Data-Driven Learning Analytics and Artificial Intelligence in Higher Education: A Systematic Review

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

***Abstract*—The responsible integration of Artificial Intelligence in Education (AIED) offers a strategic opportunity to align learning environments with the principles of Society 5.0, fostering human–technology synergy in support of quality education and social well-being. This study presents a systematic review of 36 peer-reviewed articles (2021–2025) focused on educational applications that employ learning analytics (LA) through data-driven approaches and integrate machine learning (ML) models as part of their empirical evidence. Each study was analyzed according to three key dimensions: the context of AIED application, the data-driven approach adopted, and the ML model implemented. The findings reveal a persistent disconnect between the AI models employed and the available educational data, which in many cases are limited to access logs or manually recorded grades that fail to capture deeper cognitive processes. This limitation constrains both the effective training of ML models and their pedagogical utility for delivering meaningful interventions such as personalized learning pathways, real-time feedback, early detection of learning difficulties, and monitoring and visualization tools. Another significant finding is the absence of psychopedagogical frameworks integrated with quality standards and data governance, which are essential for advancing prescriptive and ethical approaches aligned with learning goals. It is therefore recommended that educational leaders foster AIED applications grounded in data governance and ethics frameworks, ensuring valid and reliable metrics that can drive a more equitable and inclusive education.**

***Index Terms*—Learning analytics, machine learning, data-driven approach, educational applications, intelligent systems, higher education, artificial intelligence, Education 4.0, educational innovation, Society 5.0.**

I. Introduction

Although interest in data-driven approaches to education is increasing, there remains a lack of psychopedagogical and ethical frameworks to guide the integration of AI and to consolidate a data architecture capable of enhancing metacognitive development and academic performance. Renz and Hilbig [1] highlight the complexity of integrating Learning Analytics (LA) and Artificial Intelligence (AI) to design adaptive teaching and learning pathways. Educational Data Mining (EDM) contributes to the development of techniques that identify student behavior patterns from various types of data: structured, semi-structured, and unstructured [2], [3]. LA and EDM have primarily focused on the measurement and analysis of student data, drawing on records such as time on task and access to materials within Learning Management Systems (LMS), with the aim of testing hypotheses related to study habits, motivation, and engagement in order to identify learning behavior patterns and optimize learning processes [4], [5], [6], and [7]. Designing a data-driven educational framework requires clear objectives, high-quality and traceable data from virtual environments, and governance structures that enable AI to effectively support learning.

Cognitive process modeling [8] supported by AI and Machine Learning (ML), enables personalized learning [9], facilitates the integration of heterogeneous data sources, and supports the development of scalable and adaptive learning ecosystems [10]. In parallel, generative AI tools are increasingly being adopted in academic contexts to provide support, guidance, and even personalized instruction [11]. To integrate AI into the learning systems of Higher Education Institutions (HEIs), it is essential to promote literacy in data governance, ethics, and deontological principles to ensure that ML models are trained on valid and reliable data, minimizing bias risks and grounded in rigorous algorithmic performance metrics. Building on these considerations, the objective of this review is to address the following research questions: (RQ1) How has the use of conceptual, empirical, and/or mixed methods evolved in studies published between 2021 and 2025 that adopt data-driven approaches and integrate AI models with ML? (RQ2) What data-driven approaches are employed in studies that also use ML models? (RQ3) What AIED applications have been implemented in HEIs? (RQ4) What types of AI models with ML have been developed?.

Received 30 April 2025; revised 7 July 2025; accepted 3 August 2025. Date of publication 29 September 2025; date of current version 18 December 2025. This work was supported by the Instituto Politécnico Nacional under Project SIP-20250146. *(Corresponding author: Amadeo José Argüelles-Cruz.)*

Laura Icela González-Pérez is with the Facultad de Psicología, Universidad Autónoma de Nuevo León, Nuevo León 64460, Mexico (e-mail: lauraicela25@hotmail.com).

Francisco José García-Peñalvo is with the Department of Computer Science and Automation, University of Salamanca, 37008 Salamanca, Spain.

Amadeo José Argüelles-Cruz is with the Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City 07738, Mexico (e-mail: aarguelles@ipn.mx).

A Spanish version of this article is available as supplementary material at https://doi.org/10.1109/RITA.2025.3615512.

Digital Object Identifier 10.1109/RITA.2025.3615512

II. Theoretical Framework

A. Learning Analytics (LA) and Educational Data Mining (EDM))

To achieve effective governance of data and AI, HEIs must align their institutional strategies with ethical governance standards while ensuring the identification and mitigation of risks associated with their use. Learning Analytics (LA) is defined as the measurement, collection, analysis, and reporting of student data to improve learning and learning environments [12], [13]. Digital maturity frameworks and data quality standards can serve as necessary guidelines to systematize and structurally enhance teaching and learning processes when integrated with AI models, thereby enabling more inclusive and less biased predictions of academic outcomes [14], [15]. The CRISP-DM model (Cross-Industry Standard Process for Data Mining) is a methodology validated across multiple sectors and applied in data mining and data science. It is conceived as an iterative cycle encompassing business and data understanding, preparation, modeling, evaluation, and the deployment of resulting architectures [16], [17]. Embedding a data governance framework at the core of a digital learning ecosystem allows for the articulation of technological platforms and fosters interoperability among diverse systems [18]. The OECD urges HEIs to validate, through scientific evidence, both the benefits and the limitations of the data employed, and to increasingly leverage big data infrastructures contextualized to educational needs [19]. For AI to provide bias-free functionalities, it is essential to rely not only on psychopedagogical models and instruments or scales, but also on algorithmic validation metrics for ML model training, thereby ensuring comprehensive measures that demonstrate the positive impact of their use.

B. AIED and Applications in Learning Processes

Identifying the roles and contexts of faculty, students, and academic leaders enables the design of competency frameworks in AI and data governance that foster sustainable and innovative educational models grounded in data-driven approaches. Instructional design incorporating emerging technological tools must integrate the measurement of cognitive processes in order to leverage AI for the creation of adaptive learning strategies with self-regulation functionalities and performance metrics [20]. Within this framework, Cognitive Diagnostic Modeling (CDM) provides a psychometric perspective to accurately identify the mastery of specific skills [21], while Cognitive Task Analysis (CTA) facilitates the representation of complex knowledge structures. When integrated into AI-powered platforms, these approaches can strengthen personalized learning and generate intelligent feedback, thereby enhancing the effectiveness of educational processes [22]. However, the absence of psychopedagogical frameworks solidly grounded in learning data limits the connection between observed interactions and expected outcomes; hence the need to advance toward integrative models that align theory, empirical evidence, and sustainable educational practices [23].

These strategies are essential for redesigning virtual learning environments aimed at developing and scaling competencies. By integrating the influence of socioeconomic, cultural, and environmental factors, it becomes possible to design learning pathways that help overcome existing gaps and strengthen equity in educational processes [24], [25], and foster the development of higher-order skills such as complex thinking, critical reasoning, creativity, problem solving, and leadership, among others [26], [27]. Training ML models with reliable educational data under rigorous ethical protocols makes it possible to more accurately understand patterns of student behavior, identify areas for improvement, and incorporate innovative practices that strengthen complex mental processes such as critical thinking, creativity, and problem solving [28].

TABLE I
CLASSIFICATION OF DATA- DRIVEN APPROACHES ACCORDING TO [41], [42]

Methods	Analytical Objective	Technique
Diagnostic Approach	Uses data to identify the underlying causes of trends and correlations. In this sense, diagnostic regression is oriented toward answering the question: *"Why did this happen?* "	Hypothesis testing, correlation vs. causation, regression analysis
Descriptive Approach	Provides information about the past, which is useful for communicating changes over time and using trends as a basis for further studies. Its central purpose is to answer the question: *"What has happened?"*	Descriptive statistics, data mining techniques.
Predictive Approach	Employs historical data and trend patterns to estimate future scenarios, focusing on the question: *"What could happen?"*	Data mining techniques, regression, ML.
Prescriptive Approach	Uses data to quantify the potential effects of future decisions and offers recommendations to answer the question: *"What should we do next?"*	Business rules, experimental design, computational modeling, ML/AI (including Natural Language Processing, NLP)

The pedagogical impact of technologies such as AI will only be sustainable if supported by institutional governance that legitimizes data use. The combination of LA or EDM with behavioral modeling refines the understanding of student profiles and behaviors [29], making it increasingly necessary to establish a governance framework that ensures the ethical and transparent use of data. Along these lines, it is essential to determine the type of interaction or educational application under analysis. Some authors [30] group AIED applications into four categories: (1) learning, (2) teaching, (3) assessment, and (4) administrative management. Other studies classify them into more specific functions, such as student profiling and performance prediction [31], assessment tools [32], adaptive and personalized systems [33], [34] intelligent tutoring systems [35], and automatic content generation [36].

C. Data-Driven Approaches

The data-driven approach is transforming the way organizations evaluate their processes and strengthen the mechanisms that enable them to generate and sustain competitive advantages. According to [37] implementing a data-driven model requires organizations to: (1) transform collected data into meaningful information and actionable knowledge; (2) develop a robust technological infrastructure along with an articulated digital ecosystem; (3) deploy data analytics and integration strategies to optimize resources; and (4) strengthen the capabilities of key stakeholders. Several studies emphasize that data-driven decision-making enables the integration and analysis of information to achieve more accurate results. This approach, grounded in the collection, analysis, and interpretation of large volumes of data [38], [39], has become a pillar of business competitiveness and innovation, gaining increasing relevance [40]. Similarly, HEIs are called to transform themselves through their capacity to convert data into actionable information that allows them to quantify, support, and manage learning, which in turn requires the development of new data architectures. As summarized in Table I, data-driven approaches can be classified into four main categories: (1) diagnostic, (2) descriptive, (3) predictive, and (4) prescriptive [41], [42].

A data-driven approach, grounded in structured information and the application of diverse analytical methods, enables a more objective examination of user behaviors and attitudes, particularly within learning management systems, and provides highly valuable evidence for decision-making and the design of pedagogical strategies. In this context, ML plays a prominent role by employing data-driven approaches to uncover, predict, and classify behavioral patterns with the purpose of guiding the continuous improvement of educational processes [43]. Supervised ML consists of using labeled data with predefined categories to train and validate predictive or classification models. Among the most frequently employed algorithms are decision trees, random forests, the Naïve Bayes classifier, and artificial neural networks [44]. However, these techniques raise ethical considerations, particularly regarding the need to validate performance metrics and verify both data architectures and labeling processes in order to ensure reliability and transparency.

Biases present in insufficiently validated instruments may be amplified in AI models, underscoring the importance of rigorous validation aligned with data governance. In this regard, QuantCrit principles critically examine prevailing practices in quantitative research, as highlighted in [45] by emphasizing: (1) the centrality of racism, (2) the lack of neutrality in numerical data, (3) the socially constructed nature of categories such as gender and race, (4) the interpretive character of data as a socially produced construct, and (5) the imperative of a commitment to social justice and equity. A robust understanding of educational phenomena, therefore, requires combining quantitative analytics with qualitative inquiry. Similarly, psychological constructs such as motivation, personality, or epistemic beliefs—often measured through person-centered methods—require strict tests of validity and reliability, since without such scrutiny they remain methodologically and ethically vulnerable [46]. In the case of ML models, including generative AI, it is essential to implement mechanisms for detecting errors, biases, and hallucinations, as these limitations may compromise the reliability and validity of metrics, as well as the safety and trustworthiness of their educational applications.

D. Artificial Intelligence Models

AI has the potential to transform educational processes, provided that solid data governance architectures are established to ensure system interoperability and support the adoption of

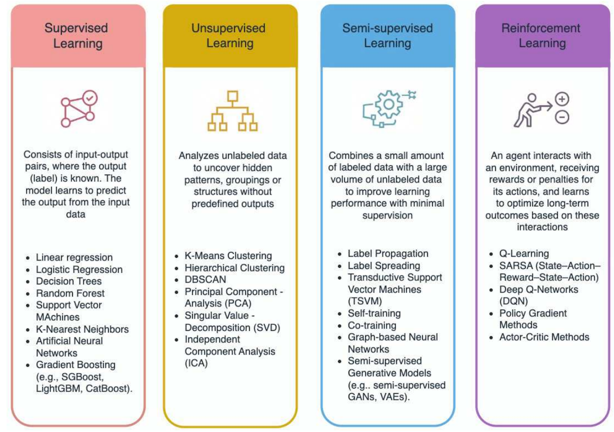

Fig. 1. ML models [54], [55].

evidence-based measures. AI can be defined as a computer system capable of adapting, learning, synthesizing information, correcting errors, and leveraging heterogeneous data to solve complex tasks [47]. Authors [48] identify four key AI technologies: (1) ML [49] (2) Natural Language Processing (NLP) [50], (3) Computer Vision [51], and (4) Robotics [52]. ML is a field of study that employs computational algorithms to transform empirical data into usable models [53], encompassing four main paradigms: (a) supervised, (b) unsupervised, (c) semi-supervised, and (d) reinforcement learning [54], [55] (Fig. 1).

The implementation of AI in HEIs may involve processing sensitive student data without fully established security mechanisms, thereby increasing the risk of inappropriate use of private information and the inadequate detection of indicators related to mental health, emotional well-being, and learning patterns. AI has evolved from programming specific and poorly structured tasks, lacking connection with interoperable approaches or models, toward methods grounded in formal frameworks. Among the latter are Constraint Satisfaction Problems (CSP), Bayesian networks, Partially Observable Markov Decision Processes (POMDP), and General Game Playing, all of which are addressed through domain-independent algorithms [56]. This paradigm supports significant advances in expert systems, intelligent agents, predictive models, and personalized learning environments [57], [58]. ML further enriches the analysis by revealing latent patterns and generating actionable knowledge [59], reinforcing the need for AI governance in education that ensures the comprehensive protection of the fundamental human rights of those involved.

Several authors highlight that the application of ML models in education has revealed three main challenges: (1) integrating heterogeneous data to identify students at risk, (2) ensuring rigorous evaluation of methodological validity, and (3) establishing reliable guidelines for model development and assessment [60]. Ahmad et al. [61] numerate key data-driven applications—performance prediction, personalized learning, and sentiment analysis—and emphasize the need to curb algorithmic bias and safeguard student privacy. The broader deployment of ML systems has also raised concerns about potential discriminatory outcomes [62]. Consequently, intelli-

TABLE II
GUIDING RESEARCH QUESTIONS

Research questions	Answer posibilities
RQ1. How has the use of empirical and/or mixed methods evolved in studies published between 2021 and 2025 that address data-driven approaches and ML in education?	Empirical, Conceptual, Mixed (2021, 2022, 2023, 2024, and 2025)
RQ2. What data-driven approaches are employed in studies that integrate AI models with ML?	Descriptive, Diagnostic, Predictive, Prescriptive

gent platforms must be grounded in evidence-based behavioral constructs [63] and designed to guarantee fairness, transparency, accountability, and responsible data management in alignment with instructional objectives.

III. METHODOLOGY

This systematic literature review (SLR) examines how conceptual, empirical, and mixed-method studies published between 2021 and 2025 employ LA and EDM in combination with data-driven approaches and ML within educational applications. The review catalogs AI models, identifies methodological trends, and evaluates the depth of ML and AI integration in educational processes. The three-stage protocol proposed by Kitchenham—planning, execution, and reporting—was followed to ensure methodological rigor [64]. In addition, the SLR guidelines proposed by Brereton et al. [65] were applied, which include predefined search protocols, pilot testing, clear inclusion and exclusion criteria, as well as transparent documentation of search strategies and study selection procedures.

A. Planning Phase

To support a rigorous analysis of the literature, identify research gaps, and formulate relevant questions, the scope of the review was defined. Accordingly, studies were selected that implemented AIED solutions—platforms, systems, interfaces, or related technologies—grounded in data-driven approaches employing ML models in educational, data-oriented environments. The research questions guiding the study are summarized in Table II.

The SLR protocol is explicitly based on the PICOC framework [66]:

1) PICOC Reference Framework:

- Population: students, faculty, and academic administrators in HEIs.
- Intervention: identification of studies applying LA or EDM methodologies based on data-driven approaches that incorporate ML to improve learning outcomes.
- Comparison: not applicable, as the review does not include comparative interventions.
- Outcome: analysis of how empirical studies integrate ML-based data analysis within AIED, highlighting methodological rigor, AI model implementation, and evidence-based results.
- Context: educational use of AIED and LA in HEI environments between 2021 and 2025.

2) Inclusion and Exclusion Criteria:

- Inclusión criteria:
- Published between January 2021 and February 27, 2025.
- Document type: Articles.
- Language: English.
- Studies applying ML models.
- Exclusión criteria:
- Articles for which full-text access was not available despite reasonable retrieval efforts.
- Studies published in languages other than English.
- Studies that do not apply ML models.

3) Search Strategy: The search process in the Scopus and Web of Science (WoS) databases began with the definition of the keywords to be used in their search engines. To select the articles that formed the database, inclusion and exclusion criteria were established according to the PICOC framework, along with the following keywords:

"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")"

The search was conducted on February 27, 2025. The 2021–2025 period was chosen to capture the recent surge in the adoption of data-driven learning technologies and AI models. This timeframe coincides with the post-pandemic context, which has accelerated the digital transformation of HEIs.

B. Review Phase

The systematic review followed the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), as illustrated in Fig. 2 [67], [68]. This widely recognized methodology encompasses four key stages: identification, selection, eligibility assessment based on exclusion criteria, and final inclusion.

Data were systematized using Excel and Tableau to generate cross-tabulations and visual representations. Initially, Excel was employed to merge datasets and remove duplicate entries. A new worksheet was then created to classify the studies, with each assigned a unique identification code based on the alphabetical order of the article title. Subsequently, the following columns were created to categorize the key aspects of each study:

- Type of research method: conceptual, empirical, or mixed.
- Type of data-driven approach: descriptive, diagnostic, predictive, or prescriptive.
- LA or EDM approach.
- ML model: supervised learning, unsupervised learning, semi-supervised learning, or reinforcement learning.
- Type of AI application in education: learning, teaching, assessment, administrative processes, or cross-cutting applications.
- Study description.
- Algorithms used.
- Observations.

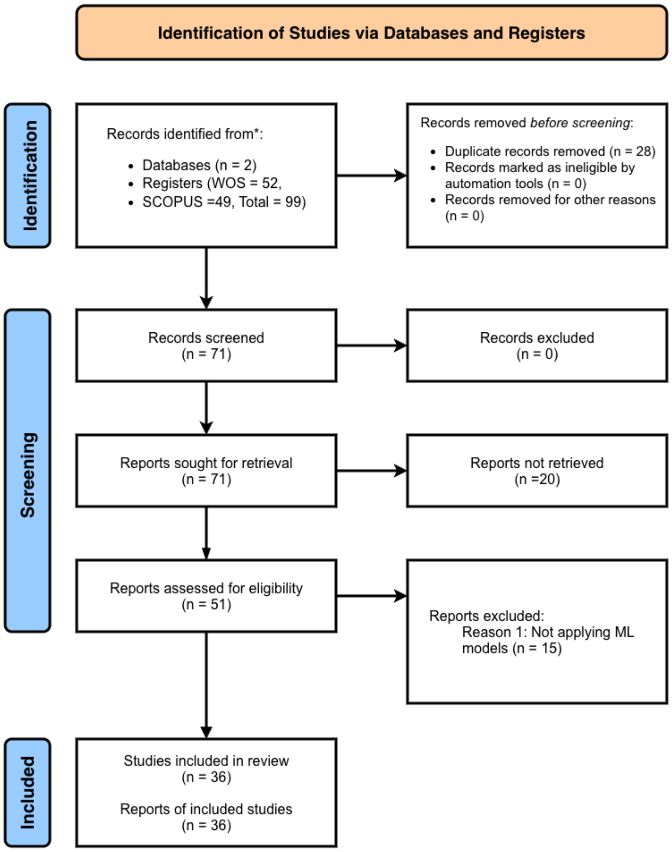

Fig. 2. PRISMA 2020 flow diagram.

C. Reporting Phase

Data processing adhered to ethical standards and was based on the evidence collected. Following Traxler's recommendations, platform observations were analyzed in relation to the research questions and the coding categories of the study [69]. Subsequently, 15 studies were excluded after it was determined through closer analysis that they did not employ ML models. The dataset is published on Zenodo at: https://zenodo.org/records/15802477

IV. Results

Table III provides an overview of the 36 peer-reviewed studies published between 2021 and 2025 that apply ML-driven data methods to AI-enabled education. Listed alphabetically by title and indexed with consecutive identifiers (Id), these works collectively document ongoing advances in the digital transformation of teaching and learning. The identifier corresponds to the article's identification number and will be used for this purpose in the following sections.

Geographic coverage was determined based on the affiliation of the corresponding author, revealing a strong presence of universities from Europe, Asia, and North America. Europe accounted for 14 studies, including Spain (S17, S44), Sweden (S25, S31), Turkey (S21, S41), Italy (S5, S20), Austria (S7), Croatia (S49), France (S3), Germany (S13), Norway (S33), and Portugal (S37). Asia contributed 12 studies, with Japan (S12, S38, S43), China (S2, S40), Indonesia (S9, S50),

TABLE III
List of Included Studies

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]

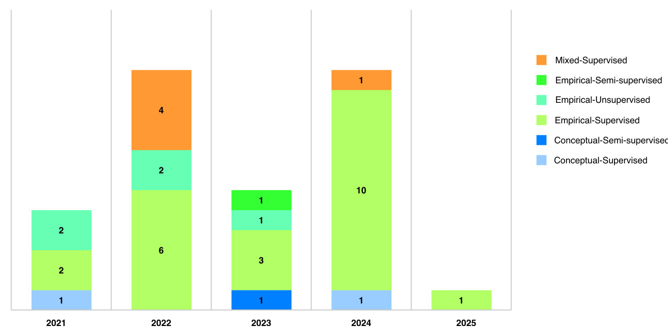

Fig. 3. Research methods and ML models (2021–2025).

Bangladesh (S11), India (S28), Malaysia (S4), Vietnam (S24), and Saudi Arabia (S23). The Americas included four studies: the United States (S26, S29, S30) and Canada (S22). Oceania was represented by four studies from New Zealand (S14, S34, S46) and Australia (S8). Finally, Africa contributed two studies, from Algeria (S35) and South Africa (S27). The absence of studies from Latin America highlights a critical gap in current research on the use of AI and ML models grounded in data-driven approaches. This reflects not only regional underrepresentation but also a scarcity of empirical documentation on their application in education. These asymmetries underscore the need to foster empirical efforts that can establish a research agenda on AIED in Latin America that is geographically, institutionally, and disciplinarily representative.

To address RQ1—*How has the use of conceptual, empirical, and/or mixed methods evolved in studies published between 2021 and 2025 related to data-driven approaches and ML in education*?—of the 36 studies analyzed, 8% (3 studies) employed conceptual methods, 78% (28 studies) were empirical, and 14% (5 studies) used mixed methods. Figure 3 illustrates this temporal evolution of both the research method and the ML model used in the analyzed studies.

To address RQ2—*What data-driven approaches are employed in studies that integrate AI models with ML?* —the

TABLE IV
CLASSIFICATION OF STUDIES ACCORDING TO DATA-DRIVEN APPROACH AND ML [41] AND [42]

Studies with Supervised ML Models (n=29)	
E–Diagnostic & Predictive (4)	**S5, S35, S38, S50**
C-Descriptive (1)	**S7**
E-Descriptive & Predictive (1)	**S27**
E-Predictive (13)	**S4, S9, S12, S13, S21, S24, S28, S30, S33, S37, S41, S43, S20**
C-Predictive (1)	**S23**
E-Predictive & Prescriptive (5)	**S25, S31, S34, S40, S44**
E- Prescriptive (4)	**S2, S11, S14, S46**
Studies with ML No supervised (n=5)	
E-Diagnostic & Descriptive (2)	**S8, S29**
E-Diagnostic & Predictive (1)	**S49**
E-Descriptive (1)	**S17**
E-Descriptive & Predictive (1)	**S26**
Studies with ML semisupervised (n=2)	
C-Prescriptive (1)	**S22**
E-Predictive (1)	**S3**

data-driven framework proposed in [41] and [42], was adopted, with details presented in Table I, along with the ML models based on [54] and [55] shown in Fig. 1. Table IV presents the studies organized according to these classifications. Each classification was coded by the research method employed: C (conceptual), E (empirical), and M (mixed).

In summary, this classification of data-driven approaches shows that the predictive approach is predominant, appearing in 19 of the 36 studies analyzed. It is followed by the diagnostic approach (7 studies), the descriptive approach (6 studies), and the prescriptive approach (6 studies). Notably, the diagnostic approach frequently appears in combination with other approaches, serving as a transversal component in the early stages of data-driven implementations.

A. *Diagnostic Analysis*

A total of seven studies were identified with this approach. Four of them (S5, S35, S38, and S50) employed supervised ML in combination with a predictive approach. Studies S8 and S29 applied unsupervised ML together with a descriptive approach. Study S49 used unsupervised ML in combination with a predictive approach. These findings highlight the importance of incorporating the diagnostic approach in the initial phases as a key element for advancing toward more sophisticated data-driven strategies, such as predictive or prescriptive approaches.

B. *Descriptive Analysis*

Six studies were identified with this approach. Two of them (S7 and S27) employed supervised ML models, while two others (S17 and S26) used unsupervised ML. It should be noted that studies S26 and S27 also combined the descriptive with the predictive approach. Studies S8 and S29 combined the descriptive with the diagnostic approach using unsupervised ML.

TABLE V
AIED APPLICATIONS WITH ML MODELS

Applications with Supervised ML (n=29)		
Administration (1)	Recommender Systems	**S44**
Learning (8)	AI-Based Career Prediction	**S23**
	AI-Based Feedback	**S11, S14, S46**
	AI-Based Feedback: Self-Regulated Learning (SRL)	**S25, S31**
	Collaborative Learning	**S2**
	Learning Style	**S35**
Teaching (1)	Student Modeling	**S40**
Cross-cutting (1)	Learning Dashboards	**S34**
Assessment (18)	Code Evaluation and Error Classification	**S38**
	Dropout Prediction	**S5**
	Sistemas de alerta temprana (EWS) → Student Dropout Prediction – Early Warning Systems (EWS)	**S7, S20, S30, S37, S43**
	Intelligent Tutoring Systems (ITS)	**S13**
	Performance and Achievement Prediction	**S27, S50, S9**
	Risk Prediction and Interpretability	**S12, S21, S24, S28, S4, S41**
	Sensor-Based Evaluation and Affective Analytics	**S33**
Applications with No Supervised ML (n=5)		
Learning (4)	Student modeling and behavior analytics	**S8, S26, S49**
	Student modeling and collaboration witn behavior analysis	**S17**
Teaching (1)	Social Interaction and Student Participation Monitorin	**S29**
Applications with Semisupervised ML (n=2)		
Learning (1)	Pedagogical Agents with Generative AI	**S22**
Assessment (1)	Student Dropout Prediction – Early Warning Systems (EWS)	**S3**

C. *Predictive Analysis*

A total of 19 studies were identified with this approach. Of these, thirteen (S4, S9, S12, S13, S20, S21, S24, S28, S30, S33, S37, S41, and S43) applied a purely predictive approach. The remaining six combined predictive with prescriptive analysis: five (S25, S31, S34, S40, and S44) employed supervised ML models, while one (S3) used a semi-supervised ML model.

D. *Prescriptive Analysis*

Within this group, study S22 stands out as the only one adopting a purely prescriptive approach, relying exclusively on a semi-supervised ML model. This singularity sets it apart from the other identified studies, which applied supervised ML models—either exclusively (S2, S11, S14, and S46) or in combination with predictive approaches (S25, S31, S34, S40, and S44)—making S22 a particularly relevant case within the analyzed corpus.

To address RQ3—*What AIED applications have been implemented in the studies?*—AIED applications were classified into five main categories: (1) learning, (2) teaching, (3) assessment, (4) administration [30] and (5) cross-cutting, as detailed in

Table V. It should be noted that, in order to incorporate multidisciplinary approaches with impact across multiple processes and domains of knowledge, the cross-cutting category was introduced. This category includes studies related to areas such as visual dashboards for learning monitoring, multi-agent systems, adaptive and personalized environments [33], [34], as well as tools for automatic content generation [35].

Table V reports 19 studies classified in the evaluation category, representing approximately 53% of all investigations that apply ML models, of which 18 employ supervised ML and one uses semi-supervised ML (S3). Although all studies fall under the evaluation category, they encompass diverse functional applications. A significant portion focuses on measuring academic performance through descriptive and predictive approaches (S27, S9), as well as on identifying risks of low performance and dropout (S12, S21, S24, S28, S4, S41). Noteworthy are studies that implement Early Warning Systems (EWS) to prevent school dropout (S7, S20, S30, S37, S43, S3), along with study S13, which focuses on the development of an Intelligent Tutoring System (ITS).

In the learning category, and due to its diagnostic nature and formative contribution within educational processes, a total of thirteen studies were identified. Five of these address personalized feedback strategies, supported by supervised ML models with a prescriptive approach (S11, S14, S46, S25, S31). Study S35 centers on identifying learning styles through a predictive approach, while S2 proposes support mechanisms for collaborative learning, and S23 focuses on predicting academic trajectories using AI. In the same category, four studies (S8, S26, S49, S17) employed unsupervised ML, targeting the analysis of student behavior and collaboration. Additionally, Study S22 employs agents to provide feedback aimed at enhancing students' self-regulated learning skills.

With respect to teaching-centered applications—aimed at supporting teaching tasks such as identifying social interaction and monitoring student participation—two studies were identified: one with a diagnostic approach (S29) and another combining predictive and prescriptive analysis for student modeling (S40). Collectively, these studies represent approximately 7% of the total analyzed. In terms of administrative applications, a single study (S44) was identified, accounting for 3% of the total, and focused on the development of a recommendation system based on diagnostic analysis. Finally, within the category of cross-cutting applications, study S34 stands out, proposing the design of learning dashboards aimed at monitoring student progress.

To address RQ4—What AI models with ML have been developed? —the results show that supervised ML predominates in the corpus of 36 studies (80.6%), whereas unsupervised (13.8%) and semi-supervised approaches (5.6%) are considerably less frequent (Fig. 4).

Within supervised ML, nearly 50% of the studies focus on educational assessment—such as performance prediction, dropout detection, and intelligent tutoring (S38, S5, S7, S13, S27, S41, among others). For teaching applications (S40), administrative management (S44), and cross-cutting applications (S34), only one study was identified in each category, representing 2.8% of the total in each case. A total of 22.2%

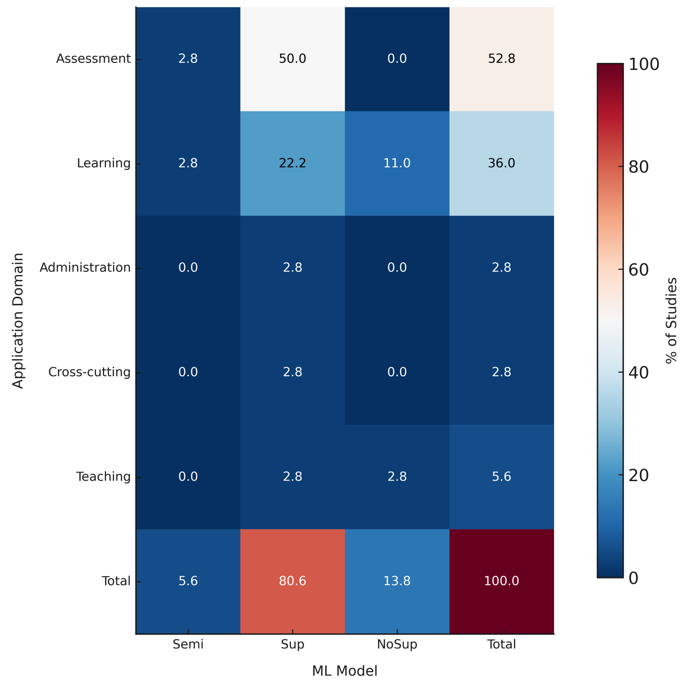

Fig. 4. IAED with models ML.

corresponded to learning applications (S2, S11, S14, S23, S25, S31, S35, and S46). In unsupervised ML, 11% addressed learning applications, primarily related to behavior analysis and student collaboration (S8, S26, S49, S17), while one study (2.8%) focused on teaching (S29). Two studies employed semi-supervised ML: S22, oriented toward personalized learning through generative AI pedagogical agents, and S3, focused on evaluation using early warning systems to prevent school dropout.

V. Discussion

The incorporation of data-driven decision models in education requires the establishment of robust management frameworks and accountability principles that guarantee transparency, trust, and respect for fundamental rights. The 36 studies presented in Table III demonstrate that data-driven decision models can be essential for integrating AI and enhancing the educational solutions offered to the academic community. In HEIs, their adoption should advance gradually, beginning with data literacy and the articulation of still-fragmented digital ecosystems. It is important to establish ethical and transparent principles for information management, as well as to promote pilot projects that, once evaluated, can be scaled to strengthen institutional management and teaching and learning processes.

To address RQ1, 78% of the studies analyzed correspond to empirical designs and 14% to mixed-method approaches, representing a combined empirical coverage of 92%, compared to only 8% conceptual (Fig. 3). This empirical predominance underscores the need to validate, through evidence, the implementations that integrate LA with AI models in teaching and learning contexts [1]. One way to simplify these processes is to begin with diagnostic approaches based on the available data and then progressively advance toward more sophisticated, higher-performing algorithmic models, grounded in prescriptive analysis and employing data-driven approaches. These models would enable the anticipation of learning behaviors on platforms—whether from labeled data or unstructured interactions—in order to achieve greater granularity of information and allow for the early detection of academic performance [29]. In this regard, the incorporation of ML models enables the anticipation of potential barriers to academic performance and the establishment of early warning systems that foster adaptive learning processes based on the available data. However, such advances must be grounded in robust data governance frameworks [9], whether for digitizing cognitive processes [8] for predicting and optimizing interactions in digital environments. Although empirical studies are the most represented, the limited availability of data restricts the traceability of cognitive processes, underscoring the need for strong governance architectures that align with pedagogical objectives and enhance the value of adaptive and personalized learning.

An effective selection of data-driven approaches must be aligned with institutional goals and increase the availability of data that enable the collection of relevant information. Table IV presents the data-driven decision methods employed and the ML models used in the reviewed studies. Diagnostic methods make it possible to identify the underlying causes of potential problems from observed correlations. Study S8 illustrates how anticipatory learning strategies can be developed from learning platform logs, while study S49 examines how data-centered diagnostic and predictive approaches employ unsupervised ML models to generate behavioral profiles and provide real-time personalized feedback. Diagnostic approaches can further increase their value if, from the outset, they integrate ethical principles such as those of QuantCrit [45], which underscore the need to interpret behavioral and psychological data within the social context in which they are applied. These principles warn of the risks that categories used to measure competencies may introduce biases that, rather than supporting learning, could reinforce educational inequalities or reproduce interpretations dependent on human judgment.

Studies employing descriptive methods have yielded relevant findings. Study S26 presents a data-centered model of individual differences that improves the prediction of performance and engagement, aligning with personalized learning strategies supported by empirical evidence. At the same time, study S17 examined Coursera interaction data—including clickstream logs and demographic information—to identify patterns of collaboration and potential cases of academic dishonesty. Data-driven methods benefit from structured frameworks such as CRISP-DM, which provides an iterative lifecycle suitable for validating educational data science initiatives and ensuring consistency between institutional questions, available data, and analytical design [16], [17].

Studies employing predictive methods include S12, S25, S28, and S31, which evaluated student performance using learning management system records. By contrast, studies S4, S9, and S41 used the OULAD dataset to assess the generalizability of model predictions. In S31, six ML algorithms were trained with data from 446 students, and the best-performing model was a prescriptive explainable AI (XAI) system that generated personalized feedback delivered through a dashboard aligned with Zimmerman's SRL model, supporting goal setting, planning, monitoring, and reflection. Study S25 applied an explainable model using learning platform data to predict performance and support self-regulation through personalized feedback and an automated dashboard, with the aim of incorporating additional time-management and scheduling functionalities. Predictive methods benefit from data science maturity frameworks, which support outcome forecasting and structured institutional decision-making [14], [15]. However, its implementation must be guided by data governance and ethical frameworks to prevent bias, ensure data privacy and security, and include oversight and monitoring mechanisms to avoid discrimination or inequality.

Studies employing prescriptive methods were identified, focusing on the use of data to recommend actions and quantify potential outcomes in order to optimize decision-making. Among them, study S22 developed a data-driven mechanism that enables pedagogical agents enhanced with generative AI to provide LA-based feedback, helping learners reflect on their learning processes and develop self-regulated learning skills. Study S14, which draws on a real dataset from an Australian HEI with 7,000 students (52% completed the course and 48% dropped out) collected from platforms such as Moodle, demonstrates how predictive modeling can be enhanced with prescriptive analytics. Through two case studies, the research shows how ChatGPT can be used to generate clear and actionable feedback for at-risk students. Prescriptive analysis requires high-quality data, reliable models, and robust processes to generate trustworthy and pedagogically sound recommendations [32], [33]. It is essential to place the data analysis layer at the core of an ecosystem that connects cyber-physical systems, organizational platforms, and learning platforms, serving as a nexus for key stakeholders in HEIs [18], and to define how data will be collected, analyzed, and interpreted, particularly when working with large-scale information [34]. The prescriptive approach not only forecasts but also recommends actions, whose effectiveness relies on reliable and validated data aligned with the educational model, making it necessary to validate the effectiveness of the algorithmic model employed to avoid false positives.

Understanding the educational context—what in the CRISP-DM framework is referred to as "Business Understanding"—is a key requirement for developing AI educational applications grounded in evidence-based solutions and aligned with quality standards. Table V classifies and presents the techniques employed in the ML models applied to the reviewed AIED applications, thus addressing RQ3. Study S22 examines the integration of generative AI agents and outlines the essential pedagogical principles to guide educators in the effective use of AI chatbots in educational contexts. Study S29 employs sensor data to support collaborative learning by detecting proximity, while Study S40 proposes an optimized data architecture model for adaptive approaches. Studies S25 and S31 demonstrate that explainable AI-based feedback can strengthen self-regulated learning and extend the analysis beyond dropout prediction to focus on competence development. The CDM and CTA frameworks are examples of how these theories can be integrated into AI platforms to personalize learning and provide intelligent feedback [21], [22]. To strengthen the use of data-driven approaches, HEIs must provide greater clarity in defining the purpose of using AI—such as conversational chatbots, intelligent tutoring systems, or predictive models of academic risk—and formulate hypotheses that yield reliable results supported by both psychopedagogical metrics and algorithmic performance measures.

The increasing adoption of AI in education highlights existing gaps in pedagogical design, ethical validation, and the integration of cognitive processes within learning analytics systems. To address RQ4, Fig. 4 shows that 80.6% of the studies reviewed employ supervised ML, of which 62% focus on applications related to educational assessment. This challenges us to move toward semi-supervised and unsupervised models and to incorporate applications in teaching, learning, and management. Among the most promising solutions are explainable AI (XAI) and intelligent dashboards. In studies S14, S25, and S31, XAI is used to provide transparent and personalized feedback that supports decision-making. Specifically, study S31 employs it for prescriptive purposes, recommending corrective actions, while study S25 uses it to identify the causes of low academic performance and strengthen self-regulation skills. Studies S14 and S34 implement intelligent dashboards that optimize real-time monitoring, foster student self-regulation, and contribute to the continuous improvement of teaching strategies. ML models enhance educational analysis by uncovering latent patterns and generating actionable insights [59]. Herefore, their application must be grounded in empirically supported behavioral constructs [63] and accompanied by rigorous methodological and algorithmic performance to ensure the reliability of model ML evaluation [60] and algorithmic performance. The integration of ML models into educational applications must be grounded in informed consent, validated instruments and constructs, and rigorous methodological evaluations that ensure the predictive validity of educational dimensions, the reliability of algorithmic performance, and adherence to ethical principles of equity and human dignity.

VI. Conclusion and Future Work

Of the 36 studies analyzed (2021–2025), the majority of AI applications in education are based on supervised ML and predictive approaches, primarily in the evaluation of low performance and dropout. The review identified only one study with cross-cutting educational applications, highlighting a significant gap in the integration of diverse tools and processes. The absence of research that methodologically documents the use, scope, and limitations of LA, ML, and AI in education underscores the urgency of establishing governance frameworks that guide their application in a safe, ethical, and responsible manner, as an essential condition for designing educational ecosystems centered on human well-being and aligned with the principles of Society 5.0. Likewise, the findings emphasize the need for HEIs to adopt comprehensive frameworks that ensure data quality and privacy. The ML models reviewed in the studies reveal opportunities to enhance cognitive skills such as self-regulation. For HEIs, adopting AI entails not only improving accuracy and pedagogical relevance but also fostering trust, protecting data, and consolidating ecosystems aligned with educational objectives. This requires robust governance for the use of generative AI through institution-specific implementations adapted to the institutional context. Functionalities such as Retrieval-Augmented Generation (RAG), the creation of curated knowledge bases, and domain-specific fine-tuning, together with the assessment of infrastructure costs, are key steps. Likewise, it is essential to integrate LLM Guardrails to ensure transparency, mitigate bias, and secure alignment with ethical and psychopedagogical principles.

Emerging research horizons must adopt approaches grounded in mixed methods, data, and security; conduct rigorous validation of assessment instruments; bring psychopedagogical theories closer to neurotechnology for the development of future competencies; and incorporate governance guidelines such as QuantCrit and inspection or observation methods oriented toward ethical use in learning environments, the detection of inappropriate language, and related social aspects, drawing on early detection tools and the use of multidimensional evaluation metrics. The challenge lies in moving educators and institutional leaders from passive adoption to the active promotion of AI applications that are ethically aligned with education and human development. This requires a progressive process of literacy that accompanies process automation and enables the evaluation of its impact on work and social awareness. AI in HEIs must be designed in accordance with social values and ethical principles, recognizing that data reflect both knowledge and bias, and that its integration, under robust governance frameworks, can contribute to a more equitable, inclusive, and dignified education.

VII. Statements and Declarations

The authors declare that there are no conflicts of interest regarding the research, authorship, or publication of this study.

References

[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. Higher 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*, S. ElAtia, D. Ipperciel, and O. R. Za"ıane, Eds., Hoboken, NJ, USA: 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 Syst. Appl.*, 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," *Interact. Technol. Smart Educ.*, vol. 19, no. 3, pp. 281–299, Aug. 2022, doi: 10.1108/itse-03-2021-0052.

[5] T. T. Nguyen, M. T. Nguyen, and H. T. Tran, "Artificial intelligent based teaching and learning approaches: A comprehensive review," *Int. J. Eval. Res. Educ. (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," *Multimedia 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*, R. K. Sawyer, Eds., Cambridge, U.K.: Cambridge Univ. Press, 2014, pp. 253–272, doi: 10.1017/CBO9781139519526.016.

[8] M. Rosala. *Task Analysis: Support Users in Achieving Their Goals*. Nielsen Norman Group. Accessed: Feb. 2, 2025. [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," *Inf. Fusion*, vol. 103, Mar. 2024, Art. no. 102135, doi: 10.1016/j.inffus.2023.102135.

[10] F. J. 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," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 8, no. 4, pp. 7–16, Dec. 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," *Innov. Educ. Teaching Int.*, vol. 61, no. 5, pp. 1029–1043, Sep. 2024, doi: 10.1080/14703297.2023.2271445.

[12] G. Siemens and D. Gašević, "Guest editorial-learning and knowledge analytics," *J. Educ. Technol. Soc.*, vol. 15, no. 3, pp. 1–2, 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. Educ. Technol.*, vol. 55, no. 2, pp. 439–459, 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 *Proc. IEEE 20th Int. Conf. Adv. Learn. Technol. (ICALT)*. 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 Comput. Sci.*, vol. 181, pp. 526–534, Jan. 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," *J. Simul.*, 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," *Technol., Knowl. Learn.*, pp. 1–34, Apr. 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*. Paris, France: OECD, Apr. 2019, pp. 1–252, 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 Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119150, 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," *Human Factors: J. Human Factors Ergonom. Soc.*, 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. Higher 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," *Frontiers Educ.*, vol. 8, May 2023, Art. no. 1186731, doi: 10.3389/feduc.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, Mexico, Tech. Rep. [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," *J. Open Innovation: Technol., Market, Complex.*, vol. 8, no. 1, p. 4, Mar. 2022, doi: 10.3390/joitmc8010004.

[28] G. Ibarra-Vazquez, M. S. Ramírez-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," *Technol., Knowl. 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," *Comput. Education: Artif. Intell.*, vol. 4, Jan. 2023, Art. no. 100118, doi: 10.1007/s10639-019-09886-3.

[31] A. H. Hoti, X. Zenuni, M. Hamiti, and J. Ajdari, "Student performance prediction using AI and ML: State of the art," in *Proc. 12th Medit. Conf. Embedded Comput. (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," *Learn., Res. Pract.*, 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," *ACM SIGKDD Explorations Newslett.*, 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. Higher 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. (EKS)*, vol. 25, Apr. 2024, Art. no. e31942, 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.*, pp. 1–52, Jan. 2025, doi: 10.1007/s10961-024-10176-5.

[38] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz. (2010). *Big Data, Analytics and the Path From Insights to Value*. MIT Sloan Management Review. [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. Operations Res.*, vol. 333, nos. 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. Control*, vol. 19, no. 4, pp. 1–16, 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, MA, USA: Harvard Bus. Review Press, 2013.

[42] C. Cote. (2021). *Types of Data Analytics to Improve Decision-Making*. Bus. Insights-Blog. [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," *J. Educ. Data Mining*, vol. 1, no. 1, pp. 3–17, 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. Informat. Decis. Making*, 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 Educ.*, 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,” *Stud. Eng. Educ.*, vol. 5, no. 2, pp. 150–174, Oct. 2024, doi: 10.21061/see.159.

[47] S. A. D. Popenici and S. Kerr, “Exploring the impact of artificial intelligence on teaching and learning in higher education,” *Res. Pract. Technol. Enhanced Learn.*, vol. 12, no. 1, p. 22, Dec. 2017, doi: 10.1186/s41039-017-0062-8.

[48] SAP.(2023). *Quées Inteligencia Artificial?*. [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,” *Educ. Media Int.*, 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 M. 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,” *Comput. Appl. Eng. Educ.*, 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, USA: Syngress, an imprint of Elsevier, 2017.

[54] B. Mahesh, “Machine learning algorithms—A review,” *Int. J. Sci. Res. (IJSR)*, vol. 9, no. 1, pp. 381–386, Jan. 2020.

[55] Geeksforgeeks.*Machine Learning Tutorial*. Accessed: Feb. 2, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/?ref=shm

[56] H. Geffner, “Artificial intelligence: From programs to solvers,” *AI Commun.*, 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. Higher 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,” *Comput. Education: Artif. Intell.*, vol. 2, Jan. 2021, Art. no. 100025, 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*, vol. 35. Amsterdam, The Netherlands: 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,” *Inf. Learn. Sci.*, 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. Learn. 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* (Accounting, Finance, Sustainability, Governance & Fraud: Theory and Application), M. Kılıç and S.B. Kahyaoğlu, Eds., Singapore: Springer Nature, 2023, 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,” *Comput. Educ.*, vol. 195, Apr. 2023, Art. no. 104724, doi: 10.1016/j.compedu.2023.104724.

[64] B. Kitchenham and S. Charters, “Guidelines for performing systematic literature reviews in software engineering,” Dept. Comput. Sci., Univ. Durham, Durham, U.K., Tech. Rep. 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,” *J. Syst. Softw.*, 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. (EKS)*, vol. 23, Apr. 2022, Art. no. e28600, doi: 10.14201/eks.28600.

[67] M. J. Page et al., “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, 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, ON, Canada: 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,” *Int. J. Environ. Res. Public Health*, 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,” *Stud. Higher Educ.*, 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,” *Frontiers Artif. Intell.*, vol. 4, Jan. 2022, Art. no. 713176, 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,” *Decis. Support Syst.*, vol. 168, May 2023, Art. no. 113940, doi: 10.1016/j.dss.2023.113940.

[74] M. Yağcı, “Educational data mining: Prediction of students’ academic performance using machine learning algorithms,” *Smart Learn. Environments*, 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. Higher 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. Zemmouri, “Learning behavior analysis to identify learner’s learning style based on machine learning techniques,” *JUCS J. Universal Comput. Sci.*, 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. Intell. Syst.*, 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. Düştegör, “Employability prediction: A survey of current approaches, research challenges and applications,” *J. Ambient Intell. Humanized 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,” *Appl. Sci.*, 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 Develop. Perspect.*, 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,” *Comput. Education: Artif. Intell.*, vol. 6, Jun. 2024, Art. no. 100244, doi: 10.1016/j.caeai.2024.100244.

[84] M. F. I. Amin, A. Shirafuji, M. 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," *J. Comput. Assist. Learn.*, 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," *Frontiers Artif. Intell.*, vol. 4, Nov. 2021, Art. no. 723447, 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," *Frontiers Educ.*, vol. 9, Dec. 2024, Art. no. 1421479, doi: 10.3389/feduc.2024.1421479.

[89] K. Akhuseyinoglu and P. Brusilovsky, "Exploring behavioral patterns for data-driven modeling of learners' individual differences," *Frontiers Artif. Intell.*, vol. 5, Feb. 2022, Art. no. 807320, doi: 10.3389/frai.2022.807320.

[90] Y. S. Balcioğlu and M. Artar, "Predicting academic performance of students with machine learning," *Inf. Develop.*, vol. 41, no. 3, pp. 896–915, 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," *Math. Problems Eng.*, 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," *South Afr. J. Inf. Manage.*, vol. 26, no. 1, pp. 1–10, 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," *Appl. Sci.*, 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. Educ. Technol.*, vol. 56, no. 3, pp. 1122–1147, 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," *J. Learn. 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," *Big Data Cognit. Comput.*, 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," *J. Data Sci.*, vol. 21, no. 4, pp. 745–760, Oct. 2022, 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," *Appl. Comput. Syst.*, 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 Gener. Comput. Syst.*, 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," *Technol., Knowl. Learn.*, vol. 29, no. 1, pp. 331–354, Mar. 2024, doi: 10.1007/s10758-023-09650-0.

[105] H. Tjahyadi, "The implementation of educational data mining in predicting students' academic achievement in mathematics at a private elementary school," *Int. J. Inf. Educ. Technol.*, vol. 15, no. 1, pp. 154–163, 2025, doi: 10.18178/ijiet.2025.15.1.2228.

Laura Icela González-Pérez received the master's degree in educational technology from the Tecnológico de Monterrey, the master's degree in neuromarketing from the Universidad Internacional de La Rioja, and the Ph.D. degree in knowledge society formation from the University of Salamanca. She is currently a Professor and a Researcher at the Faculty of Psychology of the Universidad Autónoma de Nuevo León (UANL), and a Level I member of Mexico's National System of Researchers. She is affiliated with JWEL and RedLaTEMx, she has received international recognition in open science, artificial intelligence, and educational innovation. Her research focuses on digital transformation, open education, and emerging technologies.

Francisco José García-Peñalvo (Member, IEEE) is currently a Full Professor with the Department of Computer Science and Automation, University of Salamanca (USAL). He leads the GRIAL research group and the Deputy Director of the Research Institute for Educational Sciences. He has held leadership roles as the Vice Rector and the Vice Dean of USAL and was a Distinguished Professor at the Tecnológico de Monterrey from 2016 to 2018. His work focuses on technology-enhanced learning and educational innovation. Included in Stanford's World's Top 2% Scientists list, he has published over 150 indexed articles from 2019 to 2024. He is the Editor-in-Chief of Education in the Knowledge Society and an associate editor of several international journals.

Amadeo José Argüelles-Cruz is a Full Research Professor at the Centro de Investigación en Computación of the Instituto Politécnico Nacional (IPN), Mexico. He is a Level II member of Mexico's National System of Researchers (SNII). His research focuses on artificial intelligence, pattern recognition, cloud computing, and datadriven intelligent systems applied to education, mobility, and urban sustainability. He has contributed to institutional, national, and international projects involving intelligent mobility, digital twins, smart campus systems, and urban analytics. He has collaborated with the MIT Senseable City Laboratory and the MIT City Science Laboratory through several seed-funded research initiatives addressing technology-enabled decision-making for complex urban environments. He has also participated in multidisciplinary programs such as Ecoemprende, developed with the Tecnológico de Monterrey and the Secretaría de Educación, Ciencia, Tecnología e Innovación of Mexico City. He actively promotes scientific development across Latin America through research networks, graduate supervision, academic collaboration, and conference organization. His work integrates computational methods, engineering principles, and societal needs to advance innovation in emerging intelligent systems.