ECLAIR: A Centralized AI-Powered Recommendations System in a Multi-Node EXAIT System

Isanka WIJERATHNE[a*], Brendan FLANAGAN[b], Yiling DAI[a] & Hiroaki OGATA[a]
[a]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
[b]*Center for Innovative Research and Education in Data Science, Kyoto University, Japan*
*Wijerathne.isanka.6z@kyoto-u.ac.jp

Abstract: Educational recommender systems are increasingly becoming a core feature of modern educational systems. Often the recommender component of a system is tightly integrated, or might be remotely located without accessing data from other local systems. This paper proposes a framework called ECLAIR in which local educational systems can work and share data with a global recommender system that spans multiple educational institutions. In particular, an AI-driven recommendation system is intricately integrated within a multi-node learning management system. Situated at the intersection of large-scale data analysis and personalized education, ECLAIR efficiently processes heterogeneous data from diverse LMS, while ensuring data security and privacy. The proposed ECLAIR's architecture, data pipeline, and processing mechanisms are explored in detail, focusing on its seamless integration with the existing infrastructure. By leveraging MongoDB change streams and relational databases, ECLAIR guarantees real-time data synchronization, secure storage, and efficient processing. Its unique Ingestor tool transforms selected xAPI data points into a relational table format, bolstering system functionality. The successful launch of ECLAIR serves as a testament to the potential of AI in enhancing personalized learning experiences, improving data security, and bolstering system efficiency. Nevertheless, the paper emphasizes the need for ongoing research, specifically concerning privacy-preserving mechanisms and efficient management of data heterogeneity. It demonstrates ECLAIR's pivotal role in the rapidly evolving landscape of eLearning, and its potential for future advancements, scalability, and adaptability, setting a new precedent for AI-powered eLearning recommendation systems.

Keywords: Artificial intelligence (AI), AI-Powered recommender system, Explainable AI, e-learning, ECLAIR, EXAIT

1. Introduction

In an era marked by technological integration, education has witnessed considerable transformations. The emergent realms of Learning Management Systems (LMS), eLearning platforms, and Artificial Intelligence (AI)-enhanced recommendation systems have significantly augmented educational experiences (Aleven et al., 2016). These advancements demonstrate potential in personalizing learning experiences, optimizing data security, and augmenting system efficiency. However, despite these gains, obstacles persist, such as managing data heterogeneity, preserving data privacy, and maintaining system scalability (Boitshwarelo, 2011). This presents a continued need for innovative approaches.

In this paper we propose the EXAIT CentraL Artificial Intelligence Recommender (ECLAIR), a centralized AI-powered recommendation system that functions within a multi-node EXAIT (Educational eXplainable AI Tools) system (Flanagan et al., 2021). ECLAIR is designed to tackle the key challenges associated with AI recommender systems in eLearning, offering a novel pathway towards personalized learning.

The ECLAIR system stands at the intersection of large-scale data analysis and personalized education. It can manage and analyze a vast array of data across various LMS platforms, furnishing personalized, efficient, and precise recommendations to learners. ECLAIR prioritizes data security and privacy, crucial considerations in the global educational sector (Polonetsky & Tene, 2013; Wijerathne et al., 2022).

The paper is methodically structured to provide an extensive understanding of the ECLAIR system, commencing with a review of related works in the field of AI-enhanced recommendation systems. This is succeeded by a comprehensive system overview, an explanation of the data pipeline and processing mechanisms, and a detailed account of the implementation and integration of ECLAIR. Throughout the paper, we underscore how ECLAIR surpasses traditional system challenges and anticipates future advancements in the rapidly evolving domain of eLearning.

2. **Related Works**

The intersection of Learning Management Systems (LMS), eLearning, and Artificial Intelligence (AI) has been a focal point of research in recent years (Ouyang et al., 2022), particularly concerning the development of recommendation systems. One of the earliest strides in this space was the adoption of collaborative filtering and content-based recommendation algorithms (Koren, 2008; Lu et al., 2015; Marlin, 2003). For instance, implemented a personalized recommendation system in eLearning platforms using a content-Lalitha et al. (Lalitha & Sreeja, 2020) based approach, which considered the learners' profiles and their learning styles. Despite the initial success, traditional methods like these often-encountered challenges such as the cold start problem and scalability issues, particularly when dealing with a vast number of users and contents (Bell & Koren, 2007; Lam et al., 2008; Lika et al., 2014).

To overcome these hurdles, subsequent research began to incorporate machine learning techniques into the design of recommendation systems. A study conducted by Huang et al. (Huang et al., 2002) exemplifies this transition. They proposed a hybrid recommendation system that combined the strengths of collaborative filtering and deep learning techniques. The system was built on an Autoencoder, a neural network model that improved the accuracy of recommendations by learning complex, non-linear user-item interactions. This study represented a substantial step forward from traditional methods, demonstrating the potential of machine learning in enhancing recommendation systems (Bai et al., 2019; Beel et al., 2013).

Numerous educational institutions traditionally exhibit a reluctance to disclose student learning data to external entities, a tendency grounded in concerns regarding privacy and data security. However, the functionality and effectiveness of recommender systems are inextricably linked to their access to extensive datasets, thereby creating a conundrum. As such, there is a pressing need to strike a delicate equilibrium between data utility, that is, the optimization of recommender systems through comprehensive data access, and the preservation of data privacy, ensuring the protection of sensitive information. This balance is paramount to harness the full potential of such systems while still adhering to ethical guidelines and legal regulations (Kobsa, 2007; Polonetsky & Tene, 2013). Ocheja et al's work introduces the Blockchain of Learning Logs (BOLL), a platform that comprehensively records learning pathways and achievements beyond the scope of traditional transcripts. This system collects data from distributed learning systems and addresses the "cold-start" issue in learning data analytic platforms by enabling the secure and verifiable transfer of learning records across institutions. Crucially, BOLL maintains a delicate balance between data privacy, sensitivity, and technology, underscoring its commitment to respecting individual privacy (Ocheja et al., 2019).

In the preliminary phase, the proposed artificial intelligence (AI)-driven recommendation system has been utilized for teaching English as a Foreign Language (EFL) to high school students (Takii et al., 2022). Furthermore, an AI-enhanced recommendation system with self-explanatory capabilities has been used for mathematics instruction (Dai et al., 2022; Flanagan et al., 2021).

These advancements underscore the potential of centralized AI-powered recommendation systems in enhancing personalized learning, improving data security, and increasing system efficiency. However, they also highlight the need for additional research in areas like privacy-preserving mechanisms and managing data heterogeneity.

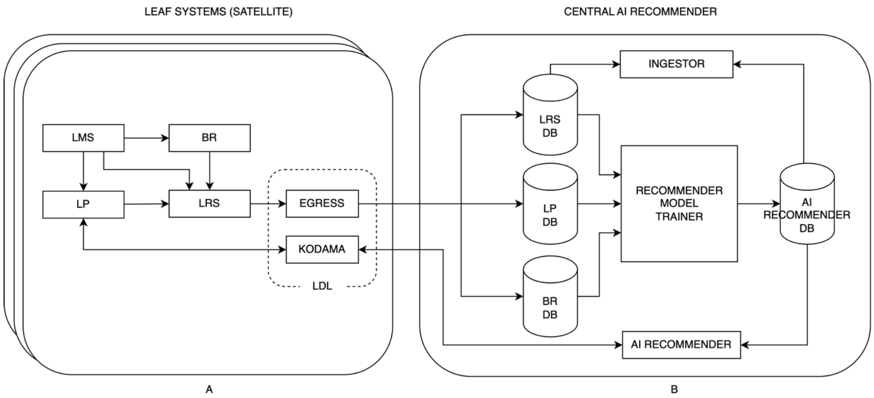

Figure 1: Overview of the proposed system

3. **System Overview**

The implementation of AI recommender systems, which inherently require substantial amounts of data for efficient operation, is thereby affected. It is pertinent to understand that the quality of these systems is directly proportional to the quantity of data available, hence a larger dataset is always advantageous.

In response to this challenge, we propose a system that amalgamates the requisite data from these distributed systems into a singular, centralized system as shown in Figure 1. The proposed system contains two main parts: Sets of multiple LEAF systems (Left side) and the central AI recommender (right side), with API connections facilitating communication between the two systems. This system then processes the accumulated data through the AI recommender system. The processed information, in turn, is channeled back to the respective LEAF system, thereby ensuring a feedback mechanism that enhances the overall learning experience.

3.1 LEAF System

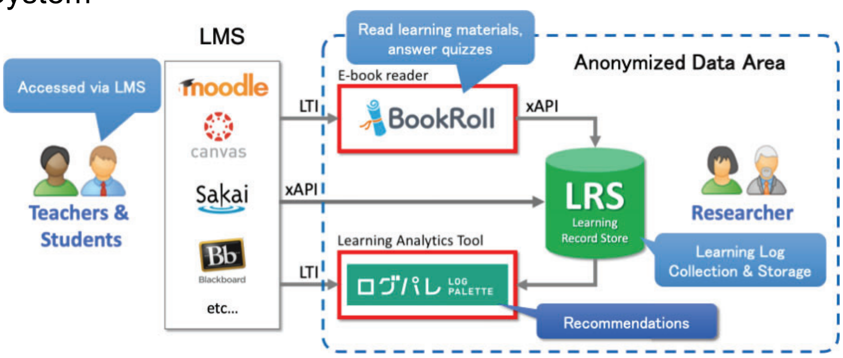

Figure 2: Overview of the LEAF system

The Learning Evidence Analytics Framework (LEAF) infrastructure shown in Figure 2(Flanagan & Ogata, 2018) is an intricate composite of several unique components, prominently including an LMS, a digital learning material reading systems Bookroll, a learning analysis tool Log palette, and a Learning Record Store (LRS). LRS, is an open-source tool for managing learning data, leveraging the scalable and high-performance capabilities of MongoDB for data storage and retrieval, and supporting data analysis and insights generation. The LMS, as implied by its nomenclature, orchestrates the planning, execution, and evaluation of the comprehensive learning process. Meanwhile, Bookroll functions as a versatile e-book reader. Intriguingly, all these systems emit Experience API (xAPI) data, which are subsequently stored within the LRS. The proposed system has multiple LEAF nodes connected to the Central AI Recommender.

3.2 Central AI Recommender

The Central AI Recommender as shown in Figure 1 is a complex construct consisting of numerous key components such as the Recommender Model Trainer, the LRS Database, the LP Database, the BR Database, the AI Recommender Database, the Ingestor, and the AI Recommender application. Data harvested from the distributed LEAF systems through the Egress API is consolidated within the LRS, LP, and BR databases. This stored data is subsequently leveraged by the Recommender Model Trainer to train the model, which is then deposited in the AI Recommender Database. The Ingestor serves as a data transformer, converting LRS MongoDB data into relational data, which is then stored in the AI Recommender Database. Finally, the AI Recommender application utilizes this data to provide recommendations through the Kodama API.

4. Data Pipeline and Processing

The LEAF system encompasses two types of data: relational and non-relational. Relational data are managed through the MySQL database and are primarily utilized within the LMS, Bookroll, and Log Palette. In contrast, non-relational data, mainly characterized by xAPI statements, are generated from every interaction within these systems and are subsequently stored in MongoDB. The proposed system, moreover, features multiple LEAF nodes distributed across various educational institutions.

The Experience API (Tin Can API[1]), initially introduced in 2013 (Gonzalez & Churchill, 2019), serves as a crucial component of learning analytics. It offers a platform-agnostic formalism for capturing events that transpire throughout any learning experience. The proposed system capitalizes on the capabilities of xAPI to record all learner-system interactions. The generated xAPI statements are committed to a Learning Record Store (LRS), which is a key element in the infrastructure for storing and retrieving learning records. This combination of relational and non-relational data storage mechanisms, coupled with the adoption of xAPI for event tracking, ensures comprehensive data coverage and fosters the potential for rich analytical insights.

For the purpose of enhancing the ECLAIR system, it is imperative that pertinent data is obtained from various LEAF data stores, which will subsequently be utilized in training its models. This process is essential to avoid congestion and to maintain a keen focus on the microservices architecture. In light of this, a system has been devised that incorporates the LEAF Data Layer (LDL). This system comprises two Application Programming Interfaces (API) endpoints, named "Egress" and "Kodama", with the later meaning echo in Japanese. These designated endpoints are instrumental in facilitating seamless data exchange between the LEAF systems and the ECLAIR.

Consequently, the ECLAIR system has been provisioned with the ability to subscribe to the LRS MongoDB databases of the LEAF systems. This subscription operation is actualized by leveraging MongoDB change streams, a feature that facilitates real-time data processing. This mechanism empowers the ECLAIR system to garner real-time data from the LRS MongoDB, and in turn, replicate this data into its own LRS MongoDB, thereby ensuring a continuous flow and synchronization of data between systems.

Moreover, the ECLAIR system is designed to extract relational data from the LEAF systems through the LDL EGRESS API. The extracted data is then transferred and securely stored within the ECLAIR's own data repositories. This methodical process ensures that the ÉCLAIR system continuously has access to updated and relevant data necessary for its

[1] https://xapi.com/

optimal operation and the achievement of its intended objectives.

The ECLAIR necessitates relational data for the comprehensive functionality and features it offers. To achieve this, selected data points from the replicated LRS MongoDB data are transformed into a relational table format using the Ingestor tool. The Ingestor system is comprised of two main components: real-time data conversion and backfilling data. In scenarios involving system failure or the incorporation of novel xAPI data points, the backfilling application can be manually executed to facilitate the propagation of relational data from the LRS MongoDB.

5. Implementation and Integration

During the implementation phase, our focus was on integrating the ECLAIR system with the existing LEAF infrastructure. This integration, which was aimed at facilitating seamless data transmission, synchronization, and processing while also guaranteeing data security and privacy, necessitated meticulous planning and stringent execution.

The ECLAIR system is based on a microservices design, which allows for versatility, scalability, and simplicity of updating. It was created and built mostly with Python, a versatile and widely used programming language for AI applications, and the Scikit-Learn[2] module, which provides a wide range of machine learning techniques for the recommendation.

Setting up the Egress and Kodama endpoints for data interchange between the LEAF system and ECLAIR was the first step in its implementation. These endpoints were built with the Flask framework, which allows for the creation of RESTful APIs. Following that, the ECLAIR system was set up to use MongoDB change streams to subscribe to the LEAF system's LRS MongoDB databases. This subscription facilitates real-time data processing and synchronization between systems. The relational data obtained from the LEAF systems through the LDL Egress API was incorporated into ECLAIR's own data repositories. We used PostgreSQL, a robust and secure relational database system, to store this data. Security measures were put in place to protect the data, including encryption at rest and in transit, regular backups, and strict access controls.

The Ingestor tool was then developed to convert the selected xAPI data points from the replicated LRS MongoDB into a relational table structure. The technology is comprising two major components: real-time data synchronization and backfilling. The real-time data synchronization component is written in Python. On the other hand, the backfilling component, responsible for filling historical data gaps, was developed using Golang due to its efficiency and performance advantages in handling large datasets. Together, these components make Ingestor a versatile tool for data conversion and synchronization.

Upon the successful establishment of the data pipeline, the commencement of the machine learning model training phase for the ECLAIR system was initiated. This intensive process was undertaken on high-performance computing systems, owing to the considerable amount of data involved. One representative example of the machine learning models employed within the ECLAIR system is the Bayesian Knowledge Tracing (BKT), a Python-based implementation of the BKT model. To ensure up-to-date integration of new learning log data, the model is refreshed at regular intervals, specifically every 30 minutes if new data has been detected by the system (Badrinath et al., 2021). Each model that is trained is then stored in the Recommender DB along with a timestamp showing when the model could first be used to give recommends. As it is continuously being updated, historical version snapshots of the model are stored and can be used by researchers to replay or recreate recommendations of particular students at a point in time. This can be useful also if a research needs to investigate in detail why and how recommendations were generated previously by the model so that explanations can be checked and validated. It also allows developers and researchers to examine problems that have occurred in the past.

Finally, when a learner uses the recommendation panel on a satellite LEAF system, this sends a request via Kodama to the recommender module in the ECLAIR system.

[2] https://scikit-learn.org/

Personalized recommendations and explanations are then generated based on the data and models stored in the central database, relayed back to the satellite LEAF system and displayed to the learner. The feedback mechanism ensured that the recommendations were continuously improved based on the learners' interactions with the system. Also, as the recommendation model is centrally located, it enables researchers and developers to update and deploy new features and targeted interventions for evaluation with efficiency, instead of having to update multiple satellite LEAF systems.

In summary, the implementation and integration of the ECLAIR system into the LEAF infrastructure was a complex yet rewarding process. The successful integration has the potential to transform the learning experience by providing personalized, efficient, and engaging AI-driven recommendations. The modularity of the system also offers opportunities for future enhancements and scalability, as the need arises.

6. Conclusion and Future Work

The development, implementation, and integration of the ECLAIR system with the existing LEAF infrastructure has demonstrated the power of combining AI, machine learning, and data processing technologies to transform the learning experience. The architecture based on microservices provides an efficient and scalable solution, while the utilization of high-performance computing systems and advanced machine learning models like BKT facilitates the processing of vast amounts of data to generate personalized learning recommendations.

The implementation of ECLAIR has successfully tackled several challenges, including seamless data transmission, real-time data synchronization, and stringent security measures for data privacy. Furthermore, it has underscored the effectiveness of Python for developing AI applications, the efficiency of Flask framework in creating RESTful APIs, and the advantage of Golang in handling large datasets.

The use of the Ingestor tool for real-time data synchronization and backfilling of historical data, along with the frequent refreshing of the machine learning model, ensures that the learning recommendations are always based on the most current data. The feedback mechanism integrated into the LEAF system allows for continuous improvement in the recommendation quality, ensuring a more engaging and personalized learning experience.

Future work on the ECLAIR system could further extend its capabilities and scalability. Potential directions include enhancing the machine learning models to handle even more complex learning scenarios and incorporating additional forms of learning data, such as biometric or psychometric data, to refine the personalization of learning recommendations. There may also be potential for integration with other LMSs or learning platforms to broaden its impact.

In summary, the ECLAIR system, through its innovative use of AI and machine learning technologies, promises to be a transformative tool in the landscape of personalized learning. Its successful integration with the LEAF system demonstrates its potential, not just for improving the learner's experience, but also for driving forward the field of AI-powered education as a whole. The flexible and scalable architecture of the system also ensures that it is well-suited for future enhancements and integrations, potentially creating an even greater impact on the education sector.

Acknowledgments

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B) JP20H01722 and JP23H01001, (Exploratory) JP21K19824, (A) JP23H00505, and NEDO JPNP20006.

References

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. *Handbook of research on learning and instruction*, *2*, 522-560.

Badrinath, A., Wang, F., & Pardos, Z. (2021). pybkt: An accessible python library of bayesian knowledge tracing models. *arXiv preprint arXiv:2105.00385*.
Bai, X., Wang, M., Lee, I., Yang, Z., Kong, X., & Xia, F. (2019). Scientific paper recommendation: A survey. *Ieee Access*, *7*, 9324-9339.
Beel, J., Langer, S., Genzmehr, M., Gipp, B., Breitinger, C., & Nürnberger, A. (2013). Research paper recommender system evaluation: a quantitative literature survey. Proceedings of the international workshop on reproducibility and replication in recommender systems evaluation,
Bell, R. M., & Koren, Y. (2007). Scalable collaborative filtering with jointly derived neighborhood interpolation weights. Seventh IEEE international conference on data mining (ICDM 2007),
Boitshwarelo, B. (2011). Proposing an integrated research framework for connectivism: Utilising theoretical synergies. *International Review of Research in Open and Distributed Learning*, *12*(3), 161-179.
Dai, Y., Flanagan, B., Takami, K., & Ogata, H. (2022). Design of a User-Interpretable Math Quiz Recommender System for Japanese High School Students. Proceedings of the 4th Workshop on Predicting Performance Based on the Analysis of Reading Behavior.
Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning*, *10*(4), 469-484.
Flanagan, B., Takami, K., Takii, K., Dai, Y., Majumdar, R., & Ogata, H. (2021). EXAIT: A Symbiotic Explanation Education System. 29th International Conference on Computers in Education Conference Proceedings, pp. 404-409.
Huang, Z., Chung, W., Ong, T.-H., & Chen, H. (2002). A graph-based recommender system for digital library. Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries,
Kobsa, A. (2007). Privacy-enhanced personalization. *Communications of the ACM*, *50*(8), 24-33.
Koren, Y. (2008, August). Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 426-434).
Lalitha, T., & Sreeja, P. (2020). Personalised self-directed learning recommendation system. *Procedia Computer Science*, *171*, 583-592.
Lam, X. N., Vu, T., Le, T. D., & Duong, A. D. (2008, January). Addressing cold-start problem in recommendation systems. In Proceedings of the 2nd international conference on Ubiquitous information management and communication (pp. 208-211).
Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert systems with applications*, *41*(4), 2065-2073.
Lu, Z., Dou, Z., Lian, J., Xie, X., & Yang, Q. (2015, January). Content-based collaborative filtering for news topic recommendation. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (pp. 217-223).
Marlin, B. M. (2003). Modeling user rating profiles for collaborative filtering. *Advances in neural information processing systems*, *16*.
Ocheja, P., Flanagan, B., Ueda, H., & Ogata, H. (2019). Managing lifelong learning records through blockchain. *Research and Practice in Technology Enhanced Learning*, *14*(1), 1-19.
Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, *27*(6), 7893-7925.
Polonetsky, J., & Tene, O. (2013). Privacy and big data: making ends meet. *Stan. L. Rev. Online*, *66*, 25.
Takii, K., Flanagan, B., Li, H., Yang, Y., & Ogata, H. (2022). Explainable English Material Recommendation Using an Information Retrieval Technique for EFL Learning. 30th International Conference on Computers in Education Conference Proceedings, pp. 561-570.
Wijerathne, I., Masako, O., Morimura, Y., & Sakai, H. (2022). *Development of MOOC Data Management Portal for Instructors and Production Team* 21st Forum on Information Technology (FIT2022) (pp. 273-280).