Nueva Zelanda (S14, S34, S46) y Australia (S8). Finalmente, África contribuyó con dos estudios, de Argelia (S35) y Sudáfrica (S27). La ausencia de estudios sobre América Latina pone de manifiesto una brecha crítica en la investigación actual sobre el uso de la IA y los modelos ML basados en enfoques *data-driven*. Esto refleja no solo la subrepresentación regional, sino también la escasez de documentación empírica sobre los resultados de su aplicación en educación. Estas asimetrías ponen de manifiesto la necesidad de impulsar esfuerzos empíricos que permitan elaborar una agenda de investigación sobre IAED en América Latina que sea representativa desde el punto de vista geográfico, institucional y disciplinar.

Para responder a la PI1—¿Cómo ha evolucionado el uso de métodos conceptuales, empíricos y/o mixtos en estudios publicados entre 2021 y 2025 relacionados con enfoques *data-driven* y ML en educación?— De los 36 estudios analizados, el 8% (tres estudios) corresponde a métodos conceptuales, el 78% (28 estudios) a estudios empíricos y el 14% (cinco estudios) utilizan métodos mixtos. En la Fig. 3 se observa la evolución temporal tanto del método de investigación como del modelo de ML utilizado en los estudios analizados.

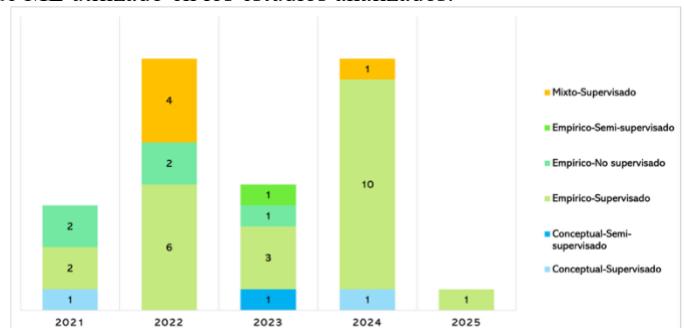


Fig. 3. Métodos de investigación y modelos ML (2021-2025)

Para responder a la PI2—¿Qué enfoques *data-driven* se utilizan en los estudios e integran modelos de IA con ML?—se



adoptó el marco de enfoques *data-driven* propuesto en [41] y [42], cuyos detalles se presentan en la Tabla I, así como por los modelos ML basados en [54] [55] en la Fig. 1. En la Tabla IV se presentan los estudios organizados de acuerdo con estas clasificaciones. Cada clasificación se codificó iniciando con el método de investigación empleado: C (conceptual), E (empírico) y M (mixto).

TABLA IV.  
CLASIFICACIÓN DE LOS ESTUDIOS POR ENFOQUE DATA-DRIVEN Y ML [41] y [42]  
Estudios con modelos de ML supervisado (n=29)

E-Diagnóstico & Predictivo (4)	S5, S35, S38, S50
C-Descriptivo (1)	S7
E-Descriptivo & Predictivo (1)	S27
E-Predictivo (13)	S4, S9, S12, S13, S21, S24, S28, S30, S33, S37, S41, S43, S20
C-Predictivo (1)	S23
E-Predictivo & Prescriptivo (5)	S25, S31, S34, S40, S44
E- Prescriptive (4)	S2, S11, S14, S46
<b>Estudios con ML no supervisado (n=5)</b>	
E-Diagnóstico & Descriptivo (2)	S8, S29
E-Diagnóstico & Predictivo (1)	S49
E-Descriptivo (1)	S17
E-Descriptivo & Predictivo (1)	S26
<b>Estudios con ML semisupervisado (n=2)</b>	
C-Prescriptivo (1)	S22
E-Predictivo (1)	S3

En síntesis, de esta clasificación de enfoques *data-driven* se observa que el predictivo es el predominante, al estar presente en 19 de los 36 estudios analizados. Seguido por los enfoques diagnóstico (7 estudios), descriptivo (6 estudios) y prescriptivo (6 estudios). Cabe destacar que el enfoque diagnóstico aparece con frecuencia combinado con otros enfoques, funcionando como componente transversal en las etapas iniciales de las implementaciones basadas en datos.

*Análisis diagnóstico.* Se identificaron un total de siete estudios con este enfoque. Cuatro de ellos (S5, S35, S38 y S50) emplearon ML supervisado en combinación con un enfoque predictivo. Por su parte, los estudios S8 y S29 aplicaron ML no supervisado junto con un enfoque descriptivo. El estudio S49 utilizó ML no supervisado en conjunto con enfoque predictivo. Estos resultados ponen de manifiesto la importancia de incorporar el enfoque diagnóstico en las fases iniciales como elemento fundamental para evolucionar hacia el diseño de estrategias más avanzadas basadas en datos, como los enfoques predictivos o prescriptivos.

*Análisis descriptivo.* Se identificaron seis estudios con este enfoque. Dos de ellos (S7 y S27) emplearon modelos ML supervisado, mientras que otros dos (S17 y S26) utilizaron ML no supervisado. Cabe señalar que los estudios S26 y S27 combinaron además el enfoque descriptivo con el predictivo. Los estudios (S8) y (S29) se combinaron con enfoque diagnóstico Y ML no supervisado.

*Análisis predictivo.* Se identificaron un total de 19 estudios con este enfoque. De ellos, trece (S4, S9, S12, S13, S20, S21, S24, S28, S30, S33, S37, S41 y S43) aplicaron exclusivamente un enfoque predictivo. Los seis restantes combinaron el análisis

predictivo con el prescriptivo: cinco de ellos (S25, S31, S34, S40 y S44) emplearon modelos ML supervisado, mientras que uno (S3) utilizó un modelo de ML semisupervisado.

*Análisis prescriptivo.* Dentro de este grupo, destaca el estudio S22 por ser el único que adoptó un enfoque puramente prescriptivo, empleando exclusivamente un modelo de ML semisupervisado. Esta singularidad lo diferencia de los demás estudios identificados, que aplicaron modelos de ML supervisado —ya sea de manera exclusiva (S2, S11, S14 y S46) o en combinación con enfoques predictivos (S25, S31, S34, S40 y S44)—, lo que convierte a S22 en un caso especialmente relevante dentro del corpus analizado.

Para responder a PI3—¿Qué aplicaciones de IAED se han implementado en los estudios?—Las aplicaciones de IAED se clasificaron en cinco categorías principales: (1) aprendizaje, (2) enseñanza, (3) evaluación, (4) administración [30] y (5) transversal, como se detalla en la Tabla V. Cabe señalar que, con el propósito de incorporar enfoques de carácter multidisciplinar y con incidencia en diversos procesos y ámbitos de conocimiento, se introdujo la categoría transversal. Esta agrupa estudios relacionados con ámbitos como tableros visuales para el monitoreo del aprendizaje, los sistemas multiagente, los entornos adaptativos y personalizados [33] [34], así como herramientas para la generación automática de contenidos [35].

TABLA V.  
MODELOS ML EN APLICACIONES IAED  
Aplicaciones con ML Supervisado (n=29)

Gestión administrativa (1)	Sistemas de recomendación	S44
Aprendizaje (8)	Predicción de carrera basada en IA	S23
	Retroalimentación basada en IA	S11, S14, S46
	Retroalimentación basada en IA: aprendizaje autorregulado (SLR)	S25, S31
	Aprendizaje colaborativo	S2
Enseñanza (1)	Estilo de aprendizaje	S35
	Modelado de estudiantes	S40
Transversal (1)	Tableros de aprendizaje	S34
	Evaluación de código y clasificación de errores	S38
Evaluación (18)	Predicción de abandono escolar	S5
	Predicción de deserción escolar - Sistemas de alerta temprana (EWS)	S7, S20, S30, S37, S43
	Sistemas de tutoría inteligente (STI)	S13
	Predicción de rendimiento y logros	S27, S50, S9
	Predicción e interpretabilidad de riesgos	S12, S21, S24, S28, S4, S41
	Evaluación basada en sensores y análisis afectivo	S33
<b>Aplicaciones con ML No Supervisado (n=5)</b>		
Aprendizaje (4)	Modelado de estudiantes y análisis de comportamiento	S8, S26, S49
	Modelado y colaboración de estudiantes Análisis del comportamiento	S17
Enseñanza (1)	Interacción social y monitoreo de participación	S29
<b>Aplicaciones con ML Semisupervisado (n=2)</b>		
Aprendizaje (1)	Agentes pedagógicos con IA generativa	S22

Evaluación (1)	Predicción de deserción escolar - Sistemas de alerta temprana (EWS)	S3
----------------	---	----

En la Tabla V se reportan 19 estudios clasificados en la categoría de evaluación, lo que representa aproximadamente el 53 % del total de investigaciones que aplican modelos de ML, de los cuales 18 emplean ML supervisado y uno, ML semisupervisado (S3). Aunque todos los estudios se enmarcan en la evaluación, abarcan diversas aplicaciones funcionales. Una parte significativa se centra en la medición del rendimiento académico mediante enfoques descriptivos y predictivos (S27, S9), así como en la identificación de riesgos de bajo rendimiento y deserción (S12, S21, S24, S28, S4, S41). Destacan estudios que implementan sistemas de alerta temprana (*Early Warning Systems*, EWS) para prevenir el abandono escolar (S7, S20, S30, S37, S43, S3), y el estudio S13, centrado en el desarrollo de un sistema de tutoría inteligente (*Intelligent Tutoring System*, ITS).

En la categoría de aprendizaje, y debido a su carácter diagnóstico y a su contribución formativa dentro de los procesos educativos, se identificaron un total de trece estudios. Cinco de ellos abordan estrategias de retroalimentación personalizada, sustentadas en modelos de ML supervisado con un enfoque prescriptivo (S11, S14, S46, S25, S31). Por su parte, el estudio S35 se centra en la identificación de estilos de aprendizaje mediante un enfoque predictivo, mientras que S2 propone mecanismos de apoyo para el aprendizaje colaborativo, y S23 se orienta en la predicción de trayectorias académicas mediante IA. En esta misma categoría se identificaron cuatro estudios (S8, S26, S49, S17) que emplean ML no supervisado, y que están enfocados en el análisis del comportamiento y la colaboración estudiantil. El estudio S22 para dar retroalimentación a través de agentes para potenciar las habilidades de aprendizaje autorregulado de los estudiantes.

En cuanto a las aplicaciones centradas en la enseñanza, orientadas a apoyar las tareas docentes, con 7% de representación, —como la identificación de la interacción social y el seguimiento de la participación estudiantil—, se identificaron dos estudios: uno con enfoque diagnóstico (S29) y otro que combina análisis predictivo y prescriptivo para el modelado de estudiantes a partir de comportamientos (S40).

Respecto a las aplicaciones administrativas, se identificó un único estudio (S44), equivalente al 3 % del total, centrado en el desarrollo de un sistema de recomendación basado en análisis diagnóstico. Finalmente, dentro de la categoría de aplicaciones transversales, se destaca el estudio S34, que propone el diseño de tableros de aprendizaje orientados la supervisión del progreso estudiantil.

Para responder a la PI4—*¿Qué modelos de IA con ML han sido desarrollados?*— los resultados muestran que el ML supervisado predomina en el corpus de 36 estudios (80.6 %), mientras que los enfoques no supervisados (13.8 %) y semisupervisados (5.6 %) son considerablemente menos frecuentes (Fig. 4).

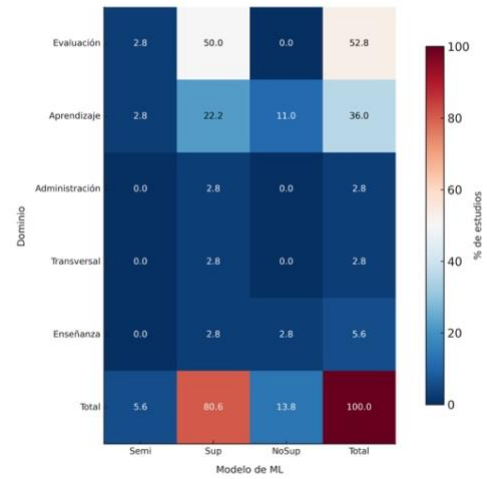


Fig. 4. IAED con modelos ML

Dentro del ML supervisado, cerca del 50 % de los estudios se enfocan en la evaluación educativa —predicción del rendimiento, detección del abandono escolar y tutoría inteligente (S38, S5, S7, S13, S27, S41, entre otros)— para las aplicaciones de enseñanza (S40), gestión administrativa (S44) y transversales (S34), se identificó únicamente un estudio por categoría, equivalente al 2.8 % del total en cada caso; El 22.2 % correspondió a aplicaciones de aprendizaje (S2, S11, S14, S23, S25, S31, S35 y S46). En ML no supervisado, el 11 % corresponde a aplicaciones de aprendizaje, vinculadas con el análisis del comportamiento y la colaboración estudiantil (S8, S26, S49, S17), y un estudio (2.8 %) a la enseñanza (S29). Con ML semisupervisado se encuentran dos estudios: S22 orientado al aprendizaje personalizado mediante IA generativa con agentes pedagógicos y S3 centrado en la evaluación con sistemas de alerta temprana para prevenir la deserción escolar.

## V. DISCUSIÓN

La incorporación de modelos de decisión *data-driven* en educación requiere establecer marcos de gestión sólidos y principios de responsabilidad que garanticen la transparencia, la confianza y el respeto a los derechos fundamentales. Los 36 estudios de la tabla III muestran que los modelos de decisión *data-driven* pueden ser fundamentales para integrar modelos de IA y beneficiar a las soluciones educativas ofrecidas a la comunidad académica. Los 36 estudios de la Tabla III muestran cómo los modelos *data-driven* son fundamentales para integrar la IA y generar soluciones educativas al servicio de la comunidad académica. En las instituciones de educación superior (IES), su adopción debe avanzar gradualmente, comenzando por la alfabetización en datos y la articulación de ecosistemas digitales aún fragmentados. Es importante establecer principios éticos y de transparencia en el manejo de la información, así como impulsar proyectos piloto que, una vez evaluados, puedan ampliarse para fortalecer la gestión institucional y los procesos de enseñanza y aprendizaje.

Para responder a la PI1, se identificó que el 78% de los estudios analizados corresponden a diseños empíricos y el 14% a enfoques mixtos, lo que en conjunto representa una cobertura empírica del 92%, frente a un 8% de carácter conceptual (Fig. 3). Este predominio empírico pone de relieve la necesidad de comprobar de manera empírica las implementaciones al

integrar LA con modelos de IA en contextos de enseñanza y aprendizaje [1]. Una vía para simplificar estos procesos consiste en iniciar con enfoques diagnósticos basados en los datos disponibles e ir avanzando progresivamente hacia modelos algorítmicos más sofisticados y con mayor desempeño, sustentados en el análisis prescriptivo, empleando enfoques *data-driven*. Estos permitirían anticipar comportamientos de aprendizaje en las plataformas, ya sea a partir de datos etiquetados o de interacciones no estructuradas, con el fin de alcanzar una mayor granularidad en la información y posibilitar una detección temprana del desempeño académico [29]. En ese sentido, la incorporación de modelos ML permite anticipar posibles barreras en el desempeño académico y establecer sistemas de alerta temprana que favorezcan procesos de aprendizaje adaptativo a partir de los datos disponibles. Sin embargo, estos avances requieren estar sustentados en marcos sólidos de gobernanza de datos [9], ya sea para digitalizar procesos cognitivos [8] o para predecir y optimizar las interacciones en entornos digitales.

Aunque los estudios empíricos son los más representados, la limitada disponibilidad de datos restringe la trazabilidad de los procesos cognitivos, lo que subraya la necesidad de arquitecturas robustas de gobernanza que se articulen con los objetivos pedagógicos y potencien el valor del aprendizaje adaptativo y personalizado. Una efectiva selección de enfoque *data-driven* debe alinearse con los objetivos institucionales y aumentar la disponibilidad de datos que permitan coleccionar información relevante.

La Tabla IV presenta los métodos de decisión *data-driven* empleados y los modelos de ML utilizados en los estudios revisados. Los métodos diagnósticos permiten identificar las causas de las posibles problemáticas a partir de correlaciones observadas. En el estudio S8 se revela cómo crear estrategias de aprendizaje de anticipación a partir de registros de plataforma de aprendizaje, mientras que el estudio S49 aborda cómo el enfoque diagnóstico y predictivo centrado en datos utiliza modelos ML no supervisado para generar perfiles de comportamiento y ofrecer retroalimentación personalizada en tiempo real. Los enfoques diagnósticos pueden incrementar su valor si, desde el inicio, integran principios éticos como los de *QuantCrit* [45], que subrayan la necesidad de interpretar los datos conductuales y psicológicos considerando el contexto social en el que se aplican. Dichos principios advierten sobre los riesgos de que las categorías empleadas para medir competencias introduzcan sesgos que, en lugar de apoyar el aprendizaje, terminen reforzando desigualdades educativas o reproduciendo interpretaciones dependientes del juicio humano.

Los estudios que emplean métodos descriptivos han arrojado hallazgos relevantes. El estudio S26 presenta un modelo centrado en datos de diferencias individuales que mejora la predicción del rendimiento y el compromiso. Este modelo se alinea con estrategias de aprendizaje personalizadas respaldadas por evidencia empírica. Al mismo tiempo, el estudio S17 examinó los datos de interacción de Coursera, que comprenden registros de flujo de clics y datos demográficos, para identificar patrones de colaboración y posibles casos de

deshonestidad académica. Los métodos *data-driven* se benefician de marcos estructurados como CRISP-DM, que ofrece un ciclo de vida iterativo adecuado para validar iniciativas educativas de ciencia de datos y garantizar la coherencia entre las preguntas institucionales, los datos disponibles y el diseño analítico [16], [17].

Se identificaron estudios que emplean métodos predictivos como S12, S25, S28 y S31 para evaluar el desempeño de los estudiantes a partir de los registros del sistema de gestión del aprendizaje. Por el contrario, los estudios S4, S9 y S41 emplean el OULAD para evaluar la generalización de las predicciones de los modelos. En S31, se entrenaron seis algoritmos de ML con datos de 446 estudiantes y el modelo de mejor rendimiento impulsado fue el sistema de IA explicable prescriptiva (XAI) que generó comentarios personalizados entregados a través de un tablero alineado con el modelo SRL de Zimmerman, que respalda el establecimiento de objetivos, la planificación, el monitoreo y la reflexión. El estudio S25 aplicó un modelo explicable utilizando datos de una plataforma de aprendizaje para predecir el rendimiento y respaldar la autorregulación a través de comentarios personalizados y un panel automatizado, con miras a incluir funcionalidades de gestión del tiempo y programación que pueden ayudar aún más. Los métodos predictivos se benefician de los marcos de madurez de ciencia de datos, que respaldan la previsión de resultados y la toma de decisiones institucionales estructuradas [14], [15]. Sin embargo, su implementación debe guiarse por la gobernanza de datos y los marcos éticos para evitar sesgos, garantizar la privacidad y seguridad de los datos e incluir una supervisión y monitoreo para evitar la discriminación o desigualdad.

Se identificaron estudios que empleaban métodos prescriptivos centrados en el uso de datos para recomendar acciones y cuantificar los resultados potenciales con el fin de optimizar la toma de decisiones. Entre ellos, el estudio S22 desarrolló un mecanismo basado en datos que permite a los agentes pedagógicos mejorados con IA generativa proporcionar retroalimentación basada en LA, lo que ayuda a los alumnos a reflexionar sobre sus procesos de aprendizaje y a desarrollar habilidades de aprendizaje autorregulado. El estudio S14, que utiliza un conjunto de datos reales de una IES de Australia con 7000 estudiantes (el 52% completó el curso y el 48% lo abandonó), procedentes de plataformas como Moodle, muestra cómo se puede mejorar el modelado predictivo con análisis prescriptivos. Mediante dos estudios de casos, el estudio demuestra cómo se puede utilizar ChatGPT para generar comentarios claros y procesables para estudiantes en riesgo. El análisis prescriptivo requiere datos de alta calidad, modelos confiables y procesos sólidos para generar recomendaciones fiables y pedagógicamente sólidas [32], [33]. Es esencial situar la capa de análisis de datos en el centro de un ecosistema que conecta sistemas ciberfísicos, plataformas organizativas y de aprendizaje, que sirven de nexo para las partes interesadas clave en las IES [18] y definir cómo se recopilarán, analizarán e interpretarán los datos, especialmente cuando se trabaja con información a gran escala [34]. El enfoque prescriptivo no solo prevé, sino que también recomienda acciones, cuya eficacia depende de datos fiables y validados alineados con el modelo



educativo, por lo que resulta necesario validar la eficacia del modelo algorítmico empleado para evitar falsos positivos.

Entender el contexto educativo —lo que en el marco CRISPD-DM se denomina “Entendimiento del negocio” — es un requisito clave para desarrollar aplicaciones educativas de IA basados en soluciones con evidencia y bajo estándares de calidad. La Tabla V clasifica y presenta la técnica empleada en los modelos de ML a las aplicaciones IAED revisadas, y con ello responde a la pregunta PI3. El estudio S22 examina la integración de agentes generativos de IA y esboza los principios pedagógicos esenciales para guiar a los docentes en el uso eficaz de los chatbots de IA en contextos educativos. El estudio S29 emplea datos de sensores que respaldan el aprendizaje colaborativo al detectar cercanía, mientras que el estudio S40 propone un modelo optimizado de arquitectura de datos para enfoques adaptativos. Los estudios S25 y S31 demuestran que la retroalimentación explicable basada en IA puede fortalecer el aprendizaje autorregulado y ampliar el análisis más allá de la predicción de abandono para centrarse en el desarrollo de competencias. Los marcos CDM y CTA son ejemplos de cómo estas teorías pueden integrarse en las plataformas de IA para personalizar el aprendizaje y ofrecer una retroalimentación inteligente [21], [22]. Para potenciar el uso de enfoques data-driven, las IES deben tener más claridad al definir el propósito para usar IA —como el uso de chatbots conversacionales, sistemas de tutoría inteligente o modelos predictivos de riesgo académico— y formular hipótesis que permitan obtener resultados fiables con métricas tanto psicopedagógica como de performance algorítmico.

La adopción cada vez mayor de la IA en la educación pone de manifiesto las lagunas existentes en el diseño pedagógico, la validación ética y la integración de los procesos cognitivos en los sistemas de análisis del aprendizaje. Para responder a la pregunta PI4, la Fig. 4 muestra que el 80.6 % de los estudios encontrados utilizan ML supervisado, de los cuales el 62 % se centra en aplicaciones relacionadas con la evaluación educativa. Esto nos reta a avanzar hacia modelos semisupervisados y no supervisados e incorporar aplicaciones de enseñanza, aprendizaje y gestión. Entre las soluciones más prometedoras destacan la IA explicable (XAI) y los cuadros de mando inteligentes. En los estudios S14, S25 y S31, la XAI se utiliza para proporcionar una retroalimentación transparente y personalizada que apoya la toma de decisiones. En concreto, el estudio S31 la emplea con fines prescriptivos, recomendando acciones correctivas, mientras que el estudio S25 la utiliza para identificar las causas del bajo rendimiento académico y fortalecer las habilidades de autorregulación. Por su parte, los estudios S14 y S34 implementan cuadros de mando inteligentes que optimizan el seguimiento en tiempo real, fomentan la autorregulación del alumnado y contribuyen a la mejora continua de las estrategias de enseñanza. Los modelos de ML fortalecen el análisis educativo al revelar patrones latentes y generar información procesable [59]. Por lo tanto, su aplicación debe basarse en constructos conductuales con base empírica [63] y estar acompañada de una evaluación rigurosa de la validez metodológica para garantizar la fiabilidad de la evaluación del modelo [60]. La integración de modelos de ML

en aplicaciones educativas debe fundamentarse en el consentimiento informado, en instrumentos y constructos validados, y en evaluaciones metodológicas rigurosas que aseguren la validez predictiva de las dimensiones educativas, la fiabilidad del desempeño algorítmico y el cumplimiento de principios éticos de equidad y dignidad humana.

## VI. CONCLUSIONES Y ESTUDIOS FUTUROS

De los 36 estudios analizados entre 2021–2025 se encontró que la mayoría de las aplicaciones de IA en educación se basan en ML supervisado y enfoques predictivos, principalmente en evaluación de bajo rendimiento y deserción. La revisión identificó un solo estudio con aplicaciones educativas transversales, lo que evidencia una brecha significativa en la integración de herramientas y procesos diversos. La ausencia de investigaciones que documenten metodológicamente el uso, alcance y limitaciones de LA, ML e IA en educación resalta la urgencia de contar con marcos de gobernanza que orienten su aplicación de manera segura, ética y responsable, como condición esencial para el diseño de ecosistemas educativos centrados en el bienestar humano y alineados con los principios de la Sociedad 5.0. Asimismo, los hallazgos enfatizan la necesidad de que las IES adopten marcos integrales que garanticen la calidad de los datos, la privacidad, la mitigación de sesgos y la protección neuroética. Los modelos de ML revisados en los estudios revelan posibilidades para mejorar habilidades cognitivas como, la autorregulación. Para las IES, adoptar la IA no solo implica mejorar la precisión y la pertinencia pedagógica, sino también generar confianza, proteger los datos y consolidar ecosistemas alineados con los objetivos educativos. Esto requiere una gobernanza sólida para el uso de la IA generativa, mediante implementaciones propias adaptadas al contexto institucional. Funcionalidades como la generación aumentada de recuperación (RAG), la creación de bases de conocimiento seleccionadas y la adaptación específica al dominio (*fine-tuning*), junto con la evaluación de los costes de infraestructura, son pasos clave. Del mismo modo, es fundamental integrar *LLM Guardrails* para garantizar la transparencia, reducir los sesgos y asegurar la alineación con principios éticos y psicopedagógicos.

Los horizontes de investigación emergentes deben adoptar enfoques basados en métodos mixtos, en datos y en seguridad, realizar una validación rigurosa de los instrumentos de evaluación, acercar las teorías psicopedagógicas y la neurotecnología para desarrollar competencias futuras e incorporar directrices de gobernanza como *QuantCrit* y métodos de inspección u observación orientados al uso ético en entornos de aprendizaje, la detección de lenguaje inapropiado y aspectos sociales relacionados, a partir de herramientas de detección temprana y el uso de métricas de evaluación multidimensionales. El desafío consiste en que educadores y líderes institucionales pasen de la adopción pasiva a la promoción activa de aplicaciones de IA éticamente alineadas con la formación y el desarrollo humano. Esto requiere un proceso progresivo de alfabetización que acompañe la automatización de procesos y permita evaluar el impacto en el trabajo y en la conciencia social. La IA en las IES debe diseñarse conforme a valores sociales y principios éticos,

reconociendo que los datos reflejan tanto conocimientos como sesgos, y que su integración, bajo marcos de gobernanza sólidos, puede contribuir a una educación más equitativa, inclusiva y digna.

#### MANIFESTACIONES Y DECLARACIONES

Los autores declaran no tener conflictos de intereses con respecto a la investigación, autoría y/o publicación de este artículo.

#### AGRADECIMIENTOS

Los autores agradecen el apoyo del Instituto Politécnico Nacional en el marco del proyecto SIP-20250146.

#### REFERENCIAS

- [1] A. Renz and R. Hilbig, "Prerequisites for artificial intelligence in further education: identification of drivers, barriers, and business models of educational technology companies," *Int J Educ Technol High Educ*, vol. 17, no. 1, p. 14, Dec. 2020, doi: 10.1186/s41239-020-00193-3.
- [2] C. Romero, R. Cerezo, A. Bogarín, and M. Sánchez-Santillán, "EDUCATIONAL PROCESS MINING: A TUTORIAL AND CASE STUDY USING MOODLE DATA SETS," in *Data Mining and Learning Analytics*, 1st ed., S. ElAtia, D. Ipperciel, and O. R. Zaiane, Eds., Wiley, 2016, pp. 1–28. doi: 10.1002/9781118998205.ch1.
- [3] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432–1462, Mar. 2014, doi: 10.1016/j.eswa.2013.08.042.
- [4] S. D. Morales Dussan, M. Leon, O. Garcia-Bedoya, and I. Galpin, "Exploring the Colombian digital divide using Moodle logs through supervised learning," *ITSE*, vol. 19, no. 3, pp. 281–299, Aug. 2022, doi: 10.1108/ITSE-03-2021-0052.
- [5] T. Tk. Nguyen, M. T. Nguyen, and H. T. Tran, "Artificial intelligent based teaching and learning approaches: A comprehensive review," *IJERE*, vol. 12, no. 4, p. 2387, Dec. 2023, doi: 10.11591/ijere.v12i4.26623.
- [6] M. M. Iqbal, Y. Saleem, K. Naseer, and M. Kim, "Multimedia based student-teacher smart interaction framework using multi-agents in eLearning," *Multimed Tools Appl*, vol. 77, no. 4, pp. 5003–5026, Feb. 2018, doi: 10.1007/s11042-017-4615-z.
- [7] R. Baker and G. Siemens, "Educational Data Mining and Learning Analytics," in *The Cambridge Handbook of the Learning Sciences*, 2nd ed., R. K. Sawyer, Ed., Cambridge University Press, 2014, pp. 253–272. doi: 10.1017/CBO9781139519526.016.
- [8] M. Rosala, "Task Analysis: Support Users in Achieving Their Goals," *Nielsen Norman Group*. [Online]. Available: <https://www.nngroup.com/articles/task-analysis/>
- [9] I. Triguero, D. Molina, J. Poyatos, J. Del Ser, and F. Herrera, "General Purpose Artificial Intelligence Systems (GPAIS): Properties, definition, taxonomy, societal implications and responsible governance," *Information Fusion*, vol. 103, p. 102135, Mar. 2024, doi: 10.1016/j.inffus.2023.102135.
- [10] F. García-Peñalvo and A. Vázquez-Ingelmo, "What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI," *IJIMAI*, vol. 8, no. 4, p. 7, 2023, doi: 10.9781/ijimai.2023.07.006.
- [11] A. M. Al-Zahrani, "The impact of generative AI tools on researchers and research: Implications for academia in higher education," *Innovations in Education and Teaching International*, vol. 61, no. 5, pp. 1029–1043, Sep. 2024, doi: 10.1080/14703297.2023.2271445.
- [12] G. Siemens and D. Gasevic, "Guest editorial-learning and knowledge analytics," *Journal of Educational Technology & Society*, vol. 15, no. 3, p. 2, 2012.
- [13] L. Márquez, V. Henríquez, H. Chevreux, E. Scheihing, and J. Guerra, "Adoption of learning analytics in higher education institutions: A systematic literature review," *Brit J Educational Tech*, p. bjet.13385, Sep. 2023, doi: 10.1111/bjet.13385.
- [14] E. Freitas, F. Fonseca, V. Garcia, R. Ferreira, and D. Gasevic, "Towards a Maturity Model for Learning Analytics Adoption An Overview of its Levels and Areas," in *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*, Tartu, Estonia: IEEE, Jul. 2020, pp. 122–126. doi: 10.1109/ICALT49669.2020.00059.
- [15] K. Fahd and S. J. Miah, "Designing and evaluating a big data analytics approach for predicting students' success factors," *J Big Data*, vol. 10, no. 1, p. 159, Oct. 2023, doi: 10.1186/s40537-023-00835-z.
- [16] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [17] R. H. Bemthuis, R. R. Govers, and A. Asadi, "A CRISP-DM-based methodology for assessing agent-based simulation models using process mining," *Journal of Simulation*, pp. 1–22, Jun. 2025, doi: 10.1080/17477778.2025.2508245.
- [18] L.-I. González-Pérez, J. A. Enciso-González, C. M. Vicario-Solorzano, and M.-S. Ramírez-Montoya, "Measuring Digital Transformation in Education 4.0 with DT-Smarty: Valid and Reliable Model," *Technology, Knowledge and Learning*, 2025, doi: 10.1007/s10758-025-09844-8.
- [19] P. K. Kuhl, S.-S. Lim, S. Guerriero, and D. Van Damme, *Developing Minds in the Digital Age: Towards a Science of Learning for 21st Century Education*. in Educational Research and Innovation. OECD, 2019. doi: 10.1787/562a8659-en.
- [20] C. Cachero, J. R. Rico-Juan, and H. Macià, "Influence of personality and modality on peer assessment evaluation perceptions using Machine Learning techniques," *Expert Systems with Applications*, vol. 213, p. 119150, Mar. 2023, doi: 10.1016/j.eswa.2022.119150.
- [21] X. Zhai, K. C. Haudek, and W. Ma, "Assessing Argumentation Using Machine Learning and Cognitive Diagnostic Modeling," *Res Sci Educ*, vol. 53, no. 2, pp. 405–424, Apr. 2023, doi: 10.1007/s11165-022-10062-w.
- [22] R. R. Hoffman, B. Crandall, and N. Shadbolt, "Use of the Critical Decision Method to Elicit Expert Knowledge: A Case Study in the Methodology of Cognitive Task Analysis," *Hum Factors*, vol. 40, no. 2, pp. 254–276, Jun. 1998, doi: 10.1518/001872098779480442.
- [23] C. Guzmán-Valenzuela, C. Gómez-González, A. Rojas-Murphy Tagle, and A. Lorca-Vyhmeister, "Learning analytics in higher education: a preponderance of analytics but very little learning?," *Int J Educ Technol High Educ*, vol. 18, no. 1, p. 23, Dec. 2021, doi: 10.1186/s41239-021-00258-x.
- [24] L. I. González-Pérez and M. S. Ramírez-Montoya, "Components of Education 4.0 in 21st Century Skills Frameworks: Systematic Review," *Sustainability*, vol. 14, no. 3, p. 1493, Jan. 2022, doi: 10.3390/su14031493.
- [25] J. Sanabria-Z, I. M. Castillo-Martínez, L. I. González-Pérez, and M. S. Ramírez-Montoya, "Complex thinking through a Transition Design-guided Ideathon: testing an AI platform on the topic of sharing economy," *Front. Educ.*, vol. 8, p. 1186731, May 2023, doi: 10.3389/educ.2023.1186731.
- [26] I. M. Castillo-Martínez and M. S. Ramírez-Montoya, "Instrumento eComplexity: Medición de la percepción de

- estudiantes de educación superior acerca de su competencia de razonamiento para la complejidad.” Tecnológico de Monterrey. [Online]. Available: <https://hdl.handle.net/11285/643622>
- [27] M. S. Ramírez-Montoya, I. M. Castillo-Martínez, J. Sanabria-Z, and J. Miranda, “Complex Thinking in the Framework of Education 4.0 and Open Innovation—A Systematic Literature Review,” *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 8, no. 1, p. 4, Mar. 2022, doi: 10.3390/joitmc8010004.
- [28] G. Ibarra-Vazquez, M. S. Ramí-rez-Montoya, and H. Terashima, “Gender prediction based on University students’ complex thinking competency: An analysis from machine learning approaches,” *Educ Inf Technol*, vol. 29, no. 3, pp. 2721–2739, Feb. 2024, doi: 10.1007/s10639-023-11831-4.
- [29] V. E. Owen and R. S. Baker, “Fueling Prediction of Player Decisions: Foundations of Feature Engineering for Optimized Behavior Modeling in Serious Games,” *Tech Know Learn*, vol. 25, no. 2, pp. 225–250, Jun. 2020, doi: 10.1007/s10758-018-9393-9.
- [30] T. K. F. Chiu, Q. Xia, X. Zhou, C. S. Chai, and M. Cheng, “Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education,” *Computers and Education: Artificial Intelligence*, vol. 4, p. 100118, 2023, doi: 10.1016/j.caeai.2022.100118.
- [31] A. H. Hoti, X. Zenuni, M. Hamiti, and J. Ajdari, “Student Performance Prediction Using AI and ML: State of the Art,” in *2023 12th Mediterranean Conference on Embedded Computing (MECO)*, Budva, Montenegro: IEEE, Jun. 2023, pp. 1–6. doi: 10.1109/MECO58584.2023.10154933.
- [32] E. Koh and S. Doroudi, “Learning, teaching, and assessment with generative artificial intelligence: towards a plateau of productivity,” *Learning: Research and Practice*, vol. 9, no. 2, pp. 109–116, Jul. 2023, doi: 10.1080/23735082.2023.2264086.
- [33] I. Zliobaite *et al.*, “Next challenges for adaptive learning systems,” *SIGKDD Explor. Newsl.*, vol. 14, no. 1, pp. 48–55, Dec. 2012, doi: 10.1145/2408736.2408746.
- [34] R. Castro, “Blended learning in higher education: Trends and capabilities,” *Educ Inf Technol*, vol. 24, no. 4, pp. 2523–2546, Jul. 2019, doi: 10.1007/s10639-019-09886-3.
- [35] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, “Systematic review of research on artificial intelligence applications in higher education – where are the educators?,” *Int J Educ Technol High Educ*, vol. 16, no. 1, p. 39, Dec. 2019, doi: 10.1186/s41239-019-0171-0.
- [36] F. J. García-Peñalvo, “Inteligencia artificial generativa y educación: Un análisis desde múltiples perspectivas,” *Educ. Knowl. Soc.*, vol. 25, p. e31942, Apr. 2024, doi: 10.14201/eks.31942.
- [37] M. Grimaldi, O. Troisi, A. Papa, and E. De Nuccio, “Conceptualizing data-driven entrepreneurship: from knowledge creation to entrepreneurial opportunities and innovation,” *J Technol Transf*, Jan. 2025, doi: 10.1007/s10961-024-10176-5.
- [38] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, “Big data, analytics and the path from insights to value,” *MIT sloan management review.*, 2010. [Online]. Available: <https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to->
- [39] A. Behl, B. Sampat, V. Pereira, N. S. Jayawardena, and B. Laker, “Investigating the role of data-driven innovation and information quality on the adoption of blockchain technology on crowdfunding platforms,” *Ann Oper Res*, vol. 333, no. 2–3, pp. 1103–1132, Feb. 2024, doi: 10.1007/s10479-023-05290-w.
- [40] M. Tang, J. Cao, D. Zhang, and I. Pandelica, “A Data-Driven Assessment Model for Metaverse Maturity,” *INT J COMPUT COMMUN, Int. J. Comput. Commun. Control*, vol. 19, no. 4, Jul. 2024, doi: 10.15837/ijccc.2024.4.6498.
- [41] T. H. Davenport and J. Kim, *Keeping Up with the Quants: your guide to understanding and using analytics*. Boston, Mass: Harvard business review press, 2013.
- [42] C. Cote, “Types of Data Analytics to Improve Decision-Making,” *Business Insights-Blog.*, 2021. [Online]. Available: <https://online.hbs.edu/blog/post/types-of-data-analysis>
- [43] R. S. J. d. Baker and K. Yacef, “The State of Educational Data Mining in 2009: A Review and Future Visions,” Oct. 2009, doi: 10.5281/ZENODO.3554657.
- [44] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, “Comparing different supervised machine learning algorithms for disease prediction,” *BMC Med Inform Decis Mak*, vol. 19, no. 1, p. 281, Dec. 2019, doi: 10.1186/s12911-019-1004-8.
- [45] D. Gillborn, P. Warmington, and S. Demack, “QuantCrit: education, policy, ‘Big Data’ and principles for a critical race theory of statistics,” *Race Ethnicity and Education*, vol. 21, no. 2, pp. 158–179, Mar. 2018, doi: 10.1080/13613324.2017.1377417.
- [46] J. Niu and D. Reeping, “Humanizing Data-Driven Methods in Engineering Education Research: A Systematic Literature Review of Four Journals From 2011 to 2021,” *Studies in Engineering Education*, vol. 5, no. 2, pp. 150–174, Oct. 2024, doi: 10.21061/sec.159.
- [47] S. A. D. Popenici and S. Kerr, “Exploring the impact of artificial intelligence on teaching and learning in higher education,” *RPTTEL*, vol. 12, no. 1, p. 22, Dec. 2017, doi: 10.1186/s41039-017-0062-8.
- [48] SAP, “Qué es Inteligencia Artificial?,” *SAP*, 2023. [Online]. Available: <https://www.sap.com/latinamerica/products/artificial-intelligence/what-is-artificial-intelligence.html>
- [49] C. Korkmaz and A.-P. Correia, “A review of research on machine learning in educational technology,” *Educational Media International*, vol. 56, no. 3, pp. 250–267, Jul. 2019, doi: 10.1080/09523987.2019.1669875.
- [50] H. A. Younis, N. I. R. Ruhaiyem, W. Ghaban, N. A. Gazem, and M. Nasser, “A Systematic Literature Review on the Applications of Robots and Natural Language Processing in Education,” *Electronics*, vol. 12, no. 13, p. 2864, Jun. 2023, doi: 10.3390/electronics12132864.
- [51] G. Bebis, D. Egbert, and Mubarak Shah, “Review of computer vision education,” *IEEE Trans. Educ.*, vol. 46, no. 1, pp. 2–21, Feb. 2003, doi: 10.1109/TE.2002.808280.
- [52] M. Á. Conde, F. J. Rodríguez-Sedano, C. Fernández-Llamas, J. Gonçalves, J. Lima, and F. J. García-Peñalvo, “Fostering STEAM through challenge-based learning, robotics, and physical devices: A systematic mapping literature review,” *Comp Applic In Engineering*, vol. 29, no. 1, pp. 46–65, Jan. 2021, doi: 10.1002/cae.22354.
- [53] T. W. Edgar and D. O. Manz, *Research methods for cyber security*. Cambridge, MA: Syngress, an imprint of Elsevier, 2017.
- [54] B. Mahesh, “Machine Learning Algorithms - A Review,” *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381–386., 2020.
- [55] Geeksforgeeks, “Machine Learning Tutorial.” [Online]. Available: <https://www.geeksforgeeks.org/machine-learning/?ref=shm>
- [56] H. Geffner, “Artificial Intelligence: From programs to solvers,” *AI Communications*, vol. 27, no. 1, pp. 45–51, 2014, doi: 10.3233/AIC-130581.

- [57] S. Z. Salas-Pilco and Y. Yang, "Artificial intelligence applications in Latin American higher education: a systematic review," *Int J Educ Technol High Educ*, vol. 19, no. 1, p. 21, Dec. 2022, doi: 10.1186/s41239-022-00326-w.
- [58] K. Zhang and A. B. Aslan, "AI technologies for education: Recent research & future directions," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100025, 2021, doi: 10.1016/j.caeai.2021.100025.
- [59] V. N. Gudivada, V. V. Raghavan, V. Govindaraju, and C. R. Rao, *Cognitive computing: theory and applications*. in Handbook of statistics, no. volume 35. Amsterdam: Elsevier, 2016.
- [60] Y. Cui, F. Chen, A. Shiri, and Y. Fan, "Predictive analytic models of student success in higher education: A review of methodology," *ILS*, vol. 120, no. 3/4, pp. 208–227, Mar. 2019, doi: 10.1108/ILS-10-2018-0104.
- [61] K. Ahmad *et al.*, "Data-Driven Artificial Intelligence in Education: A Comprehensive Review," *IEEE Trans. Learning Technol.*, vol. 17, pp. 12–31, 2024, doi: 10.1109/TLT.2023.3314610.
- [62] T. Soysal, "Can the Right to Explanation in GDPR Be a Remedy for Algorithmic Discrimination?," in *Algorithmic Discrimination and Ethical Perspective of Artificial Intelligence*, M. Kılıç and S. Bozkuş Kahyaoğlu, Eds., in Accounting, Finance, Sustainability, Governance & Fraud: Theory and Application. , Singapore: Springer Nature Singapore, 2024, pp. 69–89. doi: 10.1007/978-981-99-6327-0\_5.
- [63] S. Cheng, K. Xie, and J. Collier, "Motivational beliefs moderate the relation between academic delay and academic achievement in online learning environments," *Computers & Education*, vol. 195, p. 104724, Apr. 2023, doi: 10.1016/j.compedu.2023.104724.
- [64] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," Technical Report EBSE-2007-01, 2007. [Online]. Available: <https://goo.gl/L1VHcw>
- [65] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, no. 4, pp. 571–583, Apr. 2007, doi: 10.1016/j.jss.2006.07.009.
- [66] F. J. García-Peñalvo, "Desarrollo de estados de la cuestión robustos: Revisiones Sistemáticas de Literatura," *Educ. Knowl. Soc.*, vol. 23, p. e28600, Apr. 2022, doi: 10.14201/eks.28600.
- [67] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.
- [68] M. J. Page *et al.*, "PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews," *BMJ*, p. n160, Mar. 2021, doi: 10.1136/bmj.n160.
- [69] J. Traxler, "Ethics and ICTD research," in *Linking research to practice in Asia: ICT for development research, management and perspectives*, Ottawa: International Development Research Centre., 2012, pp. 68–81.
- [70] L. Zhang, X. Wang, T. He, and Z. Han, "A Data-Driven Optimized Mechanism for Improving Online Collaborative Learning: Taking Cognitive Load into Account," *IJERPH*, vol. 19, no. 12, p. 6984, Jun. 2022, doi: 10.3390/ijerph19126984.
- [71] M. Cannistrà, C. Masci, F. Ieva, T. Agasisti, and A. M. Paganoni, "Early-predicting dropout of university students: an application of innovative multilevel machine learning and statistical techniques," *Studies in Higher Education*, vol. 47, no. 9, pp. 1935–1956, Sep. 2022, doi: 10.1080/03075079.2021.2018415.
- [72] K. Sharma, S. Lee-Cultura, and M. Giannakos, "Keep Calm and Do Not Carry-Forward: Toward Sensor-Data Driven AI Agent to Enhance Human Learning," *Front. Artif. Intell.*, vol. 4, p. 713176, Jan. 2022, doi: 10.3389/frai.2021.713176.
- [73] M. Phan, A. De Caigny, and K. Coussement, "A decision support framework to incorporate textual data for early student dropout prediction in higher education," *Decision Support Systems*, vol. 168, p. 113940, May 2023, doi: 10.1016/j.dss.2023.113940.
- [74] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, p. 11, Dec. 2022, doi: 10.1186/s40561-022-00192-z.
- [75] T. Susnjak, G. S. Ramaswami, and A. Mathrani, "Learning analytics dashboard: a tool for providing actionable insights to learners," *Int J Educ Technol High Educ*, vol. 19, no. 1, p. 12, Dec. 2022, doi: 10.1186/s41239-021-00313-7.
- [76] J. A. I. S. Masood, N. S. Kalyan Chakravarthy, D. Asirvatham, M. Marjani, D. Abdulkareem Shafiq, and S. Nidamanuri, "A Hybrid Deep Learning Model to Predict High-Risk Students in Virtual Learning Environments," *IEEE Access*, vol. 12, pp. 103687–103703, 2024, doi: 10.1109/ACCESS.2024.3434644.
- [77] D. H. Chang, M. P.-C. Lin, S. Hajian, and Q. Q. Wang, "Educational Design Principles of Using AI Chatbot That Supports Self-Regulated Learning in Education: Goal Setting, Feedback, and Personalization," *Sustainability*, vol. 15, no. 17, p. 12921, Aug. 2023, doi: 10.3390/su151712921.
- [78] Z. Mehenaoui, Y. Lafifi, and L. Zemouri, "Learning Behavior Analysis to Identify Learner's Learning Style based on Machine Learning Techniques," *Jucs*, vol. 28, no. 11, pp. 1193–1220, Nov. 2022, doi: 10.3897/jucs.81518.
- [79] C. Blundo, G. Fenza, G. Fuccio, V. Loia, and F. Orciuoli, "A time-driven FCA-based approach for identifying students' dropout in MOOCs," *Int J of Intelligent Sys*, vol. 37, no. 4, pp. 2683–2705, Apr. 2022, doi: 10.1002/int.22414.
- [80] N. Mezhoudi, R. Alghamdi, R. Aljunaid, G. Krichna, and D. Düşteğör, "Employability prediction: a survey of current approaches, research challenges and applications," *J Ambient Intell Human Comput*, vol. 14, no. 3, pp. 1489–1505, Mar. 2023, doi: 10.1007/s12652-021-03276-9.
- [81] M. V. Martins, L. Baptista, J. Machado, and V. Realinho, "Multi-Class Phased Prediction of Academic Performance and Dropout in Higher Education," *Applied Sciences*, vol. 13, no. 8, p. 4702, Apr. 2023, doi: 10.3390/app13084702.
- [82] L. Horn, M. Karsai, and G. Markova, "An automated, data-driven approach to children's social dynamics in space and time," *Child Dev Perspectives*, vol. 18, no. 1, pp. 36–43, Mar. 2024, doi: 10.1111/cdep.12495.
- [83] P. Xuan Lam, P. Q. H. Mai, Q. H. Nguyen, T. Pham, T. H. H. Nguyen, and T. H. Nguyen, "Enhancing educational evaluation through predictive student assessment modeling," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100244, Jun. 2024, doi: 10.1016/j.caeai.2024.100244.
- [84] Md. F. I. Amin, A. Shirafuji, Md. M. Rahman, and Y. Watanobe, "Multi-Label Code Error Classification Using CodeT5 and ML-KNN," *IEEE Access*, vol. 12, pp. 100805–100820, 2024, doi: 10.1109/ACCESS.2024.3430558.
- [85] H. Lahza, H. Khosravi, and G. Demartini, "Analytics of learning tactics and strategies in an online learnersourcing environment," *Computer Assisted Learning*, vol. 39, no. 1, pp. 94–112, Feb. 2023, doi: 10.1111/jcal.12729.
- [86] M. Afzaal *et al.*, "Explainable AI for Data-Driven Feedback and Intelligent Action Recommendations to Support Students Self-Regulation," *Front. Artif. Intell.*, vol. 4, p. 723447, Nov. 2021, doi: 10.3389/frai.2021.723447.
- [87] X. Liu, "Online Learning Based on Learning Analytics in Big

- Data for College English Language Teaching,” *Int J Comput Intell Syst*, vol. 17, no. 1, p. 192, Jul. 2024, doi: 10.1007/s44196-024-00594-6.
- [88] M.-R. Borna, H. Saadat, A. T. Hojjati, and E. Akbari, “Analyzing click data with AI: implications for student performance prediction and learning assessment,” *Front. Educ.*, vol. 9, p. 1421479, Dec. 2024, doi: 10.3389/educ.2024.1421479.
- [89] K. Akhuseyinoglu and P. Brusilovsky, “Exploring Behavioral Patterns for Data-Driven Modeling of Learners’ Individual Differences,” *Front. Artif. Intell.*, vol. 5, p. 807320, Feb. 2022, doi: 10.3389/frai.2022.807320.
- [90] Y. S. Balcioglu and M. Artar, “Predicting academic performance of students with machine learning,” *Information Development*, p. 02666669231213023, Nov. 2023, doi: 10.1177/02666669231213023.
- [91] M. Hooda, C. Rana, O. Dahiya, A. Rizwan, and M. S. Hossain, “Artificial Intelligence for Assessment and Feedback to Enhance Student Success in Higher Education,” *Mathematical Problems in Engineering*, vol. 2022, pp. 1–19, May 2022, doi: 10.1155/2022/5215722.
- [92] J. Roestenburg, C. J. Kruger, M. Nel, and Z. Janse Van Rensburg, “Exploring factors influencing academic literacy – A data-driven perspective,” *S. Afr. j. inf. manag.*, vol. 26, no. 1, Mar. 2024, doi: 10.4102/sajim.v26i1.1729.
- [93] S. Peng and K. Nagao, “Recognition of Students’ Mental States in Discussion Based on Multimodal Data and its Application to Educational Support,” *IEEE Access*, vol. 9, pp. 18235–18250, 2021, doi: 10.1109/ACCESS.2021.3054176.
- [94] S. Leelaluk *et al.*, “Attention-Based Artificial Neural Network for Student Performance Prediction Based on Learning Activities,” *IEEE Access*, vol. 12, pp. 100659–100675, 2024, doi: 10.1109/ACCESS.2024.3429554.
- [95] S. Gaftandzhieva *et al.*, “Exploring Online Activities to Predict the Final Grade of Student,” *Mathematics*, vol. 10, no. 20, p. 3758, Oct. 2022, doi: 10.3390/math10203758.
- [96] E. J. Delahoz-Domínguez and R. Hijón-Neira, “Recommender System for University Degree Selection: A Socioeconomic and Standardised Test Data Approach,” *Applied Sciences*, vol. 14, no. 18, p. 8311, Sep. 2024, doi: 10.3390/app14188311.
- [97] M. W. H. Spitzer, M. Ruiz-Garcia, and K. Moeller, “Basic mathematical skills and fraction understanding predict percentage understanding: Evidence from an intelligent tutoring system,” *Brit J Educational Tech*, p. bjet.13517, Aug. 2024, doi: 10.1111/bjet.13517.
- [98] A. Mallavarapu, L. Lyons, and S. Uzzo, “Exploring the Utility of Social-Network-Derived Collaborative Opportunity Temperature Readings for Informing Design and Research of Large-Group Immersive Learning Environments,” *Learning Analytics*, vol. 9, no. 1, pp. 53–76, Mar. 2022, doi: 10.18608/jla.2022.7419.
- [99] G. Ramaswami, T. Susnjak, and A. Mathrani, “Supporting Students’ Academic Performance Using Explainable Machine Learning with Automated Prescriptive Analytics,” *BDC*, vol. 6, no. 4, p. 105, Sep. 2022, doi: 10.3390/bdcc6040105.
- [100] T. Susnjak, “Beyond Predictive Learning Analytics Modelling and onto Explainable Artificial Intelligence with Prescriptive Analytics and ChatGPT,” *Int J Artif Intell Educ*, vol. 34, no. 2, pp. 452–482, Jun. 2024, doi: 10.1007/s40593-023-00336-3.
- [101] Y. Lee, “Identifying Prerequisite Courses in Undergraduate Biology Using Machine Learning,” *Journal of Data Science*, pp. 745–760, 2023, doi: 10.6339/22-JDS1068.
- [102] I. Šarić-Grgić, A. Grubišić, and B. Žitko, “The Impact of Note-Taking on the Learning Process in Intelligent Tutoring System Tutomat,” *Applied Computer Systems*, vol. 26, no. 1, pp. 31–37, May 2021, doi: 10.2478/acss-2021-0004.
- [103] J. A. Ruipérez-Valiente, D. Jaramillo-Morillo, S. Joksimović, V. Kovanović, P. J. Muñoz-Merino, and D. Gašević, “Data-driven detection and characterization of communities of accounts collaborating in MOOCs,” *Future Generation Computer Systems*, vol. 125, pp. 590–603, Dec. 2021, doi: 10.1016/j.future.2021.07.003.
- [104] M. Afzaal, A. Zia, J. Nouri, and U. Fors, “Informative Feedback and Explainable AI-Based Recommendations to Support Students’ Self-regulation,” *Tech Know Learn*, vol. 29, no. 1, pp. 331–354, Mar. 2024, doi: 10.1007/s10758-023-09650-0.
- [105] H. Tjahyadi and K. Tude, “The Implementation of Educational Data Mining in Predicting Students’ Academic Achievement in Mathematics at a Private Elementary School,” *IJJET*, vol. 15, no. 1, pp. 154–163, 2025, doi: 10.18178/ijet.2025.15.1.2228.

**Laura Icela González Pérez** obtuvo el Doctorado en Educación en la Sociedad del Conocimiento por la Universidad de Salamanca. Asimismo, cuenta con una Maestría en Tecnología Educativa por el Tecnológico de Monterrey y una Maestría en Neuromarketing por la Universidad Internacional de La Rioja. Es Profesora e Investigadora en la Facultad de Psicología de la Universidad Autónoma de Nuevo León (UANL) y miembro Nivel I del Sistema Nacional de Investigadores de México. Su investigación se centra en la transformación digital, la educación abierta y las tecnologías emergentes. Está afiliada a JWEL y RedLaTE-Mx, y ha recibido reconocimiento internacional en ciencia abierta, inteligencia artificial e innovación educativa.

**Francisco José García-Peñalvo** es Catedrático en el Departamento de Informática y Automática de la Universidad de Salamanca (USAL). Dirige el grupo de investigación GRIAL y se desempeña como Subdirector del Instituto de Ciencias de la Educación. Ha ocupado cargos de liderazgo como Vicerrector y Vicedecano en la USAL, y fue Profesor Distinguido en el Tecnológico de Monterrey (2016–2018). Su trabajo se centra en el aprendizaje mediado por tecnología y la innovación educativa. Incluido en la lista *World’s Top 2% Scientists* de Stanford (2019–2024), ha publicado más de 150 artículos indexados y es Editor en Jefe de *Education in the Knowledge Society* y Editor Asociado de varias revistas internacionales.

**Amadeo José Argüelles Cruz** obtuvo el Doctorado en Ciencias de la Computación en el Instituto Politécnico Nacional (IPN), México. Actualmente es Profesor Investigador Titular en el Centro de Investigación en Computación del IPN, donde también ha fungido como Jefe de Departamento y Coordinador Académico. Su investigación abarca inteligencia artificial, reconocimiento de patrones y cómputo en la nube. Es miembro Nivel II del Sistema Nacional de Investigadores de México y está afiliado a IEEE, ACM, AAAI y otras sociedades. Su trabajo reciente explora sistemas inteligentes para educación, movilidad y sostenibilidad urbana. A nivel internacional, ha participado en iniciativas conjuntas con el *Senseable City Lab* y el *City Science Lab* del MIT, así como en proyectos como Ecoemprende con el Tecnológico de Monterrey y SEC.