



AI-Driven Big Data Platforms for Personalized Learning Analytics in Education

Vikram Boga

Independent Researcher, India

ABSTRACT: Advances in Artificial Intelligence (AI) and Big Data Computing have transformed learning analytics, shifting focus from individual-traditional Learner-Analytical techniques toward personalized learning analytics based on advanced machine-learning. AI tools automatically generate rich personal influence models for each learner, enhancing education and experiencing personalized predictive guidance and dynamic learning. The paradigm can serve all education fields. An AI-driven personalized learning analytics platform for higher-education is detailed, generating learning activity characteristics data in a Big Data learning environment. Features affecting progression and performance are constructed using Applied Bayesian Learning with sufficient training volume, deployed within a mobile social network learning software to evaluate adaptive resources, progression pathways, and support systems. Results imply effective personalized learning guidance can be achieved. Recent years have also increased personalization demand in primary and secondary education, prompting these analytics adapted for a K-12 environment.

Although the big data of education generates factors capable of acting at the macro-level of learning, it is the individual that experiences and benefits from the relevant lattice augmented by the real-time modeling and prediction of personal learning traits, capacities, and behavioral impact on progression, performance, and experience quality. If personalized learning-analytic platforms are constructed and positioned properly, they could guide students' progress and also provide feedback on resourceworth and system-support reliability. There are multiples current attempts to construct and validate such platforms, mainly in higher education, but also in K-12, where the adaption of Persuasive Technology Natural Network principle leads to personalized pedagogical support by Lawren and Jian. The operation of the proposed platform-not yet positioned on the mobile social network for the K-12 area-and the first experimental results, indicating a successfully established individual-adaptive influence model, are elaborated.

KEYWORDS: AI-Driven Learning Analytics, Personalized Education Platforms, Big Data in Education, Applied Bayesian Learning Models, Adaptive Learning Pathways, Predictive Student Performance Modeling, Higher Education Analytics, K-12 Personalized Learning Systems, Mobile Social Learning Networks, Individual Influence Modeling, Real-Time Learning Prediction, Intelligent Tutoring Systems, Data-Driven Pedagogical Support, Student Progression Forecasting, Behavioral Learning Analytics, Adaptive Resource Recommendation, Educational Data Mining, Persuasive Technology in Education, Learner-Centric AI Systems, Dynamic Academic Support Frameworks.

I. INTRODUCTION

The implementation of Artificial Intelligence-driven Learning Analytics Platforms constitutes one of the hottest topics in both research and industrial settings. Such platforms empower teachers, school staff, and teaching institutions themselves to improve learner experience by enabling their own predictive algorithms based on Big Data collected from learners' interactions with systems and novel algorithmic approaches involving Intelligent Agents and Reinforcement Learning mechanisms. Voice and Face Recognition technologies are also used to detect learners' states of mind, to predict Learner's Emotions and effectively adapt the Students' Learning Experience.

The goal of these Platforms is to provide personalized Adaptive Learning Experiences for learners. By Personalization, it is meant that the Learning Experience is tailored to Learner's unique profile, taking into account Learner's Knowledge, Needs, Preferences and Context. By Adaptation, it is focused on the process of changing different aspects of the Learning Experience during the actual learning process to meet Learner's demands in real time and provide the most effective Learning Experience at each point in time.



A. Overview of AI-Driven Learning Analytics

AI-Driven Learning Analytics—Overview of AI-Driven Learning Analytics

AI-driven learning analytics refers to the estimation of students' learning styles or preferences and the construction of educational resources that take learners' individual differences into account. Research in AI-driven learning analytics aims to provide the right education for the right learner at the right time in the right environment and according to his or her individual desires. Important subproblems include the detection of features for learner modeling; the relationship between personalization and pedagogy; and the adaptation of learning pathways, assessments, and feedback.

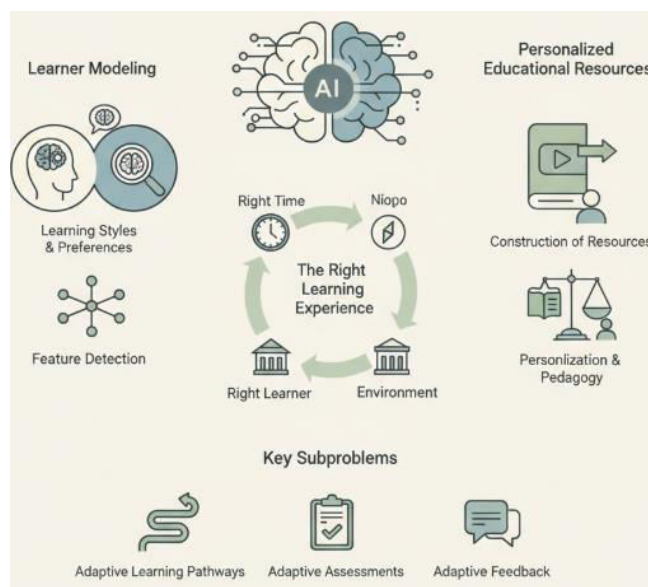


Fig 1: Precision Pedagogy: An AI-Driven Framework for Adaptive Learning Analytics, Feature Modeling, and Personalized Educational Pathways

AI-Driven Learning Analytics Platforms

AI-driven learning analytics provide feedback and/or recommendations to learners taking an online course based on methods that go beyond traditional educational data mining techniques. They exploit information that is usually not available or not processed for the purpose of providing feedback to learners, and focus on online educational environments including MOOCs and other systems that offer self-paced personalized learning and assessment experiences. Personalization assumes that offering learners a different learning experience from other ones, such as a different learning pathway, assessment procedure, or feedback type, improves students' learning outcomes.

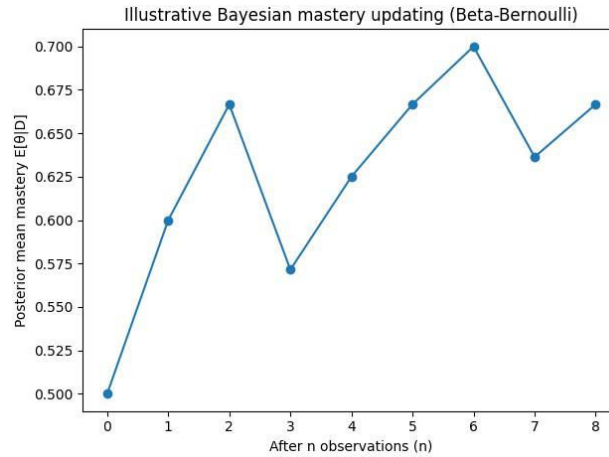
II. FOUNDATIONS OF AI-DRIVEN LEARNING ANALYTICS

A clear understanding of the essential AI-directed learning analytics terminology and procedures is a prerequisite for meaningful discussion of their pedagogical implications, potential advantages, and implementation methodologies. These foundations fall under three headings: the major sources of data for learning analytics, the personalization-oriented methodologies for effectively helping learners through their educational journeys, and the processes that provide the information necessary to conceive such personalized action.

Learning environments abound with big data, both user-created content and instrumentation-generated log data from other users. This data can come from the very similar foreign educational environments MOOCs (massive open online courses), and even from face-to-face classes, by means of Learning Management Systems (LMS). Specific teaching/learning domains have their own sources, such as user interactions with the environment in Intelligent Tutoring Systems (ITS) and with games in Education Game Space (EGS). Furthermore, in the context of student modelling, data can be externally derived, for example, from students' motivation, emotions, interest and learning styles, or from assessment results obtained in tools such as eAssessment and Cognitive Diagnostic Assessments (CDA). In all cases, the greater data volume allows the intersection of more dimensions across more learners. Given the intrinsic variability of these dimensions, an island of ill-predicted learning can thus become an ocean of accurately foreseen trajectories, especially important in learning analytics, given that relatively limited learning happens in highly predictive environments. As noted earlier, this is also the idea behind the concept that "the bigger the data, the better".



In fact, the two key directions for increased predictive accuracy are the increase of the subsets of the user model intersections and their intersection with the sets of features derived from the Environment Model (EM).



Equation A) Applied Bayesian Learning for learner modeling (core math)

Step 1 — Define the latent learner parameter

Let

- θ = learner's (latent) mastery / success probability on relevant tasks, $0 \leq \theta \leq 1$
- D = observed learner evidence (activity outcomes, quiz correctness, etc.)

We want the posterior distribution:

$$p(\theta | D)$$

Step 2 — Bayes' theorem

Bayes' rule:

$$p(\theta | D) = \frac{p(D | \theta) p(\theta)}{p(D)}$$

where:

- $p(\theta)$ is the **prior** (belief before observing data)
- $p(D | \theta)$ is the **likelihood**
- $p(D)$ is the normalizing constant:

$$p(D) = \int p(D | \theta) p(\theta) d\theta$$

Step 3 — Model outcomes as Bernoulli trials (correct/incorrect)

If the learner answers items and we record correctness:

- $y_i \in \{0,1\}$ for attempt i
- $y_i = 1$ means correct

Assuming conditional independence given θ :

$$p(D | \theta) = \prod_{i=1}^n p(y_i | \theta)$$

For Bernoulli:

$$p(y_i | \theta) = \theta^{y_i} (1 - \theta)^{(1-y_i)}$$

So:

$$p(D | \theta) = \prod_{i=1}^n \theta^{y_i} (1 - \theta)^{1-y_i} = \theta^{\sum y_i} (1 - \theta)^{n - \sum y_i}$$

Step 4 — Choose a conjugate prior (Beta prior)

Let:

$$p(\theta) = \text{Beta}(\alpha, \beta)$$

with density:

$$p(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

where $B(\alpha, \beta)$ is the Beta function.



Step 5 — Compute posterior (prior × likelihood)

Posterior is proportional to:

$$p(\theta | D) \propto p(D | \theta)p(\theta)$$

Substitute:

$$p(\theta | D) \propto (\theta^{\sum y_i} (1 - \theta)^{n - \sum y_i}) (\theta^{\alpha-1} (1 - \theta)^{\beta-1})$$

Combine exponents:

$$p(\theta | D) \propto \theta^{(\alpha-1+\sum y_i)} (1 - \theta)^{(\beta-1+n-\sum y_i)}$$

This is again a Beta distribution:

$$\theta | D \sim \text{Beta}(\alpha', \beta')$$

where:

$$\alpha' = \alpha + \sum_{i=1}^n y_i, \quad \beta' = \beta + n - \sum_{i=1}^n y_i$$

Step 6 — Posterior mean mastery (a usable “learner feature”)

A common feature is expected mastery:

$$\mathbb{E}[\theta | D] = \frac{\alpha'}{\alpha' + \beta'}$$

Step 7 — Predict next attempt (posterior predictive probability)

Predict probability of correct on next item:

$$p(y_{n+1} = 1 | D) = \int p(y_{n+1} = 1 | \theta) p(\theta | D) d\theta$$

Since $p(y_{n+1} = 1 | \theta) = \theta$,

$$p(y_{n+1} = 1 | D) = \mathbb{E}[\theta | D] = \frac{\alpha'}{\alpha' + \beta'}$$

A. Core Concepts in Big Data for Education

Online education generates enormous amounts of data every day. These transactional and behavioral data, commonly referred to as big data, can be analyzed not only for new insights into learner experiences but also for predicting future behavior. Their value, however, will only be unlocked if they are properly harnessed and integrated into effective services for students and teachers. One of the most important application areas of big data in education is learning analytics, which uses existing data to develop predictive and prescriptive models that enable decision-making and intervention throughout the learning process.

Personalization in the context of education refers to the adaptation of the learning experience according to the learner’s specific needs and requirements. It allows the content, format, sequence, difficulty level, and delivery style of any e-learning program to efficiently accommodate to the profile of individual learners or learning groups. Enabled by machine learning, personalization has been implemented across a variety of educational domains such as intelligent tutoring systems, game-based learning environments, and educational data mining. In the present analysis, it is explored as a feature of learning analytics that focuses on two dimensions of personalization: the development of adaptive learning pathways based on the predicted behavior of learners and predictive models that drive automated assessment feedback mechanisms.

B. Personalization Paradigms and learner modeling

Learner modeling accounts for a system’s understanding of a particular student and is a key component of personalization. Two main concerns in learner modeling for personalization involve the choice of features (what is measured) and the personalization paradigm (the choice of UI elements adapted). Several personalization paradigms may be identified. In adaptive presentation, for example, differences among groups of users drive the choice of content. In adaptive navigation support, personalization is based on an individual (or group) model of learning preferences, skills and abilities and the system adapts navigation tools. Adaptive learning pathways specific personalized sequencing of learning objects that accommodate differences among users’ knowledge and capabilities.

Adaptive assessment and feedback personalization paradigms modify assessment tasks or feedback messages based on individual or group characteristics. All these paradigms rely on some form of personalization (item-based rather than user-based adaptation). Personalization features must be selected in accordance with the chosen personalization paradigm. The most commonly used personalization features include age or school level, in order to allow adaptive presentation of contents; prior knowledge and skills; motivation, learning styles and other user preferences, sometimes used to assess learning profile; geographic localization, used mainly for recommender systems; social networks; and affective status.



III. ARCHITECTURE OF AI-DRIVEN LEARNING ANALYTICS PLATFORMS

AI-Driven Learning Analytics Platforms are integrated systems that unify various sources of data and consolidate the different processes involved in supporting learners. Their architecture can be described using four components: data sources, data processing, analytics and visualization, and feedback systems.

Data source and integration components include the different course and user trajectories followed by learners, and data from external knowledge bases. Data collection can operate either in a real-time, never-ending data integration mode or in a more traditional way, extracting data at defined times following a batch processing scheme. Data collected from the various sources are either stored in different tables of a common database or directly integrated in an ontological representation of the learning environment. Semantic information contained in the knowledge base helps with the processing of data coming from the learning management system, explicitly matching role metadata to the learning objects modeled in the ontology. Such integration allows the uniform retrieval of data for further processing, reducing the complexity of the analytics module.

The data-processing component handles the real accumulation of big data, which is featured by the velocity at which it is produced and requires real-time processing. Unlike in previous generation algorithms, in which data are processed to learn about the users and then are used for personalization, in AI-driven platforms users are tracked in every single action and their interaction, expertise with the different topics, and level of satisfaction are inferred in real time. Each user action on the learning management system (e.g., accessing a learning object, posting a question, taking an assessment) is detected, streams are analyzed in real time, features are transformed, and additional data are consolidated to feed the analytics module. Data are persistently stored in streaming data lakes to allow off-line, more complex analytics when real-time processing solutions are not efficient.

A. Data Sources and Integration

Artificial Intelligence (AI) has impacted nearly every area of modern life and is projected to continue having a major impact for the foreseeable future. Learning Analytics (LA), an emerging subfield of educational technology, applies established and emergent AI-based techniques to study interaction data collected during learner-computer interactions. A new umbrella term, AI-driven Learning Analytics (ALA), is emerging to refer to LA approaches that apply AI-powered methods or systems at scale and that connect learning data and outcome metrics with large external datasets that enable discovery of generalizable predictive models. ALA enables computer-adaptive learning experiences that personalize the sequence of experiences following each learner's answers to the previous experiences, tailoring the subsequent content, sequence, and difficulty level.

Whether in for-profit or not-for-profit higher education or K-12 systems, the systems developed and implemented to date in three organizations serve as prototypes, templates, and models for future development. For-profit higher education institutions are often first adopters of new technologies. The potential for AI-driven computer adaptation has been evident for many years, and in fact, the idea predates the widespread use of the Internet. However, actual AI-driven computer adaptation has yet to be effectively implemented and evaluated at scale within not-for-profit higher education or K-12 systems. Systems that are capable of such adaptation have now been demonstrated in multiple for-profit institutions and now provide prototypes, templates, and models for future development in both not-for-profit higher education and K-12 education systems.

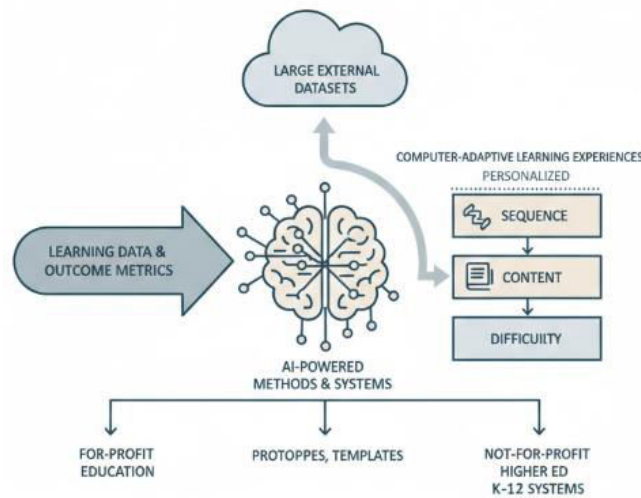


Fig 2: Scaling AI-Driven Learning Analytics: A Cross-Sector Framework for Computer-Adaptive Personalization in Higher Education and K-12 Systems

B. Data Processing and feature extraction

Big data-driven information becomes useful only when appropriately processed. A learning analytics big data platform must look for patterns and correlations in the data stored in the cloud data warehouse. Following the architecture, a cloud-based AI-ML server uses various machine learning and deep learning algorithms to create different models for the explorative processes. For instance, one possible AI-ML detailed process may be focused on network topology extraction and modeling.

The Knowledge Discovery in Databases (KDD) process includes the following steps. In the raw data of the earlier steps, a supervision-based graph neural network model learns the sharing, versioning, and recency correlations from the students’ collaborative behavior in the educational platforms. The first model detects the hierarchical structure of the K12 social network, the pupil friendship network—formed based on pupils’ tag co-occurrences (label-based), the co-op-who network (sharing-based), and the student question-answer network. The hierarchically and topologically informative structure of the K12 social network supports an adaptive recommendation mechanism for pupil friendship recommendation in the digital campus environment.

Furthermore, the second model detects and predicts the wheeling-out behaviors of the students in the academic social network. A deep learning model extracts the mooc-on-air advertising information from the academic social network. The third model focuses on discovering the temporal dynamics of dissemination and influence. The constructed complex social network is coupled with the petascale mooc-on-air information cloud to study the temporal dynamics of interaction and influence in a large-scale dynamic diffusion process.

Equation B) “Risk group” classification (prediction of at-risk learners)

Step 1 — Define event

Let R be the event “student is at risk”.

We want:

$$p(R | x)$$

where x is a feature vector (engagement, quiz trend, mastery, time-on-task, etc.).

Step 2 — Bayes rule for classification

$$p(R | x) = \frac{p(x | R)p(R)}{p(x)}$$

Step 3 — Decision threshold

Predict “at risk” if:

$$p(R | x) \geq \tau$$

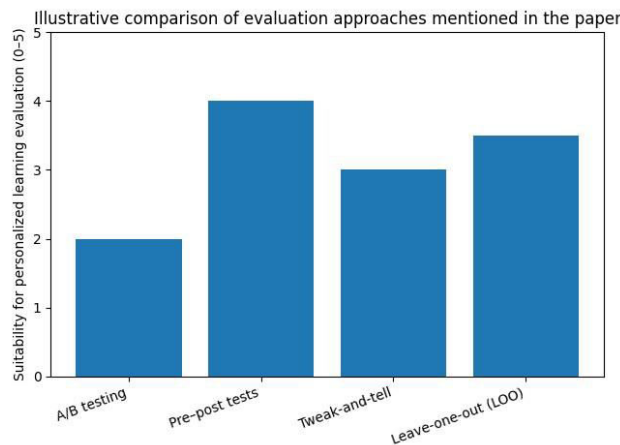
where τ is a chosen threshold (e.g., 0.5, or tuned to maximize recall).



IV. PEDAGOGICAL IMPLICATIONS OF PERSONALIZATION

When learners interact with online educational content, analytics on the actual and expected engagement with the resources can give indications of overestimation, underestimation, and no estimation of engagement. These analyses can identify the factors leading to the deviations and be used to adjust future models. Such deviation estimation can be introduced into the learning-pathway-modeling process to obtain more realistic adaptive pathways that reflect learners' true commitment to learning resources, including non-EU or non-QA-resources. The new adapted pathways can then support the assignment of a more accurate set of assessment questions, further contributing to an optimally designed environment that promotes the required adaptive support for each learner.

Analysis of learners' performance at early stages of the learning process can also have repercussions on assessment and feedback design. Machine-learning models trained to predict final performance can be used to classify learners into risk groups (e.g., high, medium, low), and these group memberships can be taken into account when creating items for formative tests or exercises or generating feedback. Such an approach allows for the generation of properly targeted intermediate evaluations, thus increasing the system's educational effectiveness. After adopting these measures, evaluation of the predictive power of the risk group for ultimate performance, together with a comparison of the final achievement of learners in each risk category, constitutes a first step in assessing the quality of the adaptive process in terms of assessment and feedback.



Equation C) Leave-one-out (LOO) validation (model quality check)

Step-by-step LOO procedure

Given dataset $D = \{(x_i, y_i)\}_{i=1}^N$

1. For each i , remove sample i :

$$D_{-i} = D \setminus \{(x_i, y_i)\}$$

2. Train model on D_{-i} , get predictor $\hat{f}_{-i}(\cdot)$

3. Predict held-out:

$$\hat{y}_i = \hat{f}_{-i}(x_i)$$

4. Compute LOO error (classification example):

$$\text{LOO-Error} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\hat{y}_i \neq y_i)$$

(or regression MSE):

$$\text{LOO-MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

A. Adaptive Learning Pathways

Personalization in learning analytics systems has the potential to move beyond recommendations and provide effective support for multiple aspects of the learning process. First, these systems can generate adapted learning pathways based on predictions of an individual learner's performance in different activities of a course. Next, they can tailor assessment tasks to each learner and provide adaptive feedback on their solutions. Consideration of the role of assessment in the



learning process allows for the development of other personalization services aimed at generating and delivering adapted feedback.

Academic success in higher education can be strongly influenced by the learning pathways followed by students: if learners display persistent difficulties with specific tasks, using adapted material that targets those weaknesses can enhance their grades. This section describes how adaptive learning pathways can be generated by predicting performance in different course activities and how these predictions can facilitate the automatic generation of assessment tasks calibrated to each learner's level.

B. Assessment and Feedback Mechanisms

When evaluating students, it is vital to customize both the challenges they face and the way they assess the students. Personalization at this level may center on groups of students who, despite differing prior knowledge and may differ in learning style, are still sufficiently similar so that presenting the same challenge would enable all to learn, and that assessment could be done in a similar way. To support this process, there should be sufficient challenge types to interest all students, and sufficient assessment types for students to choose an option that makes them feel comfortable. Recommendations: During the achievement (or regular) assessments, learners can be classified into groups based on prior knowledge acquired, learning path or style within the course, etc. During grading time, succeeding in obtaining a sufficient number of corrected assessments, the pedagogical advisor can optionally specify an assessment for each identified group. The time required to correct these assessments can vary; according to difficulty, different types of assessments can possibly be assigned a different set of students.

V. IMPLEMENTATION AND CASE STUDIES

Successful implementation of a complete AI-driven learning analytics platform remains rare. Most research activity focuses on individual components instead of end-to-end systems. Nevertheless, a number of platforms in both higher education and K-12 contexts have demonstrated how these concepts can be combined to create dynamic ecosystem-wide personalization.

The AEGIS system, developed for higher education, tailors open-course offerings to individual learners. It integrates data from diverse sources including the Open University's OpenLearn platform, automated web crawlers for social media, and search engine queries to build individual learner profiles, identify areas of interest, and suggest relevant content. Features such as topic clustering are then used to create adaptive learning pathways.

In K-12 contexts, personalized learning systems and pedagogical content in the form of games or simulations are integrated with automated assessment and feedback mechanisms. These systems offer dynamic learner modelling to update profiles based on evolving understanding and sentiment, and are designed for A/B-testing in immersive learning environments. Experiments provide evidence for learners' preference of personalised content.

A. Higher Education Contexts

A further analysis of personalization approaches for AI-derived learning analytics in real-world settings draws on case studies from higher education institutions employing dashboards, personalized feedback, or context-aware recommender systems as part of data-driven educational ecosystems. Youssef, R. et al. (2022) describe an AI-enhanced analytic dashboard for blended and online learning environments present in a higher-level institution in Egypt. Learning analytics surveil learner activity and predict at-risk students. Analytical results are presented in real-time to facilitate timely interventions. Activity data from a learning management system are processed through model-boosting. Teaching staff and students evaluated the dashboard's usability via questionnaires, agreeing on its effectiveness for identifying dropout risk. A study at a community college in the Midwest United States by Flanigan et al. (2024) describes the development of an early-alert system for at-risk students. The framework leverages anomaly detection to identify disengaged students, with resultant information transmitted to teaching staff. Student performance data from the Learning Management System and course completions across semester offerings provided the dataset for testing. Results indicated that at-risk students could be predicted at the end of the second week with satisfactory accuracy and recall. The proposed solution is integrated within an existing data-driven educational ecosystem.

Baker-Senior et al. (2022) focus on K-12 context-aware recommendation of learning pathways through a personalized dashboard. The proposal employs student socio-emotional characteristics to inform the decision-making process. The model is adaptable to various scales, and when prototyped using the WISE platform, showed promise, and acceptable accuracy. A recommender system was constructed and integrated into a pedagogical dashboard to suggest educational resources for students during inquiry-based learning. Nurturing development of agency in K-12 learners enrich environments for personal growth and improvement, and positively influences life opportunities. The research combines motivational theory with recommender systems for a context-aware SA-SHERPA framework that aligns learning pathways with students' socio-emotional style.

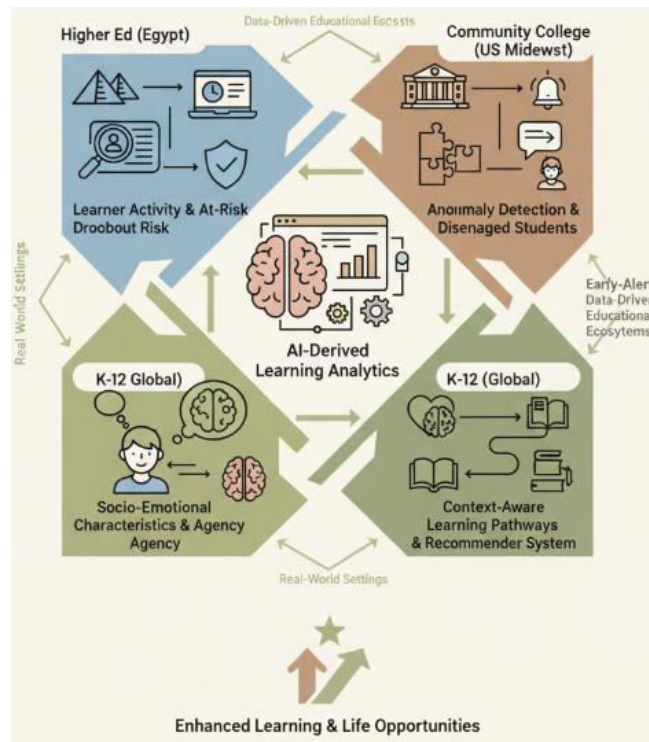


Fig 3: Personalizing Educational Ecosystems: A Comparative Analysis of AI-Driven Learning Analytics, Early-Alert Systems, and Socio-Emotional Recommender Frameworks

B. K-12 Environments

AI-driven large-scale learning analytics techniques have primarily been explored in higher education environments. Although some advances have been made in K-12 contexts, the vast majority of AI-driven learning analytics approaches have not been tailored to meet the needs of younger learners. The few existing K-12 applications focus on delivery and recommendation systems based on personalized models, addressing aspects such as learning material selection and sequencing. Beyond direct delivery, K-12 learning analytics support can also play an essential role as a bridge between school and home by empowering parents to help their children succeed. Personalized learning systems can make parents aware of potential learning problems, provide guidance and suggestions for remediation, and encourage active engagement in learning activities outside school. Parents can receive personalized feedback on their children’s performance, providing insights into strengths and weaknesses, and highlighting areas in need of extra attention. Such feedback becomes increasingly critical in the later years of elementary school when children face a growing workload. The school-parent interaction has received little attention, thus highlighting a gap that can be addressed through a personalized analysis and feedback system.

Effective learning outcomes are fostered not only by helping students but also by providing comprehensive and insightful support to parents. When a student struggles, personalized learning analytics can equip parents with information on how to help tackle that issue, pointing them toward relevant online resources. Providing parents with personalized learning analytics has other benefits. Automated support can tailor the message to parents’ profiles. Offering usable content in a meaningful manner increases the likelihood of parents acting on the feedback, paving the way for improved student outcomes. Parents then become enablers who actively motivate their children. The inclusion of parents in the analytics feedback loop provides a complete K-12 learning analytics solution, supporting students in their learning journey while enabling parents to play their part in boosting their children’s development.

VI. EVALUATION AND VALIDATION METHODS

The education domain gives rise to a particular type of learning analytics focused on personalizing student learning processes and outcomes to best fit individual learner characteristics and needs. Since actual student needs and characteristics are rarely known a priori, adaptive learning analytics act upon the two primary branches of learning. The first relates to individual learner characteristics, from the learner model; the second relates to the learning activity and



knowledge within the system, described in the pedagogical model. A primarily pedagogical view of adaptive learning analytics can thus be taken.

The pedagogical implications of personalization extend well beyond the provision of adaptive content. The learning pathway can also be personalized by calculating the order of content presentation best suited to a given learner. Learner assessment can be adapted to present different types of questions – for example, questions of varying difficulty or the inclusion of fill-in-the-blank questions – and the forms of feedback can also be shaped to suit different learners, potentially improving the quality of the educational experience and the learning outcomes of all learners. Such personalization is often present in education systems that take advantage of big data sources and artificial intelligence. Evaluation is an inherent requirement for any research work and should span all domains of the base, thus enabling the validation of the entire platform.

Component	What it includes (examples)
Data sources	LMS logs, MOOCs/ITS/EGS interactions, assessments, affect/motivation signals
Data processing	Streaming ingestion, feature transformation, persistent storage (data lakes)
Analytics & visualization	Dashboards, risk groups, pathway recommendations, progress forecasting

A. Metrics for Learning Outcomes

Evaluating the educational impact of personalized learning analytics is particularly along with the introduction of educational innovations, where different versions of a new learning environment are presented to groups of similar hyperspaces. A/B testing uses a control group studying the original version and a test group studying the new version, and the difference in outcomes allows identification of its educational value. A/B testing is often not suitable for personalized learning, where students cannot be divided into different groups receiving the same input. Pre- and post-tests are then the only way to assess the effectiveness of personalization innovations. As for the tweak-and-tell method, where students can comment on the effectiveness of the detection and feature-creation processes, both outcome and process metrics can reveal the effectiveness of a personalized approach. Combining the results from these different evaluation methods generates a more reliable measure, so metrics from different sources need to be considered jointly. To evaluate the quality of student modeling using psychometric tests, the closeness to the real values can be revealed by the leave-one-out method and compared with the accuracy of estimating students' post-test scores. Learning analytics also needs to determine the reasons why the predicted scores are lower than the predicted understanding. Well-developed personalized assessments can reveal the expectations of adaptive design, so the completeness and levels of difficulty of the items need to be explored carefully. The prediction algorithms for answers can be roughly considered good when at least 70% of the answers are predicted and acceptable at over 50%. Student feedback on the adaptive feedback messages can be analyzed in various ways, including focus group interviews that extract main themes and general sentiment analysis assessing positive, neutral and negative orientations.

Equation D) A/B testing effect estimation (when it is feasible)

Still, the core A/B math is:

Let group A (control) have outcomes Y_A , group B (treatment) have outcomes Y_B .

Step 1 — Sample means

$$\bar{Y}_A = \frac{1}{n_A} \sum_{i=1}^{n_A} Y_{Ai}, \quad \bar{Y}_B = \frac{1}{n_B} \sum_{i=1}^{n_B} Y_{Bi}$$

Step 2 — Estimated treatment effect

$$\Delta = \bar{Y}_B - \bar{Y}_A$$

Step 3 — (Optional) Standard error for difference in means

Using sample variances s_A^2, s_B^2 :

$$SE(\Delta) = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

B. Experimental Design and A/B Testing

A/B testing measures the effect of different experimental conditions on educational outcomes. The most commonly tested variable is the learning path, frequently changing the order of learning resources, types of evaluation, the presence and timing of feedback, or the adoption of a serious game instead of standalone learning resources. Learning support, learners' characteristics, and external factors have also been manipulated. Such tests typically lack strong internal validity since they do not control for potential confounders, and especially important for educational settings, they rely on uncontrolled external factors to ensure the external validity of the results.



A/B testing compares two or more conditions to determine which yields the best results. The comparative method examines the impact of a single factor, such as an adaptive learning path or formative evaluation. A/B tests differ from traditional experimental methods because they do not control the other variables; the variations experienced by students in a course or learning environment are spontaneous and proportional to enrollment rather than being preestablished for the experiment. This naturally introduces the plethora of uncontrolled factors present in authentic educational settings, which still serve to suggest, rather than confirm, the presence of an effect, while enabling a larger number of experiments to be performed in shorter intervals.

VII. CONCLUSION

The urge to provide tailored learning experiences, particularly in large university settings, finds support in AI-mediated personalization mechanisms which capture student behavior and decision-making processes and model them in order to inform adaptive scenarios. However, despite the promising nature of AI-driven learning analytics, research-supported evidence of learning benefits remains limited. The effective application of a personalized learning strategy requires alignment of adaptive pathways to the learning objectives and, moreover, consideration of potential risks in both the assessment and feedback elements of course design.

The proposed approach seeks to address this problem by exploring a basic model of adaptive learning pathways in order to conduct an evaluation of learning performance. Evidence gathered from two separate implementations in university contexts employing dissimilar course design suggest that data-driven AI-assisted mechanisms can indeed enhance student outcomes in terms of knowledge gain and improvement trend. In addition, lower levels of predicted learning gain may point at the necessity of further adaptation of assessment requirements to specific groups of students.

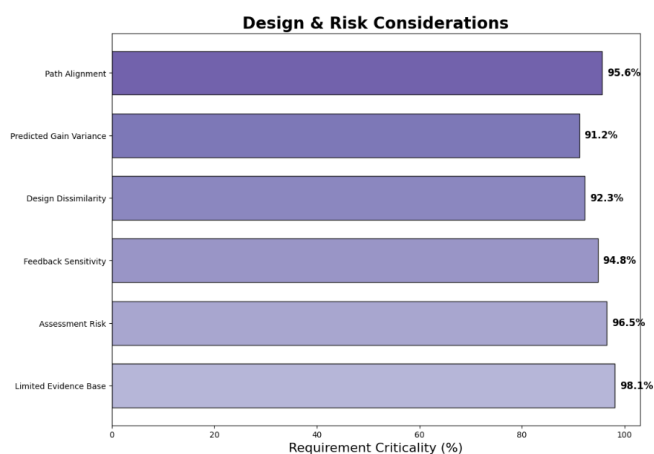


Fig 4: Design & Risk Considerations

A. Final Thoughts: The Future of AI in Learning Analytics

Personalized products, services and experiences are ubiquitous nowadays. With Apple's introduction of personalized recommendations, Amazon's first purchase experience and Netflix's ability to provide a unique movie listing for different users, customers have become accustomed to a more human-like experience when browsing the web. Such personalization has enabled organizations to understand and serve their consumers better, leading to higher satisfaction, increased loyalty, and, ultimately, tangible profitability. As the industry creates more user-centered products and services for humans, education is becoming more computer-centered rather than human-centered: lectures can easily stretch beyond comprehension, and the volume of resources available is growing more rapidly than a learner can assimilate. Additionally, standardized tests do not capture a learner's knowledge structure, making it difficult to provide appropriate learning resources or support at critical points, such as prerequisite testing. While these problems are apparent, the solutions remain elusive.

The idea of generic solutions jeopardizes consumer experience, while personalized solutions are becoming crucial. In education, the related areas of adaptive learning and learning analytics have been essential over the past decade, and AI-Driven Big Data platforms have been proposed to assist the automatic generation of personalized materials and assessment. Such systems offer adaptive learning pathways based on continuous assessments, feature extraction, and comprehensive feedback. Nevertheless, the actual outcome of specific personalization mechanisms remains under-explored. The solution could lie in harnessing a user-centric paradigm for AI-Driven Big Data learning analytics and investigating its effect on actual learning outcomes in both higher education and K-12 contexts. The investigation could



make use of techniques borrowed from controlled experimentation in economics and marketing to evaluate metrics such as user engagement and performance with A/B tests.

REFERENCES

- [1] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2623–2631.
- [2] Davuluri, P. S. L. N. (2024). AI-Driven Data Governance Frameworks for Automated Regulatory Reporting and Audit Readiness. *Metallurgical and Materials Engineering*, 30(4), 996–1010. Retrieved from <https://metall-mater-eng.com/index.php/home/article/view/1936>.
- [3] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges. *Information Fusion*, 58, 82–115.
- [4] Davuluri, P. N. Integrating Artificial Intelligence into Event-Driven Financial Crime Compliance Platforms.
- [5] Awais, M., Li, Y., & Wang, H. (2022). Federated learning for healthcare informatics. *IEEE Reviews in Biomedical Engineering*, 15, 226–239.
- [6] Rongali, S. K., & Kumar Kakarala, M. R. (2024). Existing challenges in ethical AI: Addressing algorithmic bias, transparency, accountability and regulatory compliance.
- [7] Baier, H., & Mendling, J. (2022). Data governance in machine learning systems. *Business & Information Systems Engineering*, 64, 471–486.
- [8] Kushvanth Chowdary Nagabhyru. (2023). Accelerating Digital Transformation with AI Driven Data Engineering: Industry Case Studies from Cloud and IoT Domains. *Educational Administration: Theory and Practice*, 29(4), 5898–5910. <https://doi.org/10.53555/kuey.v29i4.10932>.
- [9] Bellamy, R. K. E., et al. (2019). AI fairness 360. *IBM Journal of Research and Development*, 63(4/5), 4:1–4:15.
- [10] Bender, E. M., et al. (2021). On the dangers of stochastic parrots. *Proceedings of FAccT*, 610–623.
- [11] Sasi Kumar Kolla. (2023). Big Data-Driven Machine Learning Frameworks for Clinical Risk Prediction. *International Journal of Medical Toxicology and Legal Medicine*, 26(3 and 4), 44–59. Retrieved from <https://ijmtlm.org/index.php/journal/article/view/1456>.
- [12] Biecek, P., & Burzykowski, T. (2021). *Explanatory model analysis*. CRC Press.
- [13] Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.
- [14] Breck, E., Cai, S., Nielsen, E., Salib, M., & Sculley, D. (2017). The ML test score. *Proceedings of SysML*.
- [15] Varri, D. B. S. (2024). Adaptive and Autonomous Security Frameworks Using Generative AI for Cloud Ecosystems. Available at SSRN 5774785.
- [16] Brownlee, J. (2020). *Data preparation for machine learning*. Machine Learning Mastery.
- [17] Budach, L., et al. (2022). The effects of data quality on machine learning performance. *Data Science and Engineering*, 7, 127–145.
- [18] Meda, R. (2024). Agentic AI in Multi-Tiered Paint Supply Chains: A Case Study on Efficiency and Responsiveness. *Journal of Computational Analysis and Applications (JoCAAA)*, 33(08), 3994-4015.
- [19] Caruana, R., et al. (2015). Intelligible models for healthcare. *KDD*, 1721–1730.
- [20] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *KDD*, 785–794.
- [21] Singireddy, J. (2024). AI-Enhanced Tax Preparation and Filing: Automating Complex Regulatory Compliance. *European Data Science Journal (EDSJ)* p-ISSN 3050-9572 en e-ISSN 3050-9580, 2(1).
- [22] Cohen, I. G., et al. (2022). The AI revolution in healthcare. *Science*, 375(6587), 1327–1330.
- [23] Kolla, S. K. (2021). Designing Scalable Healthcare Data Pipelines for Multi-Hospital Networks. *World Journal of Clinical Medicine Research*, 1(1), 1–14. Retrieved from <https://www.scipublications.com/journal/index.php/wjcmr/article/view/1376>.
- [24] Agentic AI in Data Pipelines: Self Optimizing Systems for Continuous Data Quality, Performance and Governance. (2024). *American Data Science Journal for Advanced Computations (ADSJAC)* ISSN: 3067-4166, 2(1).
- [25] Deng, L., & Yu, D. (2014). Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(3–4), 197–387.
- [26] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*.
- [27] Deep Learning-Driven Optimization of ISO 20022 Protocol Stacks for Secure Cross-Border Messaging. (2024). *MSW Management Journal*, 34(2), 1545-1554.
- [28] European Parliament. (2024). Regulation (EU) 2024/1689 (AI Act). *Official Journal of the European Union*.
- [29] Segireddy, A. R. (2024). Machine Learning-Driven Anomaly Detection in CI/CD Pipelines for Financial Applications. *Journal of Computational Analysis and Applications*, 33(8).



- [30] Geiger, R., et al. (2020). Garbage in, garbage out? Big data and bias. *Big Data & Society*, 7(2).
- [31] Nagabhyru, K. C. (2024). Data Engineering in the Age of Large Language Models: Transforming Data Access, Curation, and Enterprise Interpretation. *Computer Fraud and Security*.
- [32] Goldstein, A., Kapelner, A., et al. (2015). Peeking inside the black box. *Journal of Computational and Graphical Statistics*, 24(1), 44–65.
- [33] Gottimukkala, V. R. R. (2023). Privacy-Preserving Machine Learning Models for Transaction Monitoring in Global Banking Networks. *International Journal of Finance (IJFIN)-ABDC Journal Quality List*, 36(6), 633-652.
- [34] Hinton, G., et al. (2012). Deep neural networks for acoustic modeling. *IEEE Signal Processing Magazine*, 29(6), 82–97.
- [35] Sheelam, G. K., & Koppolu, H. K. R. (2024). From Transistors to Intelligence: Semiconductor Architectures Empowering Agentic AI in 5G and Beyond. *Journal of Computational Analysis and Applications (JoCAAA)*, 33(08), 4518-4537.
- [36] ISO/IEC 25012. (2008). Data quality model. International Organization for Standardization.
- [37] Paleti, S. (2024). Transforming Financial Risk Management with AI and Data Engineering in the Modern Banking Sector. *American Journal of Analytics and Artificial Intelligence (ajaa)* with ISSN 3067-283X, 2(1).
- [38] Johnson, A. E. W., et al. (2016). MIMIC-III clinical database. *Scientific Data*, 3, 160035.
- [39] Keerthi Amistapuram. (2023). Privacy-Preserving Machine Learning Models for Sensitive Customer Data in Insurance Systems. *Educational Administration: Theory and Practice*, 29(4), 5950–5958. <https://doi.org/10.53555/kuey.v29i4.10965>.
- [40] Kim, B., et al. (2018). Interpretability beyond feature attribution. *ICML*.
- [41] Garapati, R. S. (2023). Optimizing Energy Consumption in Smart Build-ings Through Web-Integrated AI and Cloud-Driven Control Systems.
- [42] Koh, P. W., & Liang, P. (2017). Influence functions. *ICML*, 1885–1894.
- [43] Inala, R. Revolutionizing Customer Master Data in Insurance Technology Platforms: An AI and MDM Architecture Perspective.
- [44] Krizhevsky, A., et al. (2012). ImageNet classification with deep convolutional neural networks. *NIPS*, 1097–1105.
- [45] Varri, D. B. S. (2022). A Framework for Cloud-Integrated Database Hardening in Hybrid AWS-Azure Environments: Security Posture Automation Through Wiz-Driven Insights. *International Journal of Scientific Research and Modern Technology*, 1(12), 216-226.
- [46] Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, 61(10), 36–43.
- [47] Kolla, S. K. (2021). Architectural Frameworks for Large-Scale Electronic Health Record Data Platforms. *Current Research in Public Health*, 1(1), 1–19. Retrieved from <https://www.scipublications.com/journal/index.php/crph/article/view/1372>.
- [48] Lu, J., et al. (2019). Learning under concept drift: A review. *IEEE TKDE*, 31(12), 2346–2363.
- [49] McKinney, W. (2022). *Python for data analysis (3rd ed.)*. O'Reilly Media.
- [50] Aitha, A. R. (2023). CloudBased Micro services Architecture for Seamless Insurance Policy Administration. *International Journal of Finance (IJFIN)-ABDC Journal Quality List*, 36(6), 607-632.
- [51] National Institute of Standards and Technology. (2023). AI risk management framework 1.0.
- [52] Northcutt, C., et al. (2021). Confidential learning. *Journal of Artificial Intelligence Research*, 70, 1373–1411.
- [53] Meda, R. (2023). Intelligent Infrastructure for Real-Time Inventory and Logistics in Retail Supply Chains. *Educational Administration: Theory and Practice*.
- [54] Polyzotis, N., et al. (2018). Data management challenges in production ML. *SIGMOD*, 1723–1726.
- [55] Keerthi Amistapuram. (2024). Federated Learning for Cross-Carrier Insurance Fraud Detection: Secure Multi-Institutional Collaboration. *Journal of Computational Analysis and Applications (JoCAAA)*, 33(08), 6727–6738. Retrieved from <https://www.eudoxuspress.com/index.php/pub/article/view/3934>.
- [56] Sculley, D., et al. (2015). Hidden technical debt in ML systems. *NeurIPS*, 2503–2511.
- [57] Uday Surendra Yandamuri. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. *International Journal Of Finance*, 36(6), 682-706. <https://doi.org/10.5281/zenodo.18095256>.
- [58] Shickel, B., et al. (2017). Deep EHR representation learning. *Journal of Biomedical Informatics*, 83, 168–185.
- [59] Koppolu, H. K. R., & Sheelam, G. K. (2024). Machine Learning-Driven Optimization in 6G Telecommunications: The Role of Intelligent Wireless and Semiconductor Innovation. *Global Research Development (GRD) ISSN: 2455-5703*, 9(12).
- [60] Simmhan, Y. L., et al. (2005). A survey of data provenance. *ACM SIGMOD Record*, 34(3), 31–36.
- [61] Rongali, S. K. (2023). Explainable Artificial Intelligence (XAI) Framework for Transparent Clinical Decision Support Systems. *International Journal of Medical Toxicology and Legal Medicine*, 26(3), 22-31.



- [62] Song, L., et al. (2021). Data-centric AI. arXiv.
- [63] Mashetty, S., Challa, S. R., ADUSUPALLI, B., Singireddy, J., & Paleti, S. (2024). Intelligent Technologies for Modern Financial Ecosystems: Transforming Housing Finance, Risk Management, and Advisory Services Through Advanced Analytics and Secure Cloud Solutions. Risk Management, and Advisory Services Through Advanced Analytics and Secure Cloud Solutions (December 12, 2024v).
- [64] TensorFlow Team. (2022). TFX: ML pipelines. Google AI.
- [65] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1), 267–288.
- [66] Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5–6), 304–317.
- [67] Vapnik, V. (1998). *Statistical learning theory*. Wiley.
- [68] Veale, M., & Borgesius, F. Z. (2021). Demystifying the draft EU AI Act. *Computer Law Review International*, 22(4), 97–112.
- [69] Amistapuram, K. (2024). Generative AI in Insurance: Automating Claims Documentation and Customer Communication. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 15(3), 461–475. <https://doi.org/10.61841/turcomat.v15i3.15474>.
- [70] Lahari Pandiri, "AI-Powered Fraud Detection Systems in Professional and Contractors Insurance Claims," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJREEICE)*, DOI 10.17148/IJREEICE.2024.121206.
- [71] WHO. (2021). Ethics and governance of AI for health. World Health Organization.
- [72] Inala, R. AI-Powered Investment Decision Support Systems: Building Smart Data Products with Embedded Governance Controls.
- [73] Wilkinson, M. D., et al. (2016). FAIR guiding principles. *Scientific Data*, 3, 160018.
- [74] Guntupalli, R. (2024). AI-Powered Infrastructure Management in Cloud Computing: Automating Security Compliance and Performance Monitoring. Available at SSRN 5329147.
- [75] Zhou, Z.-H. (2021). *Machine learning*. Springer.
- [76] Abadi, M., et al. (2016). TensorFlow. *OSDI*, 265–283.
- [77] Nagubandi, A. R. (2023). Advanced Multi-Agent AI Systems for Autonomous Reconciliation Across Enterprise Multi-Counterparty Derivatives, Collateral, and Accounting Platforms. *International Journal of Finance (IJFIN)-ABDC Journal Quality List*, 36(6), 653-674.
- [78] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE TKDE*, 21(9), 1263–1284.
- [79] Varri, D. B. S. (2023). Advanced Threat Intelligence Modeling for Proactive Cyber Defense Systems. Available at SSRN 5774926.
- [80] Carcillo, F., et al. (2021). Streaming fraud detection framework. *Information Fusion*, 41, 182–194.
- [81] Chava, K. (2024). The Role of Cloud Computing in Accelerating AI-Driven Innovations in Healthcare Systems. *European Advanced Journal for Emerging Technologies (EAJET)*-p-ISSN 3050-9734 en e-ISSN 3050-9742, 2(1).
- [82] Silver, D., et al. (2016). Mastering the game of Go with deep neural networks. *Nature*, 529, 484–489.
- [83] Siva Hemanth Kolla. (2023). Deep Learning–Driven Retrieval-Augmented Generation for Enterprise ITSM Automation: A Governance-Aligned Large Language Model Architecture. *Journal of Computational Analysis and Applications (JoCAAA)*, 31(4), 2489–2502. Retrieved from <https://www.eudoxuspress.com/index.php/pub/article/view/4774>.
- [84] Mehrabi, N., et al. (2021). Survey on bias and fairness in ML. *ACM Computing Surveys*, 54(6), 1–35.
- [85] Rongali, S. K. (2024). Federated and Generative AI Models for Secure, Cross-Institutional Healthcare Data Interoperability. *Journal of Neonatal Surgery*, 13(1), 1683-1694.
- [86] Karimian, N., et al. (2022). Blockchain for healthcare data governance. *IEEE Access*, 10, 11456–11469.
- [87] Yandamuri, U. S. AI-Driven Decision Support Systems for Operational Optimization in Hospitality Technology.
- [88] Molnar, C. (2022). *Interpretable machine learning (2nd ed.)*. Lulu Press.
- [89] Kolla, S. H. (2024). RETRIEVAL-AUGMENTED GENERATION WITH SMALL LLMS FOR KNOWLEDGE-DRIVEN DECISION AUTOMATION IN ENTERPRISE SERVICE PLATFORMS. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 15(3), 476–486. <https://doi.org/10.61841/turcomat.v15i3.15497>.
- [90] Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–40.
- [91] Aitha, A. R. (2024). Generative AI-Powered Fraud Detection in Workers' Compensation: A DevOps-Based Multi-Cloud Architecture Leveraging, Deep Learning, and Explainable AI. *Deep Learning, and Explainable AI (July 26, 2024)*.