

An AI Companion for Self-Regulated Learning: Integrating Conversational AI, Predictive Analytics, and Explainability

Hana Ivančić, Tomislav Jagušć

University of Zagreb

Faculty of Electrical Engineering and Computing

Unska 3, 10000 Zagreb, Croatia

{hana.ivancic, tomlav.jagust}@fer.unizg.hr

Abstract. Artificial intelligence in higher education often focuses on conversational assistants rather than broader learning processes. This paper presents ongoing research towards FER AI Companion, a self-regulated learning (SRL) platform integrating retrieval-augmented generation, predictive analytics, and explainable AI, with reflection and gamification planned as extensions. The system follows Zimmerman's SRL cycle and is evaluated using historical data from five first-year courses (3,958 students). Tree-based models achieve MAE of approximately 3.0 ISVU points at mid-semester and 1.8–2.0 at the late checkpoint, outperforming a training-mean baseline (11.5). Inter-model disagreement also provides a relative error signal. No user study has been conducted.

Keywords. predictive learning analytics, self-regulated learning, retrieval-augmented generation, explainable artificial intelligence, gamification, higher education

1 Introduction

Artificial intelligence is rapidly transforming higher education by enabling personalized learning experiences and providing students with continuous access to academic support (Zawacki-Richter, 2019). Recent advances in large language models (LLMs) have led to the widespread adoption of AI-powered assistants capable of answering questions, explaining course concepts, and assisting with administrative tasks (McGrath et al., 2025; Kasneci et al., 2023). While such systems improve information accessibility, they typically function as conversational interfaces and provide limited support for broader learning processes such as planning, monitoring, and reflecting on learning (Lan & Zhou, 2025).

Self-regulated learning (SRL) emphasizes that successful students actively set learning goals, monitor their progress, and evaluate their performance through a continuous cycle of forethought, performance, and self-reflection (Zimmerman, 2002). Supporting these

phases requires more than question answering alone. Predictive learning analytics can provide students with early feedback regarding expected academic outcomes, explainable artificial intelligence (XAI) can clarify why predictions were made, and gamification can encourage sustained engagement with learning activities (Khosravi et al., 2022). Integrating these complementary technologies within a unified system has the potential to provide substantially richer educational support than standalone AI chatbots.

Despite the rapid emergence of AI-powered educational tools, current systems are typically optimized for isolated functions (most commonly question answering or administrative support) rather than for sustained learning process orchestration (McGrath et al., 2025). This creates a structural gap between what learning science identifies as essential for student success and what AI systems currently deliver in practice. In particular, self-regulated learning requires continuous interaction between goal setting, performance monitoring, adaptive feedback, and reflection (Panadero, 2017), yet such integration remains uncommon in the literature (Khosravi et al., 2022). This work addresses this gap by proposing a unified design approach in which predictive analytics, explainable AI, conversational assistance, and behavioral scaffolding are not treated as independent modules, but as coordinated components of an SRL-aligned feedback loop. In doing so, the system moves beyond the paradigm of AI as an information provider toward AI as an active learning process companion.

This paper presents the first phase of the development of FER AI Companion, an AI-supported web application designed to support first-year university students within an SRL framework. The proposed architecture combines retrieval-augmented generation (RAG) (Lewis et al., 2021) for grounded conversational assistance, machine learning models for grade prediction, AI-generated personalized study plans, automated reflection prompts, and gamification mechanisms within a single platform structured around Zimmerman's SRL model (Zimmerman, 2002). Prediction transparency is achieved through SHAP (SHapley Additive exPlanations), an explainable AI

method that identifies the features contributing most to individual predictions. The contribution of this work is twofold. First, it presents an SRL-aligned architecture that connects predictive analytics and prediction explanations with grounded conversational assistance, while explicitly identifying reflection and gamification as future extensions. Second, it provides an initial empirical evaluation of the predictive component on a chronological held-out cohort, including baseline comparisons and analysis of inter-model disagreement. The paper therefore contributes an initial technical and empirical foundation for an SRL-oriented learning companion rather than evidence of educational effectiveness.

This paper addresses the following research questions: (RQ1) How accurately can final course outcomes be predicted at different semester checkpoints using machine learning models trained on historical student data? (RQ2) Does disagreement among predictive models provide a useful indication of prediction error? (RQ3) How are the implemented and planned system functions mapped to theoretically defined SRL processes? (RQ4) What components of the proposed platform are implemented and technically evaluated at the current stage of development? The present study empirically addresses RQ1 and RQ2 through predictive-model evaluation; RQ3 is addressed at the design level in Section 3; and RQ4 through the implementation and prototype descriptions in Sections 4 and 5.

2 Related Work

The application of artificial intelligence in higher education has gained significant momentum in recent years, with a growing number of systems designed to support students through conversational interfaces. Recent reviews indicate that most applications focus on conversational assistance, information retrieval, writing support, or administrative guidance, while a relatively small number of systems are explicitly grounded in learning theory or designed to support the complete learning process (McGrath et al., 2025; Zawacki-Richter, 2019; Lo, 2023). Some early institutional deployments include Georgia State University's chatbot that provided students with timely administrative information (Mainstay, 2026), Deakin University's Genie assistant (Deakin University, 2026), the University of Michigan's Maizey advising assistant (LSA Newnan Advising Center, 2026), and the Croatian AIVA system, which integrates a generative AI advisor into Moodle and explicitly adopts Zimmerman's self-regulated learning framework (Faculty of Organization and Informatics, 2026). While these systems demonstrate the potential of AI in student support, they share a common limitation: their primary function is answering questions and delivering information, without integrating predictive analytics,

explainability mechanisms, or systematic support for the full self-regulated learning cycle.

Predictive learning analytics has become one of the most extensively studied applications of artificial intelligence in higher education. Recent systematic reviews report that machine learning methods are widely used to predict academic performance, dropout risk, and student success, with ensemble techniques such as Random Forest and XGBoost consistently achieving strong predictive performance on educational datasets (Sghir et al., 2023; Namoun & Alshantiti, 2021). The shift from traditional statistical approaches toward ensemble machine learning methods and gradient boosting algorithms was driven by their ability to capture complex relationships within heterogeneous educational data (Sghir et al., 2023; Namoun & Alshantiti, 2021). Beyond improving prediction accuracy, recent research increasingly emphasizes the practical usefulness of predictive learning analytics in supporting decision-making by both students and instructors. To achieve this objective, predictive models must not only produce accurate estimates but also provide outputs that are sufficiently transparent and actionable to inform educational interventions (Sghir et al., 2023).

In educational settings, predictive accuracy alone is insufficient. Khosravi et al. propose the XAI-ED framework, arguing that explanations should serve a pedagogical purpose and be tailored to different types of users rather than functioning merely as technical model outputs (Khosravi et al., 2022). More generally, explainable artificial intelligence has emerged as an essential prerequisite for responsible AI, enabling users to understand, trust, and appropriately act upon model predictions (Barredo Arrieta et al., 2020). Among existing methods, SHAP (SHapley Additive exPlanations) provides theoretically grounded local explanations applicable to both tree-based and linear models (Lundberg & Lee, 2017) and has been successfully applied in educational prediction tasks to identify the factors most strongly influencing individual student outcomes. Despite these advances, predictive analytics and explainability are still rarely integrated with conversational AI or broader learning support systems that help students interpret predictions and translate them into concrete learning actions.

Self-regulated learning (SRL) is widely recognized as one of the most influential theoretical frameworks for understanding successful learning. Zimmerman (2002) introduced the three-phase cycle of forethought, performance, and self-reflection, while subsequent reviews have demonstrated that different SRL models largely converge on these core regulatory processes (Panadero, 2017). Recent research suggests that AI technologies can effectively support planning, monitoring, and reflection when designed to augment rather than replace learners' self-regulation (Lan & Zhou, 2025). However, existing AI applications often support only individual SRL activities instead of providing coordinated assistance throughout the

complete learning cycle. Prior studies have shown that digital systems supporting all three SRL phases can improve academic outcomes compared with systems focused solely on information delivery (Rahal & Zainuba, 2016).

Gamification has become an established approach in higher education, for increasing learner engagement and motivation in digital learning environments (Khaldi et al., 2023; Navarro-Espinosa et al., 2022). Meta-analyses demonstrate statistically significant positive effects on cognitive, motivational, and behavioral learning outcomes when game elements are meaningfully aligned with educational objectives (Sailer & Homner, 2020). In higher education, systematic literature reviews identify points, badges, and leaderboards as the most commonly implemented mechanisms, while emphasizing that their effectiveness depends on pedagogically meaningful integration rather than the addition of isolated reward systems (Khaldi et al., 2023; Koivisto & Hamari, 2019). Nevertheless, gamification has usually been studied as a standalone intervention rather than as an integrated component of AI-supported learning systems, limiting its potential to reinforce adaptive learning behaviors within a broader self-regulated learning process (Khaldi et al., 2023).

Table 1 summarizes the key functional dimensions of representative existing systems in comparison to the system developed in this work. Representative systems were selected based on their documented use of AI-supported student assistance in higher education and the availability of publicly accessible descriptions of their functionality. The selection was intended as an illustrative comparison rather than a systematic review or exhaustive survey. The comparison indicates that none of the selected systems simultaneously integrates AI-assisted conversation, grade prediction, prediction explainability, systematic SRL support, and gamification within a unified platform, highlighting the research gap addressed in this work.

Table 1. Comparison of selected AI-based student support systems and the proposed architecture

System	AI	Pred	XAI	SRL	Game
Georgia State	Yes	No	No	No	No
Deakin Genie	Yes	No	No	Partial	No
AIVA	Yes	No	No	Partial	No
Maizey	Yes	No	No	No	No
WGU Labs	Yes	Partial	No	Partial	No
This work	Yes	Yes	Yes	Partial	Planned

Legend:

AI conversation = conversational interaction using generative or AI-based dialogue

Predictive analytics = prediction of academic outcomes based on student data

XAI = explicit explanation of individual model predictions

SRL support = explicit support for one or more SRL regulatory processes

Gamification = game elements such as points, badges, streaks or leaderboards

3 System Architecture

The FER AI Companion is implemented as a three-tier web application consisting of a React 18 frontend, a FastAPI backend, and a dual-database persistence layer consisting of a PostgreSQL relational database hosted on Supabase and a ChromaDB vector store for semantic search over course materials. The backend coordinates the implemented AI components: a large language model (OpenAI GPT-4o) for conversational and generative tasks, locally stored machine learning models for predictive analytics, and a SHAP-based explainability layer. Communication between frontend and backend is exclusively via REST API with JWT authentication.

The architecture is organized around the three phases of Zimmerman's self-regulated learning cycle (Zimmerman, 2002), with each system component explicitly mapped to one or more phases, as illustrated in Fig. 1. At the current development stage, the predictive analytics, RAG, and study-planning components are implemented, while reflection and gamification mechanisms represent planned extensions incorporated into the architectural design.

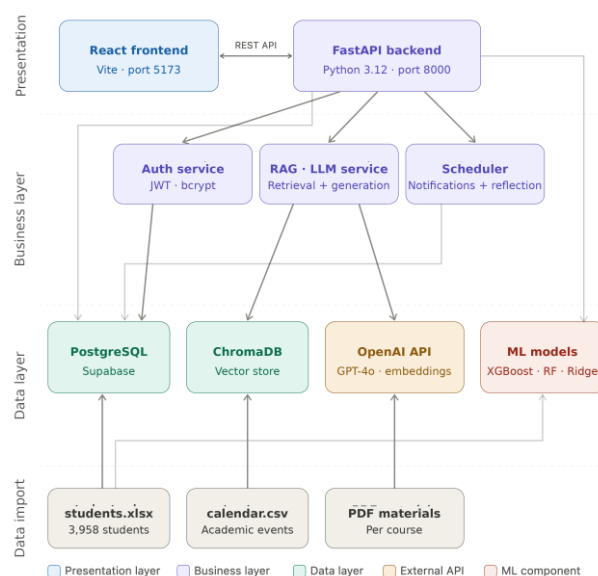


Figure 1. The system's architecture

3.1 Forethought Phase: Goal Setting and Study Planning

In the forethought phase, students set target grades for each of the five first-year courses (Digital Logic, Linear Algebra, Mathematical Analysis 1, Introduction to Programming, and Communication Skills) and request personalized study plans. The system maps the desired grade to a minimum score threshold using per-course grading scales and passes this information, together with the student's current scores and upcoming calendar events, to the language model as a structured prompt context. The model generates a markdown-formatted study plan that is stored in the database for subsequent retrieval.

The AI assistant component is built on a Retrieval-Augmented Generation (RAG) architecture (Lewis et al., 2021; Gao et al., 2024). Course materials in PDF format are segmented into approximately 500-word chunks with 50-word overlap and converted into 1536-dimensional vector representations using the text-embedding-3-small model, then stored in ChromaDB per-course collections. At query time, the five semantically closest chunks are retrieved via cosine similarity and injected into the language model prompt alongside the student's academic context and the last ten conversational turns. This architecture grounds the language model in verified course material and is expected to reduce the likelihood of hallucinations in comparison to a general-purpose LLM without retrieval.

3.2 Performance Phase: Prediction, Explainability, and AI Assistance

The predictive analytics component trains 45 models – five courses $\times$ three checkpoints $\times$ three algorithms (XGBoost ($n_estimators=300$, $learning_rate=0.05$, $max_depth=4$), Random Forest ($n_estimators=200$, $max_depth=8$, $min_samples_leaf=3$), Ridge regression ($\alpha=1.0$)) – on historical student data from academic years 2020-2024, evaluated on the 2025 cohort as a held-out test set. Three prediction checkpoints reflect the information available at different points in the semester: before the semester (enrollment features only), after the midterm exam, and after the final exam. For courses with detailed component scores available (MATAN1, LINEALG, UUP), models directly predict ISVU points on a 0-100 scale (score mode); for courses with only final grades available (DIGLOG, VJEKOM), models predict the grade on a 1-5 scale with subsequent interpolation to points (grade mode). At the mid-semester checkpoint, the midterm score is included as a predictor despite contributing directly to the final target, as the goal is to provide the best available estimate based on all information known at that point; at the late checkpoint, some continuous assessment elements may still be pending, so results should be interpreted as a consolidation estimate rather than a fully prospective prediction.

The consensus mechanism combines predictions from all three algorithms: the spread between the highest and lowest prediction serves as a model

disagreement indicator, classified as high agreement (spread ≤ 5 points), medium agreement (5-12 points), or low agreement (>12 points). These thresholds were defined with reference to the average 12-point span associated with a single grade category on the 0-100 point scale used in the system. Thus, a spread greater than 12 points indicates that the models differ by more than the approximate width of one grade category and is therefore interpreted as low model agreement. The thresholds are intended as practical, interface-level categories of inter-model disagreement rather than as calibrated measures of predictive uncertainty. This multi-model architecture is motivated not only by ensemble accuracy benefits but by the need to communicate prediction confidence, a dimension that no single model can provide on its own.

For prediction explainability, tree-based models (XGBoost, Random Forest) use SHAP TreeExplainer to compute feature-level Shapley-value-based contributions, while Ridge regression uses weighted feature contributions computed as the product of each coefficient and the mean-centered feature value, which approximates the Shapley value for linear models under the assumption of feature independence. The six features with the largest absolute contributions are returned alongside their values and directions (positive/negative), converted into human-readable explanations (e.g., "Mathematical Analysis 1 – midterm exam: 47 pts, contribution: +8.3") and visualized as a horizontal bar chart for every model separately. This implementation follows the XAI-ED pedagogical function principle (Khosravi et al., 2022): the explanation is designed to help students understand which features contributed to their prediction and potentially use this information when planning subsequent study activities.

During the performance phase, the AI assistant responds to student questions in Croatian or English, identifies the relevant course from the query, retrieves matching course material segments, and generates responses grounded in those materials. The system prompt explicitly prohibits the model from providing direct solutions to exam problems and instructs it to acknowledge uncertainty rather than fabricate information.

3.3 Self-Reflection Phase: Planned Reflection and Gamification Extensions

The components described in this section are planned extensions and have not yet been implemented. Their design illustrates the intended SRL mapping. In the self-reflection phase, future development will introduce automated reflection support and gamification mechanisms aligned with students' learning activities. A scheduler service will be designed to periodically analyze the academic calendar and generate reflection prompts for students within 24 hours after completed exams or laboratory exercises,

provided that they are enrolled in the corresponding course.

Gamification elements are planned to be integrated directly into existing student activities rather than implemented as a separate interaction flow. The proposed design includes awarding points for actions such as daily login, interaction with the AI assistant, exercise completion and viewing the progress dashboard. Badges are planned to be introduced based on predefined milestones, such as first login, first question, first completed exercise, etc. This planned approach follows the widely adopted points-and-badges model identified in the gamification literature (Khaldi et al., 2023) while aiming to reinforce self-monitoring and reflection behaviors within the SRL framework.

4 Implementation and Initial Evaluation

Since the predictive analytics component constitutes the core decision-support mechanism of the proposed architecture, the present study focuses on its initial quantitative evaluation. The remaining components are integrated into the proposed architecture, while their educational effectiveness will be investigated in future user studies.

The system is implemented in Python 3.12 (FastAPI backend) and React 18 (frontend), with SQLAlchemy as the ORM layer, Pydantic for schema validation, and JWT-based stateless authentication with bcrypt password hashing. Machine learning models are trained using scikit-learn and XGBoost and serialized as .joblib files alongside JSON metadata containing evaluation metrics and feature lists. The training dataset (students_enriched.xlsx) comprises records for 3,958 student-year observations from the 2020-2025 academic years across five first-year courses, with a chronological train/test split: models are trained on the 2020-2024 cohorts and evaluated on the 2025 cohort (663 student-year observations) to simulate real deployment conditions and avoid data leakage. Because the available dataset does not contain a persistent student identifier, the study cannot determine whether observations from different academic years correspond to the same individual. Missing values in assessment components (e.g., midterm scores for students who did not sit the exam) are imputed with zero using SimpleImputer prior to model training, reflecting the assumption that absent scores indicate non-participation rather than missing data. Five-fold cross-validation with shuffling (KFold, n_splits=5, shuffle=True, random_state=42) was applied to the training cohorts (2020–2024) for model comparison purposes. Folds were not stratified by academic year; instead, student-course records from all training years were randomly distributed across folds. Feature preprocessing is applied within each fold to

prevent data leakage from the validation set into the imputation statistics. Hyperparameters were not tuned via cross-validation; instead, values were selected based on standard recommendations from the literature for tabular datasets of comparable size.

Table 2 reports both average cross-validation performance on the training cohorts (2020-2024) and held-out test performance on the 2025 cohort for the three predictive models on score-mode courses (MATAN1, LINEALG, and UUP), addressing RQ1. Two naive baselines are included for context: a training-set mean baseline, which predicts the training average score for every student ($MAE = 11.49$, $R^2 = -0.05$ on the test set at all checkpoints), and a current-score baseline, which uses only the student's accumulated score at the given checkpoint as a single predictor (not applicable before the semester, as no current-semester scores are yet available). As expected, prediction accuracy improves substantially as additional assessment information becomes available. For the three score-mode courses reported in Table 2, average test-set R^2 across tree-based models increases from 0.27 before the semester to 0.84 at mid-semester and 0.91 at the late checkpoint, suggesting that students' actual academic performance data provides increasingly informative predictors of final outcomes.

Table 2. Average CV and held-out test metrics by checkpoint and algorithm for the three score-mode courses (MATAN1, LINEALG, and UUP); XGB = XGBoost, RF = Random Forest, RR = Ridge Regression, TM = training-set mean baseline; CS = current-score baseline (student's accumulated ISVU points at checkpoint)

Alg.	before_sem		mid		late	
Cross-validation						
	MAE	R^2	MAE	R^2	MAE	R^2
XGB	9.76	0.21	3.38	0.82	2.37	0.88
RF	9.69	0.23	3.26	0.82	2.33	0.88
RR	10.29	0.23	6.56	0.61	5.68	0.68
Test						
	MAE	R^2	MAE	R^2	MAE	R^2
XGB	9.41	0.25	3.05	0.84	2.00	0.90
RF	9.23	0.29	2.98	0.84	1.83	0.91
RR	9.92	0.19	7.99	0.49	5.83	0.66
TM	11.49	- 0.05	11.49	- 0.05	11.49	- 0.05
CS	n/a	n/a	7.45	0.55	6.55	0.64

Among the evaluated algorithms, Random Forest consistently achieves the best overall test-set performance, although its advantage over XGBoost is small (test MAE difference ≤ 0.25 points at all checkpoints). Both tree-based methods considerably outperform Ridge regression, particularly at the mid-semester checkpoint where the test MAE gap reaches approximately 5 points, suggesting that nonlinear

relationships among enrollment characteristics, continuous assessment results, and examination scores play an important role in predicting student performance. These findings are consistent with previous studies reporting the effectiveness of ensemble methods for educational prediction tasks. Importantly, both XGBoost and Random Forest outperform the training-set mean baseline at all checkpoints, and outperform the current-score baseline at the mid-semester and late checkpoints; at the mid-semester checkpoint, tree-based models achieve a test MAE of approximately 3 points compared to 7.45 for the current-score baseline, suggesting that the additional enrolment and cross-course features contribute meaningful predictive information beyond what a student's current accumulated score alone can provide.

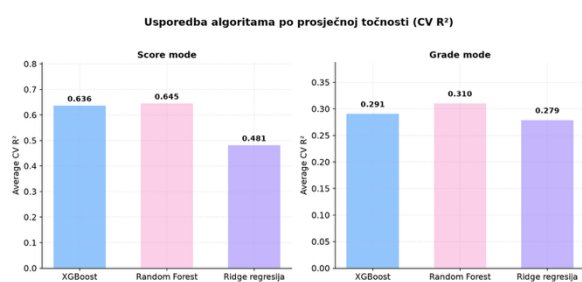


Figure 2. Comparison of predictive performance ($CV R^2$) across XGBoost, Random Forest, and Ridge Regression for score-mode and grade-mode courses.

To better understand the contribution of individual feature groups, a feature ablation study was performed on MATAN1 at the mid-semester checkpoint. Using only enrollment-related features yields a $CV R^2$ of 0.257. Adding the midterm examination score increases performance substantially to 0.908, while incorporating all remaining activity components (continuous assessment, homework, and auditory exercises) further improves performance only marginally to 0.926. This result suggests that the midterm examination is the most informative predictor among the evaluated feature groups.

Since students are presented with predictions from three different models, the prototype reports their inter-model disagreement as the spread between the highest and lowest prediction. The three agreement categories (high: ≤ 5 points; medium: 5-12 points; low: > 12 points) are used only as interface-level descriptors rather than calibrated measures of uncertainty. At the mid-semester checkpoint, high-, medium-, and low-agreement predictions yield mean absolute consensus errors of 3.1, 4.1, and 5.0 points, respectively, with the same ordering observed at the other checkpoints. Thus, greater model disagreement is associated with higher prediction error, but the agreement measure should be interpreted as a relative disagreement signal rather than as calibrated prediction confidence. Spearman correlation between prediction spread and absolute consensus error is negligible before the semester ($\rho =$

0.049, $p = 0.081$) but increases to a strong positive association at mid- and late-semester checkpoints for score-mode courses (mid: $\rho = 0.561$; late: $\rho = 0.717$, both $p < 0.001$), while remaining non-significant for grade-mode courses at all checkpoints (mid: $\rho = 0.050$, $p = 0.102$), suggesting that spread functions as a reliable uncertainty proxy only when direct score prediction is possible.

An important limitation revealed by the evaluation concerns early identification of at-risk students. Although the regression models achieve satisfactory overall prediction accuracy, they perform poorly when detecting students who ultimately fail the course. Recall for failing students is zero at both the pre-semester and mid-semester checkpoints, increasing to only 0.194 at the final checkpoint (precision = 0.75; 6 of 31 failing students correctly identified). This finding reflects the optimization objective of regression models, which minimize overall prediction error rather than emphasizing minority cases. This challenge is consistent with previous findings in student-risk prediction research, where imbalanced outcomes make metrics such as recall and precision more informative than accuracy alone (Prencak et al., 2020). Future work should therefore complement the current regression approach with dedicated classification models and appropriate class-balancing techniques for early at-risk student detection.

5 Prototype and Discussion

The developed prototype provides students with a unified interface accessible through any web browser. Fig. 3 shows the prediction view, where the student sees predicted scores from all three models alongside the consensus and the agreement level.

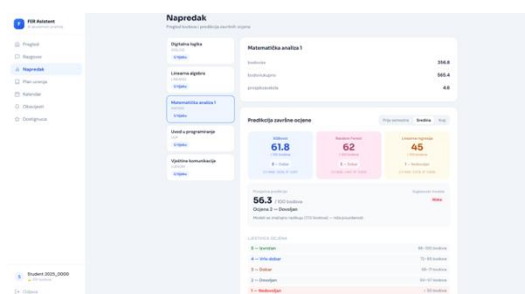


Figure 3. Prediction dashboard showing outputs of the three predictive models, consensus estimate and agreement level

The primary aim of the current stage of this research is to integrate multiple complementary AI-supported functionalities within a single student support platform. The intended practical implication of this integration is qualitatively different from what a chatbot alone can offer. For example, the prototype can present a predicted score together with model disagreement and feature-level explanations, which are

intended to provide context for subsequent study planning, while future reflection mechanisms are designed to use similar information after relevant learning activities. The planned gamification mechanisms would further connect these interactions with motivational support. The intended value of the system emerges from their integration around the SRL framework rather than from any individual component alone. The RAG architecture is intended to reduce unsupported responses by grounding generation in indexed course materials; this effect remains unevaluated. The agreement mechanism serves a similar transparent function for the prediction component: rather than presenting a single authoritative number, the system communicates its own inter-model disagreement to the student.

Limitations of the current prototype include the weaker predictive performance for courses with limited input features (DIGLOG, VJEKOM), the absence of real-user evaluation (prototype testing has not yet been conducted with students, while predictive models were evaluated using historical student data), and the inability of the current regression-based risk detection to reliably identify at-risk students before the final exam. In addition, shuffled cross-validation was used only for model comparison within the 2020–2024 training data; temporal generalization is assessed solely through the held-out 2025 cohort. The disagreement measure is a relative inter-model signal rather than calibrated uncertainty. Future work should therefore add course- and time-specific uncertainty analysis, dedicated classification for at-risk detection, broader cohorts, and user evaluation of the explanations, study plans, and planned SRL mechanisms.

6 Conclusion

This paper presented the design and initial implementation of FER AI Companion, an SRL-oriented student learning platform integrating retrieval-augmented generation, predictive learning analytics, explainable artificial intelligence, and AI-assisted study planning. Reflection support and gamification are currently represented as planned architectural extensions rather than implemented components.

Initial results suggest the technical feasibility of integrating these components into a unified web application. Experimental evaluation of the predictive models suggests that prediction accuracy improves as additional assessment data become available across three semester checkpoints, with tree-based models (XGBoost, Random Forest) outperforming Ridge regression in all evaluated conditions. The model agreement mechanism provides a relative indicator of inter-model disagreement, with high-agreement predictions exhibiting lower absolute error than low-agreement predictions.

The work presented in this paper represents the first stage of a broader research effort. While the current study focuses on the design, implementation, and technical evaluation of the platform, future work will investigate its educational effectiveness through user studies with students, evaluate the impact of explainable predictions and AI-generated study plans on self-regulated learning, and assess long-term engagement supported by the gamification mechanisms. Potential directions for future research include extending the platform to additional courses and academic years, investigating dedicated classification models for earlier identification of at-risk students, and exploring integration with existing university learning management systems.

References

- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Benetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Deakin University. (2026). Deakin's Genie: a virtual digital assistant out of the bottle. Retrieved from <https://www.deakin.edu.au/about-deakin/news-and-media-releases/articles/deakins-genie-a-virtual-digital-assistant-out-of-the-bottle>
- Faculty of Organization and Informatics. (2026). AIVA – prvi pedagoški utemeljeni AI savjetnik za studente integriran u Moodle razvijen na FOI-ju. Retrieved from <https://www.foi.unizg.hr/hr/novosti/aiva-prvi-pedagoski-utemeljeni-ai-savjetnik-za-studente-integriran-u-moodle-razvijen-na-foi>
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv*. <https://doi.org/10.48550/arXiv.2312.10997>
- Hodgson, J. (2025). How WGU Labs built an AI student support assistant. Retrieved from <https://www.wgulabs.org/posts/exploring-multi-agent-and-single-agent-systems-how-labs-is-building-an-ai-powered-student-support-assistant>
- Kasnezi, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Khalidi, A., Bouzidi, R., & Nader, F. (2023). Gamification of e-learning in higher education: A systematic literature review. *Smart Learning*

- Environments, 10(1).
<https://doi.org/10.1186/s40561-023-00227-z>
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074.
<https://doi.org/10.1016/j.caeai.2022.100074>
- Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45, 191–210.
<https://doi.org/10.1016/j.ijinfomgt.2018.10.013>
- Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *npj Science of Learning*.
<https://doi.org/10.1038/s41539-025-00319-0>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). Retrieval-augmented generation for knowledge-intensive NLP tasks. *arXiv*.
<https://doi.org/10.48550/arXiv.2005.11401>
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 410.
<https://doi.org/10.3390/educsci13040410>
- Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv*.
<https://doi.org/10.48550/arXiv.1705.07874>
- LSA Newnan Advising Center. (2026). LSA Maizey AI advising. University of Michigan. Retrieved from <https://lsa.umich.edu/advising/connect-with-an-advisor/lsa-maizey-ai-advising.html>
- Mainstay. (2026). Georgia State University chatbot supports every student. Retrieved from <https://mainstay.com/case-study/how-georgia-state-university-supports-every-student-with-personalized-text-messaging/>
- McGrath, C., Farazouli, A., & Cerratto-Pargman, T. (2025). Generative AI chatbots in higher education: A review of an emerging research area. *Higher Education*. <https://doi.org/10.1007/s10734-024-01288-w>
- Namoun, A. & Alshanqiti, A. (2021). Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. Retrieved from <https://www.mdpi.com/2076-3417/11/1/237>
- Navarro-Espinosa, A., et al. (2022). Gamification as a promoting tool of motivation for creating sustainable higher education institutions. *International Journal of Environmental Research and Public Health*, 19(5), 2599.
<https://doi.org/10.3390/ijerph19052599>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 422.
<https://doi.org/10.3389/fpsyg.2017.00422>
- Prenkaj, B., Veltri, P., & others. (2020). A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses. *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA)*.
<https://dl.acm.org/doi/10.1145/3388792>
- Rahal, A., & Zainuba, M. (2016). Improving students' performance in quantitative courses: The case of academic motivation and predictive analytics. *The International Journal of Management Education*, 14(1), 8–17.
<https://doi.org/10.1016/j.ijme.2015.11.003>
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77–112.
<https://doi.org/10.1007/s10648-019-09498-w>
- Sghir, M., et al. (2023). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and Information Technologies*. <https://doi.org/10.1007/s10639-022-11536-0>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39.
<https://doi.org/10.1186/s41239-019-0171-0>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70.
https://doi.org/10.1207/s15430421tip4102_2