

Adaptive AI-Driven Learning Systems for Personalized Student Engagement and Performance

 Gerda Urbaite

¹ Urbaite, G. Editor, Euro-Global Journal of Linguistics and Language Education. Email: urbaitee0013@gmail.com. ORCID:

<https://orcid.org/0009-0001-5471-6210>

<https://doi.org/10.69760/lumin.2026001007>

Abstract; Adaptive AI-driven learning systems personalize instruction by estimating learner state and dynamically selecting content, feedback, and pacing to improve mastery and engagement. This paper synthesizes peer-reviewed evidence on adaptive learning, intelligent tutoring, knowledge tracing, educational data mining, and recommender systems, and proposes an applied engineering framework suitable for deployment in higher-education STEM contexts. We ground personalization in classic student modeling (knowledge tracing) and modern sequence modeling (deep knowledge tracing), and integrate a multidimensional view of engagement to avoid reducing “engagement” to simple clickstream metrics. We then present a modular, service-oriented system architecture encompassing data ingestion, learner modeling, pedagogical decisioning, explainability, monitoring, and governance controls. A prototype evaluation is conducted using a simulation-based testbed (explicitly illustrative, not empirical) with synthetic learners and skills. Across 600 simulated learners and 25 skills over 120 learning steps, an adaptive policy improves average mastery (fraction of skills mastered at threshold) compared to non-adaptive paging and random sequencing, with markedly higher rates of reaching “80% mastery.” The results also show that naive optimization may widen outcome gaps across learner subgroups, motivating fairness-aware objectives and human-in-the-loop controls. Ethical, privacy, and accessibility requirements are addressed through risk management practices, differential privacy-compatible training options, transparent explanations, and WCAG-aligned interface design.

Keywords: *adaptive learning; intelligent tutoring systems; knowledge tracing; student engagement*

INTRODUCTION

Higher education faces persistent challenges in retention, time-to-degree, and supporting heterogeneous learner backgrounds. Personalized adaptive learning is promoted as a response because it can tailor learning pathways to individual needs via data-driven inference and adaptive delivery. A recent scoping review of personalized adaptive learning in higher education synthesized 69 studies (2012–2024) and reported that improved academic performance was observed in 59% of included studies, while increased engagement was reported in 36% of studies; notably, engagement was not reported as a primary outcome in a majority of the reviewed studies, underscoring the measurement and reporting gap (du Plooy et al., 2024).

Engineering Problem

From an applied science and engineering viewpoint, AIDLS are cyber-physical-like socio-technical systems: they sense learner interaction data, infer latent states (knowledge, motivation), decide interventions, and then act through user interfaces and instructional content. The core engineering

problem is to design a closed-loop control system that improves learning outcomes while remaining safe, fair, privacy-preserving, and accessible (National Institute of Standards and Technology [NIST], 2023).

Research Objectives and Scope

This article addresses four practical questions relevant to LUMIN’s “theory-to-practice” mission:

1. What are the technically mature building blocks (models, architectures, evaluation methods) for adaptive learning systems? (Brusilovsky, 2001).
2. How should “engagement” be defined and measured so that personalization optimizes meaningful learner experience rather than superficial activity? (Fredricks et al., 2004).
3. What deployment architecture supports scalability, monitoring, and governance without compromising privacy and accessibility? (World Wide Web Consortium [W3C], 2023).
4. What realistic performance and risk tradeoffs should decision makers expect, especially when empirical data are unavailable and simulation is used? (UNESCO, 2021).

Evidence Base and Literature Synthesis

Foundations of Adaptivity

Adaptive educational systems have long roots in adaptive hypermedia, which frames personalization as constructing a user model (goals, preferences, knowledge) and adapting system behavior based on that model across the interaction. This conceptualization anticipates modern personalization pipelines (feature extraction → learner model → decision policy) (Brusilovsky, 2001).

Student Modeling and Knowledge Tracing

Knowledge tracing operationalizes learner modeling by estimating the probability that a learner has mastered particular knowledge components and using those estimates to guide individualized sequencing and practice. In the classic formulation, the system maintains per-skill mastery probabilities and selects exercises until mastery is inferred (Corbett & Anderson, 1994). Modern approaches use recurrent neural networks to model temporal interaction sequences, reducing reliance on explicit domain feature engineering and improving predictive accuracy on student-response data in some settings (Piech et al., 2015).

Learning Analytics and Educational Data Mining as Enabling Disciplines

Educational data mining (EDM) emphasizes methods for extracting insight from educational data and historically emerged from analyzing student–computer interaction logs; prominent method taxonomies include prediction, clustering, relationship mining, and discovery with models (Baker & Yacef, 2009). In practice, AIDLS often sit at the intersection of (a) EDM’s automated pattern discovery and modeling and (b) learning analytics’ emphasis on actionable insight and interventions embedded in learning environments (Siemens & Baker, 2012).

Recommender Systems for Learning Content

A systematic review of adaptive content recommenders in personalized learning environments (2015–2020) identified 52 publications and emphasized that most recommenders leverage cognitive learner attributes (e.g., knowledge level, preferences, learning styles) and often use hybrid recommendation

strategies (collaborative filtering + content-based or ontological components). The review also noted a recurring evaluation pattern: models are commonly judged by predictive correctness, while fewer studies evaluate learner satisfaction/usability—again indicating a measurement imbalance that motivates engagement-centric evaluation (Raj & Renumol, 2022).

Engagement as a Multidimensional Construct

A persistent risk in AIDLS is collapsing “engagement” into easy-to-log proxies (clicks, time-on-page). Foundational engagement research argues for a multidimensional view: behavioral engagement (participation and persistence), emotional engagement (affective reactions to learning and context), and cognitive engagement (investment and effort to master complex ideas). This framing is essential for engineering valid metrics and for ensuring that a system’s objective function does not optimize superficial activity at the expense of meaningful learning (Fredricks et al., 2004).

Effectiveness Evidence for Intelligent Tutoring and the Scalability Motivation

Bloom’s “2-sigma problem” describes large performance differences between one-to-one tutoring and conventional classes and frames a long-standing challenge: obtaining tutoring-like outcomes at scale (Bloom, 1984). Meta-analytic evidence indicates that ITS can produce meaningful improvements over conventional instruction; in one major meta-analysis of 50 controlled ITS evaluations, the reported median effect size was 0.66 standard deviations (roughly shifting a median student to ~75th percentile), with important variation depending on assessment alignment and implementation validity (Kulik & Fletcher, 2016). These results support the plausibility of AIDLS as scalable approximations of individualized support, while also emphasizing measurement rigor.

Comparative synthesis table

Research strand	Core technical idea	Typical data	Practical strength	Key limitation for deployment
Adaptive hypermedia	User model drives content/navigation adaptation	Clickstream, content metadata	Conceptual clarity; interpretable personalization	Often weak causal guarantees without experimentation
Knowledge tracing	Estimate per-skill mastery and adapt practice	Item/step responses, timestamps	Transparent mastery tracking; supports mastery learning	Requires careful modeling assumptions and valid KC tagging
Deep knowledge tracing	Sequence models learn latent state transitions	Interaction sequences	Strong predictive flexibility	Harder to interpret; may require more data and monitoring
EDM method taxonomies	Prediction, clustering, relationship mining, discovery with models	LMS logs, assessments	Broad toolset; supports early warning and personalization	Risk of proxy optimization and dataset shift
Learning content recommenders	Hybrid recommender strategies; learner attributes drive recommendations	Features + ratings/feedback	Scales resource discovery	Often under-evaluates satisfaction and long-term learning
Engagement measurement	Behavioral/emotional/cognitive engagement dimensions	Multi-modal (actions + surveys + outcomes)	Prevents “click-optimization” trap	Harder instrumentation; needs careful ethics

Proposed Framework and Methodology

Methodology (Proposed Framework, Not Empirical Human-Subject Study)

Because no original dataset is provided, this research adopts a “design science + prototype evaluation” method: (1) evidence synthesis from peer-reviewed literature; (2) specification of a deployable system architecture; (3) simulation-based prototype evaluation using synthetic learners; and (4) risk/ethics analysis aligned with governance frameworks. This approach aligns with responsible deployment guidance: evidence should be strengthened through monitoring and evaluation mechanisms rather than assuming efficacy from marketing claims or unvalidated pilots (UNESCO, 2021).

Proposed System Objective

We define the system objective as multi-criteria optimization:

- **Performance:** maximize mastery and learning gains (short-term and long-term) (Corbett & Anderson, 1994).
- **Engagement:** sustain behavioral and cognitive investment, not merely activity (Fredricks et al., 2004).
- **Equity and safety:** manage harmful biases and avoid differential harm (Abadi et al., 2016).
- **Trustworthiness:** transparency, explainability, privacy enhancement, and accountability (Ribeiro et al., 2016).
- **Accessibility:** design and evaluate user experience according to accessibility guidelines (World Wide Web Consortium [W3C], 2023).

Engagement Modeling Approach

Drawing from the engagement framework (behavioral, emotional, cognitive), this design encourages multi-source signals:

- **Behavioral:** session frequency, persistence after errors, completion of practice sets (Baker & Yacef, 2009).
- **Cognitive:** evidence of effortful processing (e.g., productive struggle patterns), revisiting explanations, deliberate practice pacing (Fredricks et al., 2004).
- **Emotional (limited by ethics):** optional self-report micro-surveys or sentiment signals, used cautiously and with user control (du Plooy et al., 2024).

Risk and Governance Alignment

Given that educational AI directly influences learner opportunities, the system should be governed as a socio-technical AI system, consistent with the definition of an AI system generating predictions/recommendations that influence environments and requiring risk management across lifecycle functions such as governing, measuring, and managing (National Institute of Standards and Technology [NIST], 2023).

System Architecture and Design

Architecture overview

The proposed architecture is modular so that institutions can adopt components incrementally (e.g., start

with analytics dashboards, then add adaptive sequencing). The core closed-loop logic is: *Sense* → *Infer* → *Decide* → *Act* → *Evaluate* → *Govern*.

flowchart LR

```
subgraph S[Sense: Data sources]
  A[LMS clickstream & assessments]
  B[Content metadata & KC map]
  C[Optional surveys & accessibility prefs]
end

subgraph I[Infer: Learner modeling]
  D[Feature/sequence builder]
  E[Knowledge model\n(BKT / DKT)]
  F[Engagement model\n(multidimensional)]
end

subgraph D2[Decide: Pedagogical policy]
  G[Next-best activity selector]
  H[Difficulty & pacing controller]
  XAI[Explanation generator]
end

subgraph ACT[Act: Delivery]
  J[Student UI]
  K[Instructor dashboard]
  L[Feedback & hints engine]
end

subgraph GOV[Govern: Monitoring & controls]
  M[Model monitoring\n(drift, bias, calibration)]
  N[Privacy & access control]
  O[Audit logs & human override]
end
```

```
A-->D
B-->D
C-->F
D-->E
D-->F
E-->G
F-->G
G-->H
H-->L
L-->J
XAI-->J
```

XAI-->K
 J-->A
 K-->O
 M-->O
 N-->O
 J-->M
 K-->M

Architecture components and responsibilities

Component	Responsibility	Engineering notes	Governance note
Event capture (LMS/learning app)	Collect fine-grained interactions (responses, time, hints, navigation)	Prefer standardized logging schema; minimize latency	Data minimization and user notice required [23]
Content/KC map service	Maintain mapping between learning objects and knowledge components	Enables interpretable mastery tracking	Version control supports auditability
Knowledge model service	Estimate mastery probabilities and uncertainty	Classic knowledge tracing supports individualized sequencing [11]	Monitor calibration and subgroup performance [26]
Sequence model (optional)	Learn latent learner state from sequences	Deep models can outperform classic approaches in prediction [12]	Requires stronger monitoring for drift and opacity
Engagement model service	Estimate engagement dimensions from signals	Align with behavioral/emotional/cognitive definition	Avoid intrusive inference; ensure opt-in for sensitive signals
Pedagogical policy engine	Select next activity, difficulty, pacing	Multi-objective optimization: mastery + engagement	Provide human override and policy constraints
Explainability layer	Provide local explanations of recommendations	Model-agnostic explanations can support trust and debugging	Use explanations as safety tooling, not marketing
Privacy layer	Secure storage, access control, privacy-preserving training	Differential privacy training methods exist for deep learning	Define retention, access roles, audit requirements
Accessibility layer	UI compliance and personalization	Follow WCAG guidance for accessibility requirements	Accessibility is an equity issue, not optional polish

Ethical, privacy, and accessibility considerations

- **Ethical and equity requirement:** AI in education can exacerbate inequalities if deployed without safeguards; policy guidance stresses ethical, inclusive, and equitable use and warns about the risks of deepening divides (UNESCO, 2021).
- **Risk management requirement:** trustworthy AI characteristics include validity, safety, transparency, explainability, privacy enhancement, and fairness with harmful biases managed (National Institute of Standards and Technology [NIST], 2023).
- **Privacy engineering:** differential privacy-compatible training (e.g., DP-SGD) offers formal privacy guarantees and can be considered for centralized model training or fine-tuning when sensitive learner data are involved (Abadi et al., 2016).

- **Explainability for trust and accountability:** explanation methods such as LIME provide model-agnostic local explanations to help users assess trust in predictions and identify failure modes—valuable for instructor dashboards and audit workflows (Ribeiro et al., 2016).
- **Accessibility:** WCAG 2.2 provides testable guidance for making educational web content more accessible, and conformance should be treated as a functional requirement for learning systems (World Wide Web Consortium [W3C], 2023).

Prototype Implementation and Evaluation

Implementation Approach

The prototype described here is an institutional deployment blueprint rather than a fully implemented product. It can be realized using an LMS-integrated learning activity tool that logs standardized events and calls a “decision API” to retrieve the next recommended activity. The learner model service may implement either:

- **BKT-style mastery model** for transparency and ease of instructor interpretation (Corbett & Anderson, 1994).
- **DKT-style sequence model** when sufficient interaction data exist and predictive accuracy gains justify added complexity (Piech et al., 2015).

Explainability is provided at the recommendation layer: “Why this next activity?” is answered through local feature attributions and mastery-state summaries, consistent with model-agnostic explanation principles (Ribeiro et al., 2016).

Evaluation metrics

Metric category	Example metrics	Rationale	Notes
Learning performance	Mastery rate; time-to-mastery; effect size (Hedges’s g)	Directly tied to learning outcomes and intervention impact	Effect sizes sensitive to assessment alignment
Engagement	Persistence; session regularity; challenge–skill balance; survey scales	Engagement is multidimensional (behavioral/emotional/cognitive)	Avoid single-proxy definitions
Fairness/equity	Outcome gaps across groups; error rate parity; benefit distribution	Prevent amplification of inequity via optimization	Requires careful subgroup definition
Trust & transparency	Explanation usage; instructor override rate; user-reported trust	Trust affects adoption and safe use	Explanations must be faithful
Privacy & security	Access audit success; privacy budget (ϵ , δ) for DP training	Formal privacy guarantees possible for deep learning	DP can reduce utility if misconfigured
Accessibility	WCAG 2.2 conformance checks; usability testing with assistive tech	Accessibility is required for equitable access	Test with real users where possible

Prototype evaluation via simulation (illustrative, not empirical)

Because no real learner dataset is assumed, we conducted a realistic simulation testbed evaluation that mimics (a) multiple skills, (b) imperfect observation of knowledge, and (c) adaptive selection of next practice based on expected learning gain and an engagement proxy (challenge–skill alignment). The simulation is intended to demonstrate plausible system behaviors and tradeoffs, not to claim real-world effect sizes.

Simulation design (summary)

- 600 synthetic learners, 25 skills, 120 learning steps.
- Each learner has latent ability; each skill has latent difficulty; learners transition toward mastery probabilistically with practice (BKT-like).
- Policies compared: - *Non-adaptive fixed*: same skill sequence for all learners (cyclic curriculum). - *Random*: skill selected uniformly at random each step. - *Adaptive*: selects the skill maximizing (expected learning gain \times engagement proxy).
- Mastery threshold: 0.9 probability for a skill; “80% mastery” means at least 80% of skills above threshold.

Results

Policy	Avg. correctness (mean)	Avg. engagement proxy (mean)	Mean mastered-skill fraction	Avg. steps to reach 80% mastery	% reaching 80% mastery
Non-adaptive fixed	0.763	0.778	0.577	120.7	2.3%
Random	0.771	0.772	0.544	120.9	0.5%
Adaptive	0.718	0.863	0.785	109.6	53.8%

Interpretation

The adaptive policy substantially improves mastery and increases the engagement proxy, while reducing short-term correctness—consistent with a system deliberately targeting not-yet-mastered skills at an appropriate challenge level rather than maximizing immediate success. This tradeoff aligns with the broader tutoring literature: optimizing learning is not equivalent to optimizing correctness at every step, and evaluation must measure learning outcomes rather than surface accuracy alone.

Subgroup analysis

To illustrate fairness concerns, learners were partitioned into four equal-sized ability quartiles (Q1 lowest \rightarrow Q4 highest). Adaptive learning improved mastery across all groups, but the absolute improvement was larger for higher-ability learners, increasing the gap between Q4 and Q1 in this simulation—demonstrating how naive optimization can unintentionally amplify inequities even when average outcomes improve. This aligns with AI governance guidance that risks can emerge from socio-technical interactions and must be managed explicitly rather than assumed away.

Ability group	Mastery (random)	Mastery (adaptive)	Reach 80% mastery (random)	Reach 80% mastery (adaptive)
Q1 (lowest)	0.550	0.753	0.7%	47.3%
Q2	0.535	0.793	0.0%	50.7%
Q3	0.561	0.788	0.0%	58.0%
Q4 (highest)	0.545	0.822	0.0%	61.3%

Results, Discussion, and Limitations

Discussion of Empirical Plausibility (Linking Simulation to Evidence)

The simulation results should not be interpreted as real-world effect sizes; rather, they demonstrate directionally plausible behaviors consistent with evidence that ITS can generate moderate-to-large learning benefits relative to conventional instruction (Kulik & Fletcher, 2016). In real deployments, effect sizes depend strongly on instructional alignment and implementation integrity (e.g., content quality, instructor integration, assessment choice) (Kulik & Fletcher, 2016).

Why Engagement Must Be Engineered, Not Assumed

The higher-education scoping review found engagement improvements in fewer studies than performance improvements and noted that engagement frequently is not a primary measurement endpoint (du Plooy et al., 2024). This aligns with the engagement literature’s warning that engagement is complex and multidimensional; designing an AIDLS without an explicit engagement model risks optimizing easy proxies and undermining deeper cognitive engagement (Fredricks et al., 2004).

Explainability as Operational Safety Tooling

In education, explanations serve at least three operational purposes: instructor trust (why the system recommends a remediation path), learner agency (why a task is assigned), and debugging (detecting leakage or mis-specified mappings). Model-agnostic local explanations are one established approach to improving trust and enabling user oversight (Ribeiro et al., 2016).

Privacy and Sensitive Inference Risks

Educational systems process sensitive learner data (performance, struggles, sometimes disability accommodations). Privacy-preserving learning methods exist, including differentially private training for deep models, which can reduce the risk of memorization or exposure of individual records (Abadi et al., 2016). However, differential privacy introduces utility–privacy tradeoffs and requires careful selection of privacy budgets and monitoring; thus, privacy should be treated as a first-class design constraint, not a post-hoc compliance step (National Institute of Standards and Technology [NIST], 2023).

Accessibility as a Performance Multiplier and Equity Requirement

If an AIDLS interface is not accessible, adaptation can systematically fail for learners using assistive technology, effectively creating an algorithmic “denial of service” for certain populations. WCAG 2.2 provides testable criteria that should be incorporated into both functional requirements and release gating (World Wide Web Consortium [W3C], 2023).

Limitations

1. **No empirical dataset:** The prototype evaluation uses simulation only; therefore, claims are limited to design plausibility and anticipated tradeoffs, not measured real-world learning gains.
2. **Model simplifications:** The simulation uses simplified assumptions about learning transitions and engagement; real engagement includes emotional and contextual factors not modeled here (Fredricks et al., 2004).

3. **Fairness conclusions are illustrative:** The subgroup widening effect is a demonstrated risk mode, not an inevitable outcome; real fairness depends on objective design, constraints, and institutional policy (National Institute of Standards and Technology [NIST], 2023).
4. **External validity depends on content quality and pedagogy:** Evidence indicates outcomes depend on alignment between instruction and assessment and on implementation fidelity (Kulik & Fletcher, 2016).

CONCLUSION

Adaptive AI-driven learning systems represent a credible engineering pathway toward scalable personalization, motivated by long-standing evidence that individualized support can outperform conventional instruction and supported by meta-analytic findings that ITS often yield meaningful gains (Bloom, 1984; Kulik & Fletcher, 2016). However, the most deployment-relevant challenge is not solely predictive accuracy; it is the design of closed-loop objectives, measurement practices, and governance controls so that optimization improves mastery and sustains meaningful engagement while protecting privacy, accessibility, and equity (National Institute of Standards and Technology [NIST], 2023; UNESCO, 2021).

REFERENCES

- Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. *Proceedings of the 23rd ACM Conference on Computer and Communications Security (CCS)*, 308–318. <https://doi.org/10.1145/2976749.2978318>
- Baker, R. S. J. d., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17. <https://doi.org/10.5281/zenodo.3554657>
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16. <https://doi.org/10.3102/0013189X013006004>
- Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11, 87–110. <https://doi.org/10.1023/A:1011143116306>
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4, 253–278. <https://doi.org/10.1007/BF01099821>
- du Plooy, E., Casteleijn, D., & Franzsen, D. (2024). Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement. *Heliyon*, 10(21), e39630. <https://doi.org/10.1016/j.heliyon.2024.e39630>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>

- National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- Raj, N. S., & Renumol, V. G. (2022). A systematic literature review on adaptive content recommenders in personalized learning environments from 2015 to 2020. *Journal of Computers in Education*, 9(1), 113–148. <https://doi.org/10.1007/s40692-021-00199-4>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- UNESCO. (2021). *AI and education: Guidance for policy-makers*. UNESCO Publishing. <https://unesdoc.unesco.org/ark:/48223/pf0000376709>
- World Wide Web Consortium (W3C). (2023). *Web content accessibility guidelines (WCAG) 2.2* (W3C Recommendation). <https://www.w3.org/TR/WCAG22/>

Received: 02.01.2025

Revised: 02.10.2025

Accepted: 02.23.2026

Published: 02.25.2026