From Log Data to GenAI Scaffolds: Annotation Design as a Lens for Learning Analytics in Human–AI Collaboration*

Wenting Sun[1,*]

[1] *Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany*

Abstract

Generative AI (GenAI) introduces new forms of human–AI collaboration in learning, requiring learners to construct prompts, evaluate outputs, and iteratively revise. These activities heighten demands on self-regulated learning (SRL). Learning analytics (LA) can make such processes visible through interaction logs, but its pedagogical value is limited by three challenges: raw logs are too low-level, analytic pipelines often prioritize prediction over interpretability, and outputs rarely translate into actionable instructional support. This study presents an empirical case using an ML-to-GenAI pipeline to examine how annotation design shapes what LA can reveal and act on in human–AI collaboration. We operationalized collaboration-relevant SRL signals through a two-tier semantic annotation scheme adapted for GenAI-assisted tasks and applied it to 37,724 action units from 45 preservice teachers completing a 90-minute lesson-plan assessment. We compared modeling pipelines and annotation granularities, analyzing their effects on classification performance and the quality of LLM-generated scaffolds. Findings show that learners engaged mainly in surface-level interactions, treating GenAI as a content provider rather than a co-regulator. Frequency-based representations with ensemble ML outperformed deep models, underscoring the importance of model–data alignment. Most importantly, annotation granularity shaped the clarity, concreteness, and pedagogical usefulness of scaffolding. We argue that annotation is not mere preprocessing but a central design decision for developing transparent, actionable GenAI–LA pipelines.

Keywords

Annotation design, Human–AI division of labour, Label granularity, Learning analytics as design material, Methodological exploration, Self-regulated learning

1. Introduction

Generative AI (GenAI) systems are increasingly integrated into learning environments not merely as tools, but as interactive partners that shape how learners plan, execute, and reflect on their work. When interacting with large language models (LLMs), learners must articulate goals through prompts, evaluate AI-generated responses, decide whether to accept or reject suggestions, and iteratively refine both prompts and artefacts. These activities reconfigure the division of labour between humans and AI, introducing new forms of human–AI collaboration and intensifying demands on self-regulated learning (SRL) capacities [1, 2].

Learning analytics (LA) has long sought to capture and support SRL through trace data such as clickstreams and system logs [3, 4]. Machine learning (ML) approaches have enabled scalable modelling of learning processes [5], yet much of this work prioritises predictive performance over semantic interpretability and pedagogical actionability [6, 7]. In GenAI-assisted contexts, this limitation becomes particularly salient: understanding how learners collaborate with AI—and how this collaboration can be productively supported—requires analytics that move beyond prediction toward interpretable representations and actionable guidance.

* *Joint Proceedings of LAK 2026 Workshops, co-located with the 16th International Conference on Learning Analytics and Knowledge (LAK 2026), Bergen, Norway, April 27 – May 1, 2026*

1* Corresponding author.

suwentin@hu-berlin.de (W. Sun)

0000-0003-3413-2450 (W. Sun)

Three challenges hinder log-based SRL research in GenAI-assisted tasks: (1) time-intensive behavior labeling, (2) complexity of interpreting human–AI interactions, and (3) automatic generation of instructional actions. Existing frameworks often omit prompt construction or rely on resource-heavy qualitative coding [8, 9]. Others use SRL summaries as LLM input [8, 10, 11], overlooking rich labeled log data.

LA promises to reveal collaboration patterns from trace data, but faces three obstacles:

1. Semantic bottleneck. Raw events (keystrokes, clicks) lack meaning without semantic aggregation; manual coding is slow and unscalable.
2. Model–data misalignment. Sparse logs lack linguistic depth; advanced Deep learning (DL) models may overfit, while simpler pipelines can be more interpretable.
3. Actionability gap. LA often stops at prediction, rarely connecting outputs to concrete instructional actions supporting co-regulation with AI.

This study treats LA as a design lens for shaping human–AI collaboration. We argue annotation —transforming logs into semantically meaningful units—is a central design decision that fundamentally shapes what forms of collaboration become visible and what kinds of instructional support can be generated. We examine this through a GenAI-assisted task where preservice teachers used an LLM to assess and revise lesson plans. Using a two-tier annotation scheme tailored to GenAI interactions, we compare modeling and labeling strategies to investigate how annotation design shapes analytic insights and LLM-generated scaffolding. Specifically, we ask:

- RQ1: What patterns of human–AI collaboration emerge when GenAI-assisted interaction logs are semantically annotated into SRL action units?
- RQ2: How do annotation granularity and labeling strategies shape (a) the performance and interpretability of log-based classification models and (b) the pedagogical quality of LLM-generated instructional scaffolds?

Rather than proposing a definitive SRL detection system, this paper aims to surface design trade-offs in annotation, modeling, and scaffolding that are critical for developing transparent, actionable learning analytics in GenAI-mediated learning environments. This exploratory methodological study uses an ML-to-GenAI pipeline as an analytic lens to examine how annotation design shapes what LA can reveal—and act upon—in human–AI collaboration. Through these questions, this study contributes methodological and design-oriented insights relevant to the GenAI-LA community, illustrating how analytic representations actively shape interpretations of collaboration and possibilities for instructional intervention.

2. Related Work

2.1. SRL and Its Transformation in the GenAI Era

Self-regulated learning (SRL) encompasses learners' planning, monitoring, strategy use, and reflection during goal-directed activity [12]. In GenAI-assisted learning, these processes manifest in new ways: prompt construction can be understood as a form of planning, evaluation of AI outputs as monitoring, and revision or rejection of AI suggestions as regulatory action [2]. These behaviors are not always directly observable as metacognitive states but can be inferred through interaction patterns with AI systems.

Prior research has examined SRL in GenAI-supported writing and problem-solving through qualitative coding of discourse or reflections [9, 13]. While theoretically rich, such approaches are labour-intensive and difficult to scale. Log-based studies have begun to explore SRL in digital environments [8, 10], yet many rely on coarse indicators or platform-specific summaries that overlook the nuanced dynamics of human–AI collaboration. These gaps highlight the need for

annotation schemes that explicitly account for GenAI-specific interactions while remaining grounded in SRL theory.

2.2. Annotation and Automation for Log-Based Behavioral Classification

Machine-learning and deep-learning techniques have been widely applied to log data to classify behaviors and predict outcomes [5, 7]. However, most approaches treat annotation as a technical prerequisite rather than a pedagogical design choice. Rule-based systems and predictive models often provide limited insight into why certain behaviors occur or how learners might be supported.

Annotation mediates between raw interaction traces and analytic interpretation. Fine-grained annotations can reveal cycles of planning, monitoring, and revision, while coarse labels may stabilise modeling at the cost of pedagogical specificity. Semi-automated and human-in-the-loop approaches, including LLM-supported co-labeling, offer scalability but raise questions about theoretical grounding and interpretability [14]. Understanding how annotation design shapes analytic and instructional outcomes remain an open methodological challenge.

2.3. LLMs for Qualitative Coding and Instructional Scaffolding

With GenAI's growing reasoning capabilities, LA increasingly adopts LLMs for qualitative coding. Prompt design and label granularity directly affect output quality [10, 15]. LLM-based coding can be deductive or inductive [14], often moving from literature to codebook to prompts [16]. Most automatic analyses use text inputs (chat logs, reflections) [16, 17]; few generate personalized instructional actions from log data.

Beyond education, LLMs classify and interpret system logs [18, 19], but educational logs must convey semantic meaning such as learning behaviors. Recent work combines labeled features with logs to synthesize SRL states for LLM input [8, 10, 11], though standardized statuses risk platform-specific scaffolds and require preprocessing. Manual SRL annotation remains costly; automated methods must stay theoretically grounded. LLMs show promise for both coding and instructional support, but their sensitivity to annotation structure raises the key question: how does granularity affect the pedagogical usefulness of GenAI-generated scaffolds?

3. Methods

3.1. Context and Participants

Data was collected in an educational technology course with 54 third-year preservice physics teachers; after excluding incomplete recordings, N = 45 remained. The 90-minute task involved self-assessment, GenAI-assisted peer review, and final revision, using ERNIE Bot 3.5 alongside a search engine to support their work.

3.2. Data Collection

Interactions were recorded on standardized Windows 10 machines via PSR.exe, producing .mht files with screenshots and textual descriptions. Custom Python scripts parsed these records into structured logs (step ID, timestamp, user, operation, operating area). After filtering noise, the dataset contained 37,724 entries, each paired with a screenshot to support semantic interpretation. Each screenshot was paired with its log entry. Annotation proceeded in two stages: fine-grained then coarse labels, requiring manual inspection and cross-verification.

3.3. Data Analysis

3.3.1. From Raw Log Entries to SRL Behavioral Units

Raw logs capture atomic events, whereas SRL and collaboration processes unfold over extended sequences of interaction. We therefore adopted a three-step mapping process:

- Step 1: Event consolidation. Merge consecutive low-level events into a single meaningful user action (e.g., typing a prompt across multiple keystrokes).
- Step 2: Action unit formation. Group consolidated events into "action units" based on functional similarity (e.g., reading AI outputs).
- Step 3: SRL annotation. Action units were annotated using a two-tier scheme consisting of 12 SRL-oriented event labels and 44 fine-grained action labels (Table 1).

This scheme synthesised constructs from established SRL frameworks [20] and prior analyses of GenAI-assisted learning [9, 13]. Importantly, these annotations operationalise observable proxies for collaboration-relevant SRL behaviors rather than claiming direct measurement of internal metacognitive states. Two coders independently annotated the data (one full dataset, one 20% subset), achieving 85% agreement.

Table 1
SRL Behavior Code Scheme in GenAI-assisted Lesson Plan Assessment

SRL Event labels (12)	Action labels (44)
Planning	First-reading task lists/evaluation rubrics
Reading	Reading AI outputs, Re-Reading AI outputs, Reading lesson plan
Navigation	File-searching
Monitoring	Re-read task lists, Check screen-records setting, Re-read evaluation rubrics
Orientation	Download & Upload-file, Refresh- Learning management system, Rename lesson plan, Login AI or Learning management system, WORD revision setting
High-Interaction with AI	Prompt-follow up, Search-engine for the same task, Interaction-other AI tools for the same task
Low-Interaction with AI	Prompt-typed, Prompt-pasted from other materials, Prompt-file uploaded, Prompt-polish words, Prompt-new, Prompt-old, Prompt-cancel, Explore AI interface
Low-AI outputs Elaboration	Copy Paste-AI outputs, Copy Paste All-AI outputs, Delete Pasted-AI outputs, Replace using AI outputs
High-AI outputs Elaboration	Add explanation based on AI outputs, Delete unrelated pasted AI outputs, Paraphrase based on AI outputs
High-modify	Add-Screenshot, Paste Not-AI outputs, Delete Not-AI outputs, Reorganize sections, Add-text, Text position modification
Low-modify	Modification Distance, Serial number modification, Modification Format, Undo, Save, Try to change
Courseware	Modify self-Courseware

3.3.2. ML Classification and Human-Machine Co-Labeling

To examine model–data alignment, log-derived action units were transformed into textual representations (operation + operating area). Features included TF-IDF and BERT embeddings (uer/chinese_roberta_L-2_H-128). Three pipelines were tested: BERT as classifier, BERT as feature extractor, TF-IDF with ML models.

Models: Decision Tree (DT), Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF); DL models: Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM). ML used stratified 10-fold cross-validation; RF (100 estimators) and SVM (linear kernel) were tuned. DL models trained with AdamW (batch=16, learning rate=5e-5, dropout=0.3, 20 epochs). BERT encoder was frozen to reduce overfitting and computational cost. Workflow is shown in Figure 1. All code, prompts and summaries of generated outputs are available via an anonymized OSF: https://osf.io/29hcp/?view_only=4771a1c8e91d495bbcfe003f3456cfd0

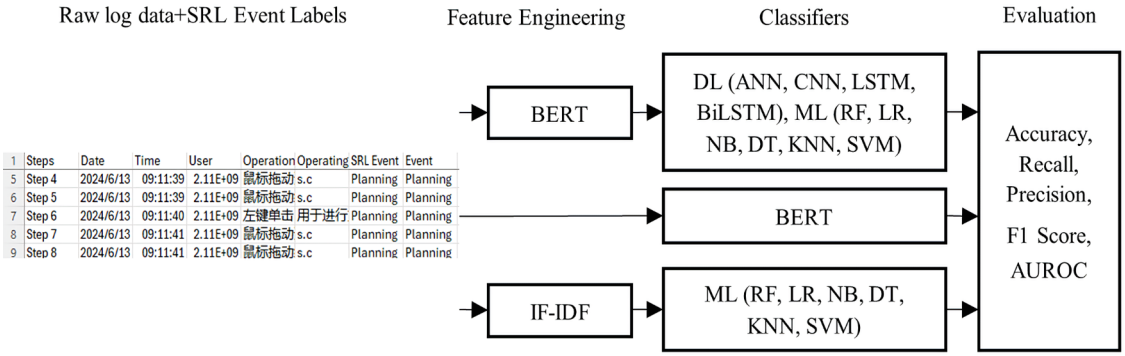

Figure 1: Overview of the Modeling Pipelines.

3.3.3. LLM-Based Instructional Action Generation

All logs were manually annotated (12 events, 44 actions); a subset trained classifiers. Rather than optimising for maximal predictive performance, the modeling stage served as a methodological probe to examine how different representations and annotation granularities interact with model behavior. The best-performing pipeline (TF-IDF + Random Forest, macro F1 and AUROC) was used to support human–machine co-labeling of the remaining data, producing confidence-scored annotations.

Five dataset versions were prepared, , varying in annotation granularity and human involvement: (1) manually labeled (12 events + 44 actions), (2) human–machine co-labeled (12+44 labels), (3) human–machine co-labeled (12 labels), (4) human–machine co-labeled (44 labels), (5) raw log data (unlabeled).

Referring to existing studies on log data interpretation and instructional action generation [11, 18, 19], structured prompts were designed to guide ChatGPT 4o in producing personalized learning reports from each dataset version. The prompts followed a four-step reasoning process: interpret logs, code SRL patterns, generate actions, validate outputs in a status–evidence–action table.

3.3.4. Evaluation

- **Classification tasks**. Assessed via Accuracy, Recall, Precision, macro F1 Score, and Area under the receiver operating characteristic curve (AUROC). Only top ML model (macro F1 Score) reported.
- **SRL status alignment**. The evaluation was based on a five-point alignment scale, which assessed the top three SRL status categories identified by the LLM against the manually labelled reference. The scale was defined as follows: 1 = No alignment (none of the top three matched); 5 = All three correctly aligned.
- **Instructional action quality**. Expert rubric (Clarity, Concreteness, Objectivity, Granularity, Specificity) adapted from [15].

4. Results

4.1. Descriptive Statistics of SRL Behaviors

From 37,724 action units, the most frequent behaviors were Reading (38.47%) and Low-modify (18.60%). AI-related behaviors were rare, with high-level AI output elaboration only 1.09%. Learners thus treated GenAI mainly as a content provider, with limited iterative prompting or critical evaluation. These patterns illuminate an imbalanced division of labour in human–AI collaboration, highlighting opportunities for targeted scaffolding.

4.2. RQ1: Classification Performance and Model-Data Alignment

On the 12-label task (Figure 2), the TF-IDF + Random Forest pipeline achieved the strongest overall performance (Accuracy = 70.33%, macro-F1 = 0.5587, AUROC = 0.93). Deep learning models underperformed, reflecting limited linguistic richness and class imbalance in log-derived text. On the 44-label task (Figure 3), performance declined across all models, illustrating the trade-off between semantic granularity and classification robustness.

Rather than interpreting performance degradation as failure, we treat it as an analytic signal: finer annotation increases semantic specificity but introduces sparsity and overlap that challenge automated classification. This trade-off is central to annotation design as a pedagogical decision.

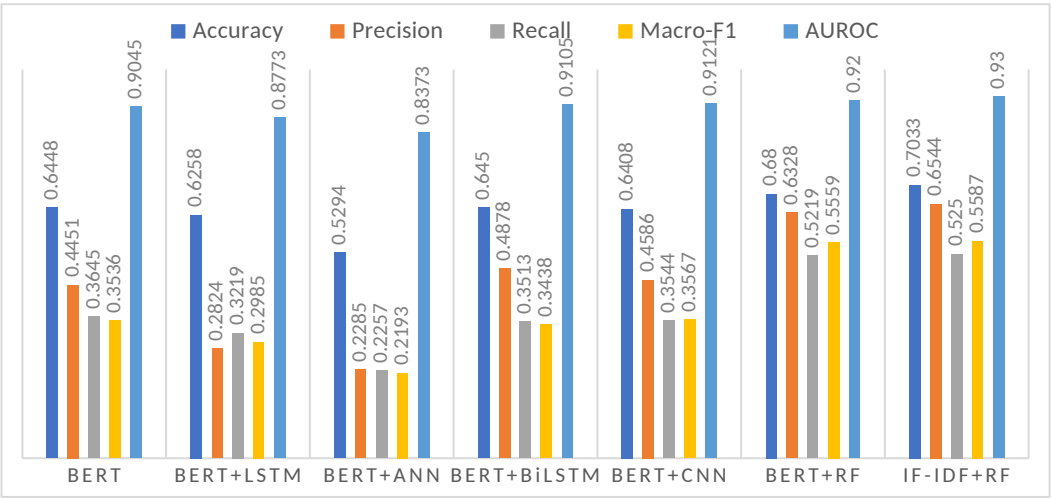

Figure 2: Performance Comparison of ML and DL Models on Log-based Classification (12 SRL Labels).

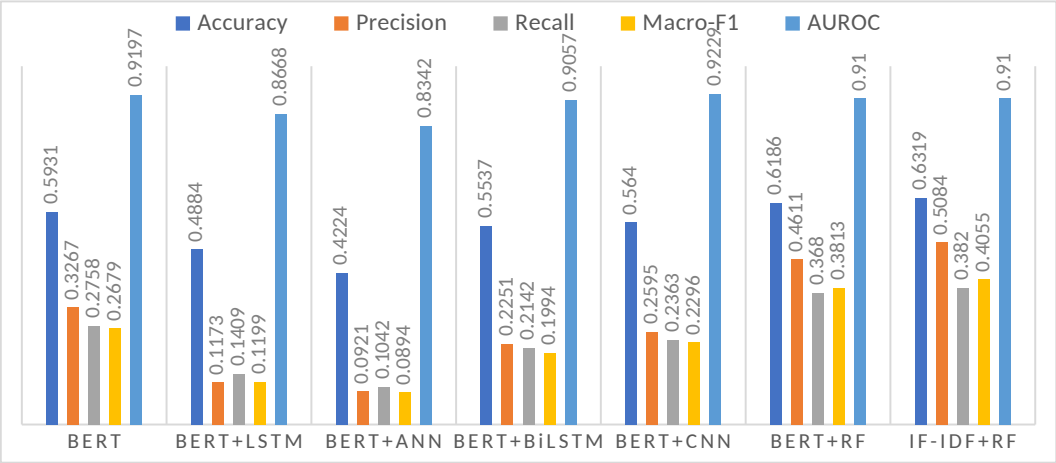

Figure 3: Performance Comparison of ML and DL Models on Log-based Classification (44 SRL Labels).

4.3. RQ2: Annotation Design and LLM-Generated Scaffolds

- **SRL status alignment.** The 12+44 dataset achieved the highest alignment with manual reference (M = 4.511, SD = 0.7869). The 12-only dataset yielded lower alignment (M = 2.622, SD = 0.4903), the 44-only dataset intermediate (M = 3.511, SD = 0.5486), and raw logs near zero (M = 0.022, SD = 0.1042). These results demonstrate Semantic annotation thus proved essential for LLMs to infer SRL states, with multi-level abstraction yielding the most reliable results.
- **Instructional action quality**. Expert ratings (Table 2) showed the manual 12+44 dataset produced the best outputs (Clarity = 4.15, Concreteness = 4.05, Specificity = 4.25). Human–machine co-labeled data approached manual quality, while raw logs yielded vague suggestions. Fine-grained labels enabled highly specific, task-level actions (e.g., "rephrase your prompt for stepwise rubric evaluation"), whereas coarse labels produced broader regulatory guidance (e.g., "improve planning"). Raw logs generated generic advice with limited instructional value. Compared with prior approaches [8, 11], our method achieved higher scores across evaluation dimensions.

Table 2
Average Scores of Evaluated Dimensions for Each Construct

Dataset version	Clarity	Concreteness	Objectivity	Granularity	Specificity
Manually labeled (12+44)	4.15	4.05	3.85	4.05	4.25
Human-machine co-labeled (12+44)	3.95	4.0	3.80	3.95	3.85
Human-machine co-labeled (12)	3.5	3.85	3.80	3.65	3.75
Human-machine co-labeled (44)	3.5	4.25	3.95	4.05	4.05
Raw log (unlabeled)	1.85	1.5	1.65	2.05	1.5

5. Discussion, Implications and Limitations

This study demonstrates how learning analytics can function as a design lens for examining and shaping human–AI collaboration in GenAI-assisted learning environments. Rather than treating annotation and modeling as purely technical steps, our findings highlight their pedagogical

consequences. We see this study as an invitation for the GenAI-LA community to reflect on annotation, modeling, and scaffolding as intertwined design decisions rather than isolated technical steps.

5.1. What LA Reveals about the Division of Labor

The dominance of reading and surface-level modification behaviors, coupled with the rarity of high-level AI interaction, suggests that learners primarily treated GenAI as a content provider rather than a collaborator. This aligns with emerging concerns that GenAI use may encourage passive consumption unless learners are explicitly supported in evaluation, iteration, and reflection. From a design perspective, learning analytics can identify such imbalances in division of labor and inform targeted scaffolds—for example, prompting learners to justify acceptance of AI outputs or to engage in comparative prompting when monitoring behaviors are absent.

5.2. Annotation Granularity as a Design Lever

A core contribution of this study is empirical evidence that annotation granularity fundamentally shapes both analytic insight and instructional actionability. Coarse labels support more stable classification and higher-level interpretations, while fine-grained labels enable specific, actionable scaffolds but introduce sparsity and modeling challenges. This trade-off suggests that annotation schemes should be designed in relation to intended pedagogical use. For real-time alerts or dashboards, coarse labels may suffice. For reflective reports or instructor-facing analytics, fine-grained labels provide richer guidance. Hybrid schemes, supported by human–machine co-labeling, offer a practical compromise.

5.3. Designing LLM Scaffolds for Productive Human-AI Collaboration

Our findings show that LLM-generated scaffolds are only as pedagogically useful as the semantic structure of their inputs. Structured evidence linking observable behaviors to collaboration-relevant SRL constructs enables LLMs to generate guidance that moves learners from passive AI use toward co-regulation and co-construction. Effective scaffolds followed a status–evidence–action structure, explicitly connecting observed behaviors (e.g., copying AI output) to regulatory interpretations (e.g., low monitoring) and concrete next steps (e.g., asking the AI to justify assumptions). Such scaffolds position GenAI not as an authority, but as a partner whose outputs must be interrogated.

5.4. Limitations

Rather than presenting a finished solution, this work contributes an empirical case that illustrates how learning analytics design choices shape human–AI collaboration, inviting further discussion and refinement within the GenAI-LA community. Several limitations should be acknowledged. First, instructional actions were generated retrospectively; future work should examine real-time deployment and learner uptake. Second, the study focused on a single task and population; generalizability across domains and GenAI tools remains to be tested. Third, class imbalance constrained fine-grained classification performance; active learning or weak supervision may improve scalability. Finally, this study used a single LLM and prompting strategy. Future research should compare models, prompts, and transparency mechanisms, and examine ethical considerations such as agency, responsibility, and equity in human–AI collaboration.

Acknowledgements

The author is deeply grateful to the late Prof. Dr. Jiangyue Liu (Soochow University, Jiangsu, China) for providing the data used in this study. His death is deeply regretted, and his support is remembered with sincere appreciation. The author also wishes to thank the organizers and

reviewers of the Third International Workshop on Generative AI and Learning Analytics, 2026 for their valuable feedback and for providing the opportunity to present and discuss this work.

Declaration on Generative AI

During the preparation of this work, the author used ChatGPT in order to: Grammar and spelling check.

References

[1] L. Markauskaite, R. Marrone, O. Poquet, S. Knight, R. Martinez-Maldonado, S. Howard, G. Siemens, Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI?, Computers and Education: Artificial Intelligence 3 (2022) 100056. doi:10.1016/j.caeai.2022.100056.

[2] L. Tankelevitch, V. Kewenig, A. Simkute, A. E. Scott, A. Sarkar, A. Sellen, S. Rintel, The metacognitive demands and opportunities of generative AI, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, ACM, 2024, pp. 1–24. doi:10.1145/3613904.3642902.

[3] P. H. Winne, Roles for information in trace data used to model self-regulated learning, in: Unobtrusive Observations of Learning in Digital Environments, Springer, Cham, 2023, pp. 175–196. doi:10.1007/978-3-031-30992-2_11.

[4] S. He, Y. Cui, A systematic review of the use of log-based process data in computer-based assessments, Computers & Education (2025) 105245. doi:10.1016/j.compedu.2025.105245.

[5] I. Osakwe, G. Chen, Y. Fan, M. Rakovic, S. Singh, L. Lim, D. Gašević, Towards prescriptive analytics of self-regulated learning strategies: A reinforcement learning approach, British Journal of Educational Technology 55 (2024) 1747–1771. doi:10.1111/bjet.13429.

[6] A. Boulahmel, F. Djelil, G. Smits, Investigating self-regulated learning measurement based on trace data: A systematic literature review, Technology, Knowledge and Learning 30 (2025) 119–156. doi:10.1007/s10758-025-09816-y.

[7] G. Gao, A. Leon, A. Jetten, J. Turner, H. Almoubayyed, S. Fancsali, E. Brunskill, Predicting long-term student outcomes from short-term edtech log data, in: Proceedings of the 15th International Learning Analytics and Knowledge Conference, ACM, 2025, pp. 631–641. doi:10.1145/3706468.3706552.

[8] T. Li, D. Nath, Y. Cheng, Y. Fan, X. Li, M. Raković, D. Gašević, Turning real-time analytics into adaptive scaffolds for self-regulated learning using generative artificial intelligence, in: Proceedings of the 15th International Learning Analytics and Knowledge Conference, ACM, 2025, pp. 667–679. doi:10.1145/3706468.3706559.

[9] M. Liu, L. J. Zhang, C. Biebricher, Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing, Computers & Education 211 (2024) 104977. doi:10.1016/j.compedu.2023.104977.

[10] K. Qian, S. Liu, T. Li, M. Raković, X. Li, R. Guan, D. Gašević, Towards reliable generative AI-driven scaffolding: Reducing hallucinations and enhancing quality in self-regulated learning support, Computers & Education (2025) 105448. doi:10.1016/j.compedu.2025.105448.

[11] A. Vujinović, N. Luburić, J. Slivka, A. Kovačević, Using ChatGPT to annotate a dataset: A case study in intelligent tutoring systems, Machine Learning with Applications 16 (2024) 100557. doi:10.1016/j.mlwa.2024.100557.

[12] P. R. Pintrich, The role of goal orientation in self-regulated learning, in: Handbook of Self-Regulation, Academic Press, 2000, pp. 451–502. doi:10.1016/B978-012109890-2/50043-3.

[13] A. Nguyen, Y. Hong, B. Dang, X. Huang, Human-AI collaboration patterns in AI-assisted academic writing, Studies in Higher Education 49 (2024) 847–864. doi:10.1080/03075079.2024.2323593.

[14] H. Zhang, C. Wu, J. Xie, F. Rubino, S. Graver, J. Cai, J. M. Carroll, Exploring inductive and deductive qualitative coding with AI: investigating inter-rater reliability between large language model and human coders, AHFE Open Access 195 (2025). doi:10.54941/ahfe1006232.

[15] X. Liu, J. Zhang, A. Barany, M. Pankiewicz, R. S. Baker, Assessing the potential and limits of large language models in qualitative coding, in: International Conference on Quantitative Ethnography, Springer, Cham, 2024, pp. 89–103. doi:10.1007/978-3-031-76335-9_7.

[16] S. Ramanathan, L. A. Lim, N. R. Mottaghi, S. Buckingham Shum, When the prompt becomes the codebook: Grounded Prompt Engineering (GROPROE) and its application to belonging analytics, in: Proceedings of the 15th International Learning Analytics and Knowledge Conference, ACM, 2025, pp. 713–725. doi:10.1145/3706468.3706564.

[17] S. Woollaston, B. Flanagan, P. Ocheja, Y. Toyokawa, H. Ogata, ARCHIE: Exploring language learner behaviors in LLM chatbot-supported active reading log data with epistemic network analysis, in: Proceedings of the 15th International Learning Analytics and Knowledge Conference, ACM, 2025, pp. 642–654. doi:10.1145/3706468.3706556.

[18] Y. Liu, S. Tao, W. Meng, J. Wang, W. Ma, Y. Chen, Y. Jiang, Interpretable online log analysis using large language models with prompt strategies, in: Proceedings of the 32nd IEEE/ACM International Conference on Program Comprehension, IEEE/ACM, 2024, pp. 35–46. doi:10.1145/3643916.3644408.

[19] L. Ma, W. Yang, S. Jiang, B. Fei, M. Zhou, S. Li, Y. Xiao, Luk: Empowering log understanding with expert knowledge from large language models, IEEE Transactions on Software Engineering (2025). doi:10.1109/TSE.2025.3594046.

[20] M. Raković, S. Iqbal, T. Li, Y. Fan, S. Singh, S. Surendrannair, D. Gašević, Harnessing the potential of trace data and linguistic analysis to predict learner performance in a multi-text writing task, Journal of Computer Assisted Learning 39 (2023) 703–718. doi:10.1111/jcal.12769.