AI-Driven Conceptual Infrastructure Model of Learning Analytics in Higher Education

Petra Vondra[a], Josipa Bađari[b], Darko Grabar[c], Barbi Svetec[d] and Blaženka Divjak[e]
University of Zagreb, Faculty of Organization and Informatics, Croatia

Keywords: Learning Analytics, Higher Education Institutions, Infrastructure, Conceptual Model, Artificial Intelligence.

Abstract: The main aim of this research is to propose an AI-driven conceptual model of infrastructure supporting learning analytics (LA) in higher education (HE). The research is motivated by the lack of studies dealing with modelling, building, using and maintaining infrastructure for LA at higher education institutions, especially in the AI era, which was confirmed by the literature review. Namely, scoping literature review (ScR) was conducted to the key aspects to be considered when developing an LA infrastructure model. Moreover, key components of a conceptual model of LA infrastructure for HE are proposed, considering the requirements of the AI era and the insights from the ScR, as well as the practical and scientific expertise of the authors. As part of the design cycle, the first round of validation of the conceptual model was conducted.

1 INTRODUCTION

In the recent years researchers, analysts and theorists have increasingly argued that higher education (HE) should harness the data generated through the use of educational technologies by utilising learning analytics (LA), in order to better understand and optimize learning and learning environments (Long & Siemens, 2011; Macfadyen, 2022). Many HE institutions (HEIs) see the educational data, arising from the HEI stakeholders' activity, as a suitable source of further analysis for different levels of decision-making at HEIs (Sclater, 2014).

However, recent developments in artificial intelligence (AI), particularly the emergence of generative AI and large language models (LLMs), are fundamentally transforming how such data can be processed, interpreted, and used in educational contexts (Gaeta et al., 2025). These developments introduce new expectations for LA systems.

Ten years ago, most institutions were still at early stages of the LA maturity model (Siemens et al., 2013), indicating that, although HEIs were aware of the potential of LA, their use was largely limited to simple statistical reports and basic visualization tools (Ferguson et al., 2016).

In the meantime, the research field of LA has significantly developed, together with new technological solutions, especially with the advent of generative artificial intelligence (GenAI). Against this background, after 15 years of research, the definition of LA has been revisited and revised. LA is currently defined as "the collection, analysis, interpretation and communication of data about learners and their learning that provides theoretically relevant and actionable insights to enhance learning and teaching" (SOLAR, n.d.). The new definition emphasized the actionable nature of LA, calling for wider implementation, for which adequate infrastructure is a prerequisite.

At the same time, contemporary LA systems are increardsingly expected to incorporate AI-driven components, such as machine learning models, multimodal data processing, contextual data retrieval, and intelligent user interfaces (Ouyang et al., 2023; Ouhaichi et al., 2023). These requirements extend beyond traditional conceptualizations of LA infrastructure, which have primarily focused on data

[a] https://orcid.org/0000-0001-9945-8668
[b] https://orcid.org/0009-0003-1289-6748
[c] https://orcid.org/0000-0002-8317-1803
[d] https://orcid.org/0000-0003-1983-2467
[e] https://orcid.org/0000-0003-0649-3267

Vondra, P., BaĂŚari, J., Grabar, D., Svetec, B. and Divjak, B.
AI-Driven Conceptual Infrastructure Model of Learning Analytics in Higher Education.
DOI: 10.5220/0015048600004021
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 18th International Conference on Computer Supported Education (CSEDU 2026) - Volume 2*, pages 1030-1040
ISBN: 978-989-758-833-4; ISSN: 2184-5026
Proceedings Copyright © 2026 by SCITEPRESS – Science and Technology Publications, Lda.

collection, storage, analysis, and visualization layers (Badari et al., 2024).

The study presented in this paper is part of the continued research conducted within several projects (TRUELA and SAILeR) at the Faculty of Organization and Informatics University of Zagreb, which have been focused on developing new AI-supported LA approaches and methods that use large amounts of data generated (primarily) in LMSs. One aspect of such development is the continuous development of the appropriate concept of the infrastructure model in line with AI era trends.

Building on prior conceptual work on LA infrastructure, this research extends the existing model towards the AI era by incorporating AI-enabled components, refining the model through a scoping literature review, and supporting it with an initial round of expert validation.

Therefore, taking into account the existing research gap, it is proposed that a conceptual model of LA infrastructure be developed at HEIs in AI era that decide to implement LA and use its possibilities for institutional development.

2 BACKGROUND

Learning analytics (LA) has provided a new perspective to the use of data in educational decision-making by focusing on new digital data on learning activity, in combination with analysis techniques related to data science and artificial intelligence (AI). LA most often relies on the data produced through student interactions with IT systems, particularly LMSs (Divjak & Maretić, 2017), administration systems (Gašević et al., 2016), and other digital educational platforms where students interact with learning materials and educational technologies (Nguyen et al., 2023). In this respect, analyses can be done using trace/log data (Gašević et al., 2015; Gray & Bergner, 2022), assessment data (Divjak et al., 2022, 2024), textual data, self-reported and survey data (Gray & Bergner, 2022), or gameplay data (Reardon et al., 2022). Some of the latest approaches also use multimodal data from both digital and physical environments (Ochoa, 2022).

Based on these data, LA extracts "insights, decisions, and actions to improve learning and teaching" (Chatti et al., 2020). Understanding students' learning processes and difficulties can help teachers identify weaknesses in teaching and learning activities, challenging topics, and provide students with constructive feedback (Gašević et al., 2015). LA also supports personalized learning (Chatti & Muslim, 2019), self-regulated learning (Schumacher & Ifenthaler, 2018), flexible learning pathways (Tsai et al., 2018), and student retention (Susnjak et al., 2022).

Despite the recognized benefits, the use of LA is not fully widespread and systematic (Tsai et al., 2018), and some of the possible reasons are related to (the lack of) trust (Tsai et al., 2021). The trustworthiness of LA relates to a number of challenges (Svetec & Divjak, 2025). These include social and ethical challenges (e.g. privacy, data protection and data security; issues of agency, autonomy, control and data ownership), human and organizational challenges (e.g. competence and expertise of stakeholders) and technological challenges (e.g. infrastructure and algorithms). Across these aspects, ensuring accountability and transparency remains an important challenge.

The technological dimension of trustworthiness is closely connected to the design of the underlying LA infrastructure. To develop successful LA infrastructure, it is crucial to select the appropriate software architecture concept, which is not always straightforward. (Ciordas-Hertel et al., 2019) LA infrastructure must support the collection, integration, processing, and analysis of data originating from multiple sources and systems within the educational ecosystem.

In some cases, especially when designed for large user numbers, LA systems need to handle large volumes of data, arriving rapidly and possibly from diverse sources, such as various virtual environments and multimodal data streams. These data may be fragmented, duplicated, non-standardly represented and identified in different ways. (Ciordas-Hertel et al., 2019) Such challenges require big data software infrastructure, and anticipation of the volume, velocity and variety of data (Taylor & Munguia, 2018). In addition, the veracity of data, i.e. its trustworthiness and accuracy, is another important aspect to consider (Ciordas-Hertel et al., 2019). However, these challenges are not equally prominent across all institutional LA systems, as many implementations operate at a smaller scale and do not necessarily deal with big data.

Artificial intelligence (AI) has become increasingly integrated with LA in HE, enabling institutions to analyse large volumes of educational data and support data-driven decision-making. The integration of AI with LA has introduced new possibilities for predictive modelling, personalized learning and adaptive educational systems. Machine learning algorithms can analyse patterns in student behaviour and academic performance to identify students at risk of failure or dropout and enable timely

interventions (Ouyang et al., 2023). In addition, AI-based systems can provide personalized feedback, adaptive learning pathways and intelligent tutoring systems that support students' individual learning needs and improve engagement in online learning environments (Alam, 2021).

The emergence of large language models (LLMs) and natural language processing techniques has provided further possibilities for extension of LA systems' capabilities. These technologies enable the analysis of textual data such as forum discussions, essays and student feedback, providing deeper insights into learning processes and student engagement (Ouhaichi et al., 2023, Gaeta et al., 2025). As a result, AI-driven LA systems can support both students and educators by improving learning outcomes, facilitating personalized learning experiences and enabling more effective educational decision-making (Sajja et al., 2023).

Despite these benefits, several challenges remain in the implementation of AI-based LA systems. Key concerns include issues related to data privacy, data ownership, algorithmic bias and the ethical use of student data. Importantly, many institutions face difficulties in integrating diverse data sources and developing the necessary technical infrastructure and expertise required to effectively utilize AI-driven analytics systems in HE (Ifenthaler & Yau, 2020; Nguyen et al., 2023).

Taking into account the limited implementation of LA systems, as well as the possibilities provided by AI, this research aims to systematically examine the existing literature and to further develop the proposed conceptual model of LA infrastructure.

3 METHODOLOGY

Two research questions guided this research:
RQ1: What are the key aspects to consider when developing an LA infrastructure model?
RQ2: What are the core components of a conceptual model of LA infrastructure for HE in the AI era?
The research followed the design cycle methodology (Wieringa, 2014).

Responding to RQ1, in the **problem investigation phase** of the design cycle, a scoping literature review (ScR) was conducted to analyse the relevant literature and examine the aspects that should be considered when designing LA infrastructure. Scoping reviews are commonly used to clarify conceptual boundaries and identify key characteristics of a research field (Peters et al., 2015). In this study, this approach was applied to systematically map the literature and identify key aspects relevant for the development of an AI-driven LA infrastructure model.

The steps for conducting qualitative research in educational sciences were followed (Newman & Gough, 2020). This involved developing the research question (RQ1), designing a conceptual framework, constructing selection criteria, developing a search strategy, selecting studies using the selection criteria, coding studies, assessing the quality of studies, synthesizing results of individual studies and reporting the findings.

After forming the research questions, a search was conducted in the Web of Science and Scopus databases. Titles, abstracts and keywords were searched by choosing a respective database category. In the first phase of this research, the initial literature review from the previous research was partially reused (Badari et al., 2024). No additional database search was conducted for that period. Instead, the identified articles were re-examined with a specific focus on AI-related aspects, and a verification was performed to determine whether AI-focused studies were present within the original time frame. Additionally, the updated analysis includes newly identified articles published between September 2023 and March 2026.

The first database search was performed in August 2023 and an additional one was conducted in March 2026. The ScR workflow presented here (Figure 1) in line with the PRISMA reporting guidelines (Page et al., 2021) presents the process of selection of articles for the analysis.

The search was led by the following query: ("learning analytics" OR "educational data mining") AND ("higher education" OR HEI OR university OR college OR undergraduate OR graduate) AND infrastructure. In total, including both phases, 97 papers (50 in the first search and 47 in the second) were identified. After deduplication, the abstracts were coded, considering the following characteristics of studies: (1) relation to LA, (2) relation to HE, (3) relation to infrastructure, (4) inclusion of AI. While the first three characteristics served as inclusion criteria, the fourth one was included to gain additional insights. Finally, 17 papers related to LA, HEI and infrastructure were thoroughly analysed.

Responding to RQ2, in the **treatment design** phase, insights from the ScR were used as the basis for the development of an AI-driven conceptual model of LA infrastructure. The model synthesizes the infrastructural components and relationships identified in the literature, but also expert inputs and experiences of the authors. The conceptual model

development was guided by the Code of Practice for Learning Analytics (Sclater, 2016) which defines that LA uses data about students and their activities to help institutions understand and improve educational processes, and provide better support to learners.

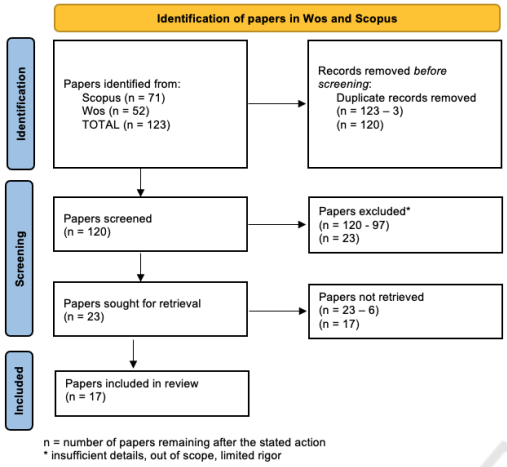


Figure 1: ScR workflow.

Finally, in the **treatment validation phase,** a preliminary validation of the proposed model was conducted by collecting expert feedback from nine participants with experience in LA and educational technologies. Feedback was collected through an online survey consisting of 20 questions. The survey aimed to capture both quantitative insights on the perceived quality of the model and qualitative inputs for further refinement.

In particular, 15 Likert-scale items (ranging from 1 – "Strongly Disagree" to 5 – "Strongly Agree") were used to evaluate the following dimensions of the model: clarity of structure, data coverage, system components, ethics, flexibility, and implementation feasibility. Moreover, five open-ended questions were used to gather qualitative feedback regarding the model's strengths, challenges, missing components, ethical concerns, and suggestions for improvement. The iterative nature of the design and validation process also reflects principles of the Plan-Do-Check-Act (PDCA) cycle (Deming, 1986) where insights from the validation stage inform potential refinement of the proposed model.

4 RESULTS AND DISCUSSION

4.1 Insights about LA Infrastructure

Responding to RQ1, this ScR indicates that relatively few studies provide a holistic view of LA infrastructure, encompassing modelling, development, use, and maintenance. Still, existing studies offer valuable insights into specific aspects of LA infrastructure, including data characteristics, design strategies, and implementation challenges.

Organizational and Strategic Aspects. The effective utilization of LA depends on the proper selection of an appropriate LA architecture strategy (Drlik et al., 2018). According to authors (Ciordas-Hertel et al., 2020), in infrastructure design, it is essential to capture the semantics, identify the need for support, and present the results or interventions to the stakeholders. Various design strategies can be used to develop LA infrastructure: experimental, custom application, data warehouse, and data lake design.

When it comes to the challenges related to development and adoption of LA at an institutional level, a cross-cases analysis of two LA initiatives was presented (Perez-Sanagustin et al., 2022), one with a bottom-up and another with a top-down approach. Authors stated that it is important to combine top-down and bottom-up approaches to facilitate the involvement of different stakeholders during the design and the implementation of a LA tools; anticipate the need for servers and data warehouses to integrate data from different sources; and inform stakeholders about the potential that LA tools could have for addressing institutional needs. LA initiatives require the coordination of both organizational aspects and infrastructures development. Furthermore, some recent studies have shown that a major impediment to the uptake of LA intervention (LAI) is the lack of access to LAI infrastructure by educators to pilot LAI, which typically requires substantial institution-wide efforts and investment to collect related data sets and develop accurate predictive models that identify at-risk students and also provide tools to facilitate interventions (Alalawi et al., 2025).

Other authors (Shacklock, 2016) point out that, to ensure students get maximum benefit from analytics, universities should use analytics systems that are: designed in consultation with students; supported by an ethical framework or policy; driven by the improvement of learning and teaching processes and student engagement; tailored to the particular needs of each institution and embedded in the institution' strategic plan. Authors (Clark et al., 2020) stated that key factors for successful implementation of LA in HE are strategy and policy at the organizational level, information technological readiness, performance

and impact evaluation, people's skills and expertise and data quality.

These findings highlight that LA infrastructure development is closely intertwined with institutional strategy, governance, and resource availability.

Data and Infrastructure Characteristics. It was pointed out (Ciordas-Hertel et al., 2020) that most LMSs do not include suitable tools to analyse data use. They also cannot be integrated with another LMS to provide a more holistic perspective on learning. For that, more advanced institutional-level infrastructure is needed. Data should be collected from various data sources, such as LMSs, student registration, survey systems, and library information systems, as well as cameras and sensors.

(Shacklock, 2016) emphasize that the HE sector has always been a data-rich sector, and universities generate and use enormous volumes of data each day. However, the sector has not yet capitalized on the enormous opportunities presented by the data revolution. In a related vein, other authors (Ciordas-Hertel et al., 2020) pointed out the high volume, high velocity and variety of data which should be considered in institutional-level infrastructure design, qualifying LA infrastructure as big data infrastructure, and bringing specific challenges. In this respect, their findings suggested the suitability of general big data infrastructure for the design of LA institutional infrastructure. However, authors (Caspari-Sadeghi, 2023) acknowledged that education lacks computational infrastructure and human capacity to fully exploit the potential of big data.

(Bergdahl et al., 2020) explored the use of LA in blended learning scenarios, pointing out that this context integrates different modalities, both physical and digital. Interestingly, they found that a majority of studies use digital traces, dispositional self-reports, interviews, relational data, not fully covering the possible data sources of the classroom. Furthermore, other studies emphasize the role of big data and assessment-related data in LA (Caspari-Sadeghi, 2023), highlighting the growing demands placed on LA infrastructure, outlining the four main LA functions in assessment, i.e. monitoring and analysis; automated feedback; prediction, prevention, and intervention; and new forms of assessment. Finally, it was found (Bergdahl et al., 2020) that data used in many LA studies were used as a proxy for what happens in the classroom. However, when studies do not include manifestations in the real classroom, they fall short of explaining learning or offer a guide on how to intervene in the classroom. Qualitative data can help better understand the reasons behind the trace data (Salonen et al., 2021). This reflects the fact that LA should account not only for the analysis of trace data from IT systems, but also other types of data, including multimodal.

Together, these studies emphasize the importance of robust data infrastructures capable of integrating and managing diverse and large-scale data sources.

Functional Components. Researchers point out that an essential factor in LA intervention infrastructure refers to developing predictive models using various data sources (student demographic, performance, engagement and other types of data). Enabling access to these kinds of datasets calls for institutional efforts accessing different IT systems, which is normally not accessible to educators individually (Alalawi et al., 2025). Furthermore, authors (Bergdahl et al., 2020) point out that, while predictive models are often able to infer future performance, they cannot explain learning or guide intervention. Some studies extend LA beyond prediction by integrating action-oriented frameworks that support targeted interventions based on student performance predictions (Alalawi et al., 2025). Such approaches highlight the need for LA infrastructure to not only generate insights but also support timely and context-aware interventions.

LA systems are not only analytical but also require careful design of user-facing components such as dashboards, which integrate data, analytics, and visualization layers. LA dashboards are important parts of LA infrastructure, as they enable visualization and end-user access to LA results. To support dashboard development, a checklist has been proposed to guide the planning, designing, implementation, and evaluation of LA dashboards (Kaliisa et al., 2023). Of particular relevance to this study is the designing dimension, in which respect the authors emphasize the consideration for pedagogical theory and evidence, learning design, user needs and competences, technical requirements and available resources, availability of data, stakeholder involvement and co-design, and ethical issues. This highlights that LA infrastructure extends beyond backend data processing to include user-facing design components that shape how insights are accessed and interpreted.

Authors (Baneres et al., 2021) emphasized the importance of AI systems in education to empower teachers during the teaching and learning process and introduced infrastructure for trustworthy systems to detect at-risk learners early on in order to help them to pass the course. Moreover, it has been noted (Pinto et al., 2025) that AI technologies such as intelligent

tutoring systems, LA, and educational chatbots can significantly support personalized learning and improve learning outcomes. However, successful implementation depends on the development of digital competencies and adequate technological infrastructure within HEIs. These findings further emphasize that AI integration introduces additional infrastructural requirements, including computational capacity, data integration, and support for adaptive learning processes.

Ethics. Some studies pointed out that data protection and privacy should be considered in the development of LA infrastructure (Ciordas-Hertel et al., 2020). Infrastructure should also incorporate privacy-preserving mechanisms, such as the ability to collect and analyse data anonymously at the group level. This can support more natural student behaviour and reduce bias in the data, thereby improving both ethical compliance and data validity. (Ameloot et al., 2022) This highlights that ethical considerations are not only normative requirements but also influence data quality and system effectiveness. Ethical considerations cut across all layers of LA infrastructure and are essential for ensuring trustworthiness and effective use.

Overall, the reviewed studies suggest that LA infrastructure is a multi-layered construct that encompasses organizational, data-related, and functional components. While most studies focus on specific aspects, such as data integration, predictive modelling, or visualization, few provide a comprehensive view that integrates these elements into a coherent framework. This fragmentation highlights the need for a conceptual model that systematically connects these components.

4.2 Conceptual Model

Regarding RQ2, the further development of a conceptual infrastructure model for LA in HE is driven by the increasing potential of LA to support learning processes, improve teaching practices, and enable data-informed institutional decision-making (Sclater, 2016). In the context of the AI era, these capabilities are further extended through the use of advanced analytics, predictive models, and intelligent support systems (Ouyang et al., 2023).

Accordingly, the proposed model aims to support the systematic collection and analysis of data on students' and teachers' activities, with the goal of enhancing learning outcomes and enabling more equitable learning opportunities. In addition, the model supports continuous monitoring and interpretation of learning processes, providing insights for institutional decision-making, improving system efficiency, and enabling timely and targeted interventions.

The proposed model builds upon previously proposed three-layer LA infrastructure model (Badari et al., 2024) and extends it by incorporating AI-enabled components and additional architectural elements.

The proposed infrastructure includes four components, aligned with the main aspects of LA, which include collecting, analysing and communicating data (SOLAR, n.d.). Accordingly, the conceptual model includes three layers: 1. Data sources, 2. Data storage and analysis, and 3. Visualization and action, and additionally an AI subsystem, linked to two layers. The infrastructure model is shown in Figure 2.

This research followed a design science research approach, operationalized through the design cycle (Wieringa, 2014). However, the iterative nature of the design and validation process also reflects principles of the Plan–Do–Check–Act (PDCA) cycle (Deming, 1986), where insights from the validation stage inform potential refinement of the proposed model. In this sense, the conceptual model of infrastructure was also developed according to the PDCA steps:

- **Plan:** literature review identifying the key aspects to consider, best practice examples; defining objectives for developing the conceptual model of LA infrastructure for HEIs
- **Do:** designing the conceptual model
- **Check:** conceptual model validation
- **Act:** final model development based on validation results, accompanied with recommendations for use in the AI era

It is important to note that the development of this conceptual model is cyclical. In this sense, in this paper, the first iterations of the Plan, Do, and Check stages are covered (i.e. the problem investigation, treatment development and treatment validation of the design cycle). However, the development goes on, and future cycles are planned. Therefore, for example, validation has so far been conducted through a survey, but further validation activities are foreseen (e.g. focus groups, interviews).

The overall AI-driven conceptual model of LA infrastructure is presented in Figure 2 and described in the following paragraphs. The 3-layered structure is presented in this order to reflect a user- and action-oriented perspective, starting from the visualization and decision-making layer as the primary point of interaction, and then tracing back through the underlying analytical, AI-driven, and data source layers that support it.

4.2.1 Visualization and Action

LA shall be actionable. To ensure that, it is very important to develop an effective and user-friendly End User Application which delivers understandable insights but also prompts users to take action. To support different evidence-based perspectives of using LA for different users (Jivet et al., 2018; Kaliisa et al., 2021; Verbert et al., 2020) it is important that End User Application provides: near real time, informative, accurate, personalized and predictive feedback visualized at dashboard; timely interventions through alert and warning system to ensure users' timely and appropriate action; students' engagement through a success planner; ability for students to take ownership and control of their data through informed consent.

The End User Application includes key functionalities that support user interaction with learning analytics results, such as providing an overview of relevant indicators for decision-making (Dashboard), enabling users to manage access to and use of their personal data (Consent Management), supporting timely interventions based on identified patterns and predictions (Alert and Intervention System), allowing users to define and monitor personal learning goals (Success Planner), and presenting analytical results through interactive visualizations that facilitate understanding and exploration of data (Visualization Platform).

In addition to traditional visualization and reporting components, the application integrates an AI agent that enables conversational access to LA insights and supports the transformation of analytical outputs into personalized explanations, recommendations, and decision-support. AI Agent is an interactive interface between end users and the LA system. It enables users to access insights and recommendations through a conversational interface. The agent receives user queries, identifies missing or relevant contextual information, and may request additional input when necessary. Based on the user's request, it communicates with the AI subsystem to retrieve and generate appropriate responses. The AI agent presents results in a structured and user-friendly format, which may include explanations of analytical results, step-by-step recommendations, warnings, or references to relevant data sources.

4.2.2 Data Storage and Analysis

This layer provides the core analytical capabilities of the system by enabling integration, processing, and interpretation of data from multiple sources. It includes components responsible for data acquisition and preparation (Analytical Database), advanced analytics and insight generation (Analytics System), orchestration of analytical workflows (Analytical Processor), and communication with external systems (Interoperability Platform).

The Analytical Database supports data collection, integration, and storage (Connectors to Data Sources, System for Data Integration, Data Storage), while the Analytics System enables both exploratory and predictive analysis (System for Deep Analysis and Predictive Analytics, Query and Report Builder). The Analytical Processor ensures coordinated execution of analytical processes, and the Interoperability Platform enables seamless data exchange across system components.

4.2.3 Data Sources

Effective implementation of learning analytics across diverse learning environments requires data sources that capture learner activities in both online and face-to-face contexts (Bergdahl et al., 2020). Ensuring data quality, reliability, and integrity is essential for building trust in learning analytics systems and enabling meaningful use of the generated insights (Svetec & Divjak, 2025). Learning analytics data is typically generated through interactions with institutional and digital systems and can be broadly categorized into system-generated, self-reported, and assessment-related data, with increasing use of multimodal data sources to support more comprehensive analysis (Ochoa, 2022).
Taking this into account, the proposed LA infrastructure considers multiple types of data sources, including system-generated data on student activities and performance, technology-supported behavioural tracking in face-to-face environments, self-reported data from learners and educators, and usage data from learning analytics applications.

4.2.4 AI Subsystem

The conceptual model includes an AI subsystem which acts as an intelligent service component supporting both analytical processing and decision-making. The AI subsystem communicates with the Data Storage and Analysis layer, where it utilizes integrated datasets and predictive models, and with the Visualization and Action layer, where it provides interpreted insights, explanations, and recommendations to end users.

The architecture of the AI subsystem follows the principles of Retrieval Augmented Generation (RAG), where GenAI models are combined with contextual data retrieval mechanisms. Instead of

relying solely on pre-trained knowledge, the system retrieves relevant information from internal knowledge bases, institutional data repositories, and external data sources, which are then used as contextual input for generating responses.

Within the Data Storage and Analysis layer, the AI subsystem supports the development and use of machine learning models for predictive analytics, such as identifying students at risk of failure or dropout and detecting patterns in learning behaviour. The subsystem may also integrate external knowledge sources and models to enhance prediction accuracy and extend analytical capabilities.

The AI subsystem supports the Visualization and Action layer by transforming analytical results into understandable and actionable insights. Through intelligent interpretation of analytics results, the subsystem can generate personalized recommendations, contextual explanations, and adaptive feedback for students, teachers, and decision-makers. The AI subsystem includes:

Gen AI Application Module. This module orchestrates the interaction between users, the AI subsystem, and various internal and external data sources. It manages the overall process of generating AI-supported responses and recommendations based on LA data. The module maintains the conversational context, including user profile information and interaction history, and determines which information sources should be used for a particular request. It may retrieve information from internal knowledge bases, structured databases, or external data sources. Furthermore, the module coordinates knowledge retrieval processes and invokes appropriate tools or services for accessing additional data sources. Based on the retrieved contextual information, the module interacts with an LLM to generate the final response, interpretation, or recommendation for the user.

Contextual Data Retrieval System. This system retrieves relevant information from internal databases and knowledge repositories to provide the AI models with appropriate contextual data.

Knowledge Base. It stores LA-relevant structured and unstructured information, such as institutional policies, learning design, pedagogical guidelines, historical analytical insights. It serves as an internal source of contextual knowledge, supporting AI-driven interpretation and recommendations.

Data Importer. This importer collects and prepares data from external data sources, e.g. external datasets, institutional repositories, or publicly available IT systems. These data may include machine learning models, educational datasets, or contextual information that can enhance predictive analytics and improve AI-based services.

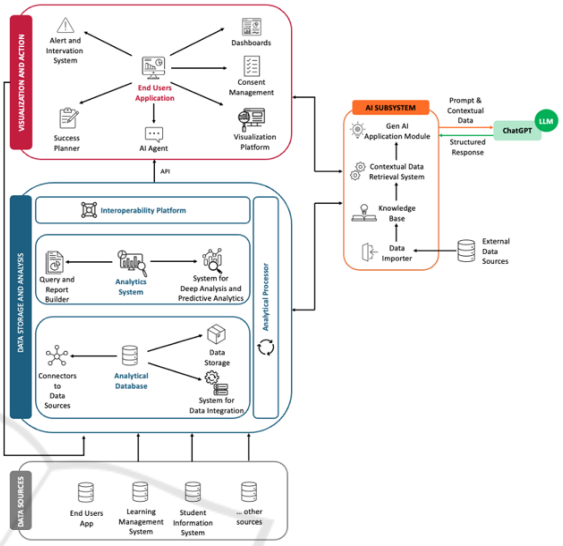


Figure 2: Components of the AI-driven LA infrastructure model.

4.3 Validation of the Model

The proposed conceptual model of LA infrastructure was validated through expert evaluation using 15 Likert-scale statements covering model structure, data coverage, ethical considerations, and implementation feasibility. Overall, the results indicate a high level of agreement, with all statements receiving mean scores above 4.0.

The highest-rated aspects relate to the clarity and structure of the model, particularly the three-layer architecture (M = 4.77), which was perceived as logical and intuitive. The model was also positively evaluated in terms of data comprehensiveness (M = 4.66) and inclusion of components supporting data processing and analysis (M ≈ 4.44–4.55).

Slightly lower, but still positive ratings were observed for aspects related to intervention mechanisms and ethical components (M = 4.00), suggesting opportunities for further refinement, particularly in terms of practical implementation and explicit support for data protection. Implementation feasibility and flexibility were rated positively (M ≈ 4.11–4.22), indicating that the model is perceived as

applicable in higher education contexts, although further real-world validation may be beneficial.

Qualitative feedback highlighted the model's clarity, layered structure, and ability to integrate diverse data sources as key strengths. Experts identified potential challenges related to data availability, system integration, infrastructure readiness, and privacy compliance, as well as organizational barriers such as limited expertise and institutional resistance.

Overall, the findings suggest that the proposed model represents a comprehensive and well-structured framework, while also indicating areas for further development, particularly regarding security considerations, ethical implementation, and practical deployment guidelines.

5 CONCLUSIONS

This paper presents an AI-driven conceptual model of learning analytics (LA) infrastructure, targeted at higher education institutions. While the existing literature on the development of such conceptual models is limited, this paper considers the available research results, complemented with practical expert insights. The proposed infrastructure model consists of four components: data sources, data storage, visualisation and action, and AI subsystem. In the first round of validation, the model has been validated through an expert survey, confirming the value of the model. This approach enables IT specialists to combine various solutions in each of the components, enabling them to create flexible infrastructure, capable of adapting to future needs. Future work shall be focused on the interoperability aspect of different data sources to ensure sound and meaningful LA.

Future work will focus on interoperability and real-world validation, as well as extending the AI subsystem with an optimization layer for orchestrating data collection and preparation, enabling context-aware AI processing. Further research will address explainable predictive models, improved contextual data retrieval, and the role of AI agents in supporting personalized and adaptive interaction.

ACKNOWLEDGEMENTS

This paper is part of the Erasmus+ SAILeR project (2025-1-HR01-KA220-HED-000355135) financed by the European Union and the TRUELA project (IP-2022-10-2854) financed by the Croatian Science Foundation.

REFERENCES

The ScR papers are marked with an asterisk.*

*Alalawi, K., Athauda, R., Chiong, R., & Renner, I. (2025). Evaluating the student performance prediction and action framework through a learning analytics intervention study. *Education and Information Technologies*, *30*(3), 2887–2916. https://doi.org/10.1007/s10639-024-12923-5

Alam, A. (2021). Should Robots Replace Teachers? Mobilisation of AI and Learning Analytics in Education. *2021 International Conference on Advances in Computing, Communication, and Control (ICAC3)*, 1–12. https://doi.org/10.1109/ICAC353642.2021.9697300

*Ameloot, E., Rotsaert, T., & Schellens, T. (2022). The supporting role of learning analytics for a blended learning environment: Exploring students' perceptions and the impact on relatedness. *Journal of Computer Assisted Learning*, *38*(1), 90–102. https://doi.org/10.1111/jcal.12593

Badari, J., Redjep, N., Grabar, D., & Vondra, P. (2024). Conceptual infrastructure model to utilize learning analytics in higher education institutions blended learning environment. In Proceedings of the ICERI2024 Conference (pp. 4321–4331). https://doi.org/10.21125/iceri.2024.1087

*Baneres, D., Guerrero-Roldán, A. E., Rodríguez-González, M. E., & Karadeniz, A. (2021). A Predictive Analytics Infrastructure to Support a Trustworthy Early Warning System. *Applied Sciences*, *11*(13), 5781. https://doi.org/10.3390/app11135781

*Bergdahl, N., Nouri, J., Karunaratne, T., Afzaal, M., & Saqr, M. (2020). Learning Analytics for Blended Learning: A Systematic Review of Theory, Methodology, and Ethical Considerations. *International Journal of Learning Analytics and Artificial Intelligence for Education (iJAI)*, *2*(2), 46. https://doi.org/10.3991/ijai.v2i2.17887

*Caspari-Sadeghi, S. (2023). Learning assessment in the age of big data: Learning analytics in higher education. *Cogent Education*, *10*(1), 2162697. https://doi.org/10.1080/2331186X.2022.2162697

Chatti, M. A., & Muslim, A. (2019). The PERLA Framework: Blending Personalization and Learning Analytics. *The International Review of Research in Open and Distributed Learning*, *20*(1). https://doi.org/10.19173/irrodl.v20i1.3936

Chatti, M. A., Muslim, A., Guesmi, M., Richtscheid, F., Nasimi, D., Shahin, A., & Damera, R. (2020). *How to Design Effective Learning Analytics Indicators? A Human-Centered Design Approach* (pp. 303–317). https://doi.org/10.1007/978-3-030-57717-9_22

*Ciordas-Hertel, G. P., Schneider, J., & Drachsler, H. (2020). Which Strategies are Used in the Design of Technical LA Infrastructure?: A Qualitative Interview

Study. *2020 IEEE Global Engineering Education Conference (EDUCON)*, 96–105. https://doi.org/10.1109/EDUCON45650.2020.9125363

Ciordas-Hertel, G. P., Schneider, J., Ternier, S., & Drachsler, H. (2019). Adopting trust in learning analytics infrastructure: A structured literature review. *Journal of Universal Computer Science*, *25*(13), 1668–1686.

*Clark, J.-A., Liu, Y., & Isaias, P. (2020). Critical success factors for implementing learning analytics in higher education: A mixed-method inquiry. *Australasian Journal of Educational Technology*, *36*(6), 89–106. https://doi.org/10.14742/ajet.6164

Deming, W. E. (1986). *Out of the Crisis*. MIT Press.

Divjak, B., & Maretić, M. (2017). Learning Analytics for Peer-assessment: (Dis)advantages, Reliability and Implementation. *Journal of Information and Organizational Sciences*, *41*(1), 21–34. https://doi.org/10.31341/jios.41.1.2

Divjak, B., Svetec, B., & Horvat, D. (2024). How can valid and reliable automatic formative assessment predict the acquisition of learning outcomes? *Journal of Computer Assisted Learning*, (January), 1–17. https://doi.org/10.1111/jcal.12953

Divjak, B., Svetec, B., Horvat, D., & Kadoić, N. (2022). Assessment validity and learning analytics as prerequisites for ensuring student-centred learning design. *British Journal of Educational Technology*, bjet.13290. https://doi.org/10.1111/bjet.13290

*Drlik, M., Skalka, J., Svec, P., & Kapusta, J. (2018). Proposal of Learning Analytics Architecture Integration into University IT Infrastructure. *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*, 1–6. https://doi.org/10.1109/ICAICT.2018.8747068

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., & Vuorikari, R. (2016). *Research evidence on the use of learning analytics: Implications for education Policy*. European Commission, Joint Research Centre. https://doi.org/10.2791/955210

Gaeta, A., Orciuoli , F., Pascuzzo, A., & Peduto, A. (2025). Enhancing traditional ITS architectures with large language models for generating motivational feedback. Computers and Education: Artificial Intelligence , 9, 100433.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gray, G., & Bergner, Y. (2022). A Practitioner's Guide to Measurement in Learning Analytics: Decisions, Opportunities, and Challenges. In *The Handbook of Learning Analytics* (pp. 20–28). SOLAR. https://doi.org/10.18608/hla22.002

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, *68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 31–40. https://doi.org/10.1145/3170358.3170421

Kaliisa, R., Gillespie, A., Herodotou, C., Kluge, A., & Rienties, B. (2021). Teachers' Perspectives on the Promises, Needs and Challenges of Learning Analytics Dashboards: Insights from Institutions Offering Blended and Distance Learning. In M. Sahin & D. Ifenthaler (Eds.), *Visualizations and Dashboards for Learning Analytics* (pp. 351–370). Springer International Publishing. https://doi.org/10.1007/978-3-030-81222-5_16

*Kaliisa, R., Jivet, I., & Prinsloo, P. (2023). A checklist to guide the planning, designing, implementation, and evaluation of learning analytics dashboards. *International Journal of Educational Technology in Higher Education*, *20*(1), 28. https://doi.org/10.1186/s41239-023-00394-6

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*. https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Macfadyen, L. P. (2022). Institutional Implementation of Learning Analytics—Current State, Challenges, and Guiding Frameworks. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), *The Handbook of Learning Analytics* (2nd ed., pp. 173–186). SOLAR. https://doi.org/10.18608/hla22.017

Newman, M., & Gough, D. (2020). Systematic Reviews in Educational Research: Methodology, Perspectives and Application. In O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, & K. Buntins (Eds.), *Systematic Reviews in Educational Research* (pp. 3–22). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27602-7_1

Nguyen, T. Tk., Nguyen, M. T., & Tran, H. T. (2023). Artificial intelligent based teaching and learning approaches: A comprehensive review. *International Journal of Evaluation and Research in Education (IJERE)*, *12*(4), 2387. https://doi.org/10.11591/ijere.v12i4.26623

Ochoa, X. (2022). Multimodal Learning Analytics: Rationale, Process, Examples, and Direction. In *The Handbook of Learning Analytics* (pp. 54–65). SOLAR. https://doi.org/10.18608/hla22.006

Ouhaichi, H., Spikol, D., & Vogel, B. (2023). Research trends in multimodal learning analytics: A systematic mapping study. *Computers and Education: Artificial Intelligence*, *4*, 100136. https://doi.org/10.1016/j.caeai.2023.100136

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to

improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-022-00372-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthcare, 13*(3), 141–146. https://doi.org/10.1097/XEB.0000000000000050

*Perez-Sanagustin, M., Hilliger, I., Maldonado-Mahauad, J., & Perez-Alvarez, R. (2022). Building Institutional Capacity for Learning Analytics: Top-Down & Bottom-Up Initiatives. *IEEE Revista Iberoamericana de Tecnologias Del Aprendizaje*, *17*(3), 281–289. https://doi.org/10.1109/RITA.2022.3191413

*Pinto, A. D. S., Longhi, M. T., & Behar, P. A. (2025). Competências digitais discentes para o uso de inteligência artificial no ensino superior a distância: Uma revisão sistemática da literatura. *RDBCI: Revista Digital de Biblioteconomia e Ciência Da Informação*, *24*, e026013. https://doi.org/10.20396/rdbci.v24i00.8679434

Reardon, E., Kumar, V., & Revelle, G. (2022). *Game learning analytics*. https://doi.org/10.18608/hla22.015

Sajja, R., Sermet, Y., Cwiertny, D., & Demir, I. (2023). *Integrating AI and Learning Analytics for Data-Driven Pedagogical Decisions and Personalized Interventions in Education* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2312.09548

*Salonen, A. O., Tapani, A., & Suhonen, S. (2021). Student Online Activity in Blended Learning: A Learning Analytics Perspective of Professional Teacher Education Studies in Finland. *Sage Open*, *11*(4), 21582440211056612. https://doi.org/10.1177/21582440211056612

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, *78*, 397–407. https://doi.org/10.1016/j.chb.2017.06.030

Sclater, N. (2014). *Learning analytics—The current state of play in UK higher and further education*. https://repository.jisc.ac.uk/5657/1/Learning_analytics_report.pdf

Sclater, N. (2016). Developing a Code of Practice for Learning Analytics. *Journal of Learning Analytics*, *3*(1). https://doi.org/10.18608/jla.2016.31.3

*Shacklock, X. (2016). *From Bricks to Clicks*. Policy Connect. https://policyconnect.org.uk/wp-content/uploads/2020/09/frombrickstoclicks-hecreportforweb.pdf

Siemens, G., Dawson, S., & Lynch, G. (2013). *Improving the Quality and Productivity of the Higher Education Sector—Policy and Strategy for Systems-Level Deployment of Learning Analytics*. Society for Learning Analytics Research (SoLAR). https://solaresearch.org/wp-content/uploads/2017/06/SoLAR_Report_2014.pdf

Society for Learning Analytics Research. (n.d.). Retrieved October 24, 2022, from https://www.solaresearch.org/about/what-is-learning-analytics/

Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: A tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, *19*(1), 12. https://doi.org/10.1186/s41239-021-00313-7

Svetec, B., & Divjak, B. (2025). Trustworthy Learning Analytics for Smart Learning Ecosystems. *InteractionDesign&Architecture(s)*.

Taylor, S., & Munguia, P. (2018). Towards a data archiving solution for learning analytics. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 260–264. https://doi.org/10.1145/3170358.3170415

Tsai, Y.-S., Moreno-Marcos, P. M., Jivet, I., Scheffel, M., Tammets, K., Kollom, K., & Gašević, D. (2018). The SHEILA Framework: Informing Institutional Strategies and Policy Processes of Learning Analytics. *Journal of Learning Analytics*, *5*(3). https://doi.org/10.18608/jla.2018.53.2

Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2021). More Than Figures on Your Laptop: (Dis)trustful Implementation of Learning Analytics. *Journal of Learning Analytics*, *8*(3), 81–100. https://doi.org/10.18608/jla.2021.7379

Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., & De Laet, T. (2020). Learning analytics dashboards: The past, the present and the future. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 35–40. https://doi.org/10.1145/3375462.3375504

Wieringa, R. J. (2014). *Design Science Methodology for Information Systems and Software Engineering* (1st ed. 2014). Springer Berlin Heidelberg : Imprint: Springer. https://doi.org/10.1007/978-3-662-43839-8