Adaptive Learning System Roadmap for Higher Education

Principal AI Architect Report

October 23, 2025

Contents

1 Executive Summary

- **Approach:** Adopt a hybrid Retrieval-Augmented Generation (RAG) architecture enhanced by adaptive bandit learning and fine-tuned pedagogical modules.
- **Grounding:** RAG ensures factual accuracy and traceability by generating answers only from retrieved trusted content.
- **Adaptivity:** A contextual bandit policy personalizes content sequencing based on demonstrated learning gains.
- **Fine-Tuning:** LoRA/PEFT adapters are used for style, grading, and distractor generation without retraining large models.
- **Cold-Start Resilience:** Zero-shot recommendations use semantic and metadata heuristics until live feedback accumulates.
- **Learning First:** Rewards and KPIs focus on *learning gains*, not click-through rates or dwell time.
- **Incremental Delivery:** A 12-week roadmap yields measurable outcomes each 2-week sprint.
- **Safety & Governance:** Grounded responses, content length caps, privacy compliance (FERPA/GDPR).
- **Modularity:** Each layer—retrieval, pedagogy, analytics, and orchestration—is independently upgradable.
- **Outcome:** A scalable, measurable, and pedagogically sound adaptive learning engine.

2 Architecture Options Comparison

Option A: RAG-first + Reranking + Agentic Orchestration

- **Core Components:** Hybrid BM25 + dense vector search, cross-encoder reranking, orchestrator managing retrieval and grounding.
- **Cost:** Low initial cost, no training required.
- **Latency:** Moderate (2–5s end-to-end).
- **Data Needs:** Minimal; operates cold-start via semantic retrieval.

- **Expected Learning Impact:** Accurate and grounded responses; low personalization.
- **Risks:** Static experience; retrieval quality depends on corpus coverage.
- **Team Fit:** Excellent; matches existing agentic-AI skills.

Option B: Lightweight Fine-Tuning (LoRA/PEFT) + RAG

- **Core Components:** Adds small LoRA adapters for tone, pedagogy, and domain vocabulary.
- **Cost:** Moderate; one-time training of adapters on few hundred samples.
- **Latency:** Similar to base LLM; negligible overhead.
- **Data Needs:** Low to medium (synthetic data or expert-labeled).
- **Expected Learning Impact:** Improves consistency and educational clarity.
- **Risks:** Requires data governance and evaluation; small risk of drift.
- **Team Fit:** Strong; within DS team's expertise.

Option C: RL/Bandits for Content Selection on RAG Baseline

- **Core Components:** Contextual bandit (UCB/Thompson) learns to select snippets maximizing learning gain.
- **Cost:** Moderate; online training infra required.
- **Latency:** Very low runtime cost (ms).
- **Data Needs:** Requires ongoing interaction data; starts with safe heuristic policy.
- **Expected Learning Impact:** High; adaptivity yields better time-to-mastery.
- **Risks:** Reward mis-specification; exploration risks mitigated via constraints.
- **Team Fit:** Excellent; RL expertise present.

Option D: Fully Fine-Tuned Task-Specific Models

- **Core Components:** Individual small models for QG, grading, and difficulty estimation.
- **Cost:** High total dev cost across multiple pipelines.

- **Latency:** Lower per-model but sequential execution adds up.
- **Data Needs:** High (thousands of labeled examples per subtask).
- **Expected Learning Impact:** Moderate to good but limited adaptivity.
- **Risks:** Maintenance overhead; model drift; integration burden.
- **Team Fit:** Feasible but resource-intensive.

Recommendation: Adopt a hybrid of A + B + C.

3 Final Recommended Architecture

Overview Diagram

```
[User Query]
   ↓
[Orchestrator Agent] → [Retrieval Layer: BM25 + Dense Search]
   ↓
[Cross-Encoder Reranker + MMR Diversification]
   ↓
[Minimal Snippet Selector (ASR/PDF Segmenter)]
   ↓
[Pedagogical Layer: QG + Hints + Grading]
   ↓
[Assessment (IRT-based  Estimation)]
   ↓
[Bandit Policy selects Next Action or Resource]
    Feedback loop
```

Core Layers

- **Retrieval:** Hybrid search (BM25 + embeddings). Chunk size 300–500 tokens with 20–30% overlap. Rerank via cross-encoder and MMR for diversity.
- **Content Minimization:** ASR + semantic segmentation for videos, section extraction for PDFs. Hard cap: videos ≤5 min, PDFs ≤3 pages. Define sufficiency score = semantic coverage / duration.
- **Pedagogical Layer:** Generates formative questions (aligned with Bloom's taxonomy), hints, distractors, and rubrics. Uses self-consistency decoding for reliability.

- **Assessment & Analytics:** IRT-based ability θ estimation (2PL/3PL), item calibration, difficulty drift detection, learning gain tracking.
- **Agentic Orchestration:** A planner coordinates retrieval, pedagogy, evaluation, and next-step decisions based on observed performance.

4 Data Plan: From Cold Start to Flywheel

- **Cold Start:** Heuristics + metadata filters, ASR semantic coverage, weak labels from Q&A overlap, and teacher-in-the-loop bootstrapping.
- **Rapid Labeling:** Human rubric for "best minimal resource"; inter-rater agreement monitoring; active learning loops for new labels.
- **Leverage Historical Logs:** Use existing Q&A/assessment data to pretrain question generator, grader, and difficulty estimation.
- **Flywheel:** Logged interactions $\rightarrow$ reward signals $\rightarrow$ updated bandit policy $\rightarrow$ improved recommendations $\rightarrow$ more data.

5 Metrics & Evaluation

- **Retrieval Quality:** nDCG@k, Recall@k, Coverage, Time-to-first-useful-resource.
- **Minimality:** Median resource length, Overkill rate (% exceeding limits), Compression ratio.
- **Question Quality:** Expert rubric (clarity, Bloom level), pass@k, factuality alignment.
- **Assessment Quality:** Item discrimination (a), difficulty (b), guessing (c), test information, reliability, θ stability.
- **Learning Outcomes:** $\Delta\theta$, normalized gain, mastery progression, time-to-mastery.
- **Safety:** Hallucination rate, refusal accuracy, bias/fairness checks.

6 12-Week Stepwise Roadmap

1. **M1 (Weeks 1–2):** RAG baseline with minimality constraints. *Acceptance:* nDCG@3 ≥ 0.7, overkill rate $< 20\%$.

2. **M2 (Weeks 3–4):** Add cross-encoder reranker, ASR/PDF segmenters, JSON outputs. *Acceptance:* nDCG@3 ≥ 0.8, median length ≤ 3 min, hallucinations <1%.
3. **M3 (Weeks 5–6):** Pedagogy tools (QG, grading, hints) + IRT-lite calibration. *Acceptance:* rubric scores $\geq 4/5$; stable item params.
4. **M4 (Weeks 7–8):** Contextual bandit for content selection. *Acceptance:* $\Delta\theta$ proxy uplift $\geq 10\%$ with overkill unchanged.
5. **M5 (Weeks 9–10):** LoRA fine-tunes for pedagogy style, grading consistency. *Acceptance:* improved expert scores, consistent grading, no latency penalty.
6. **M6 (Weeks 11–12):** Production hardening (safety, bias, compliance). *Acceptance:* privacy checks, dashboard metrics stable, ready for release.

7 Reinforcement Learning Design

$$R = w_1(\Delta\theta) + w_2(\text{Minimality Bonus}) - w_3(\text{Hallucination Penalty}) - w_4(\text{Latency/Cost Penalty})$$

- **Algorithm:** Start with contextual bandits (Thompson Sampling or LinUCB) using retrieval candidates as arms.
- **Reward Inputs:** Post-quiz correctness uplift (proxy for $\Delta\theta$), content brevity bonus, safety penalties.
- **Evaluation:** Off-policy via IPS/DR estimators on logged data.
- **Safety:** Constrain arms by length and confidence; allow uncertainty-aware deferrals.
- **Escalation:** Move to full RL only after sufficient data and plateaued bandit performance.

8 Fine-Tuning Policy

- **No full-model fine-tuning early.** Focus on LoRA/PEFT for modular adaptability.
- **Candidate modules:** question generator, rubric grader, distractor generator, short-explainer style.
- **Entry Criteria:** N_k high-quality samples, plateaued prompt-only results, projected inference savings, governance approval.

- **Migration Plan:** Adapter transfer to new base models; versioned adapters; fall-back to prompt baseline.

9 Risks & Mitigations

- **Cold-start:** Mitigate with heuristics + teacher validation.
- **Over-long Resources:** Enforce hard caps; prefer coverage-per-minute ranking.
- **Hallucinations:** Grounded verification and refusal policy.
- **Difficulty Drift:** Regular IRT recalibration and anchor items.
- **Privacy:** Encrypted logs, role-based access, FERPA/GDPR compliance.
- **Bias:** Diversity checks and fairness audits.
- **Model Drift:** Continuous monitoring and version control for adapters.

10 Deliverables per Milestone

- **M1:** Prototype notebook, data/model cards, baseline metrics, red-team report.
- **M2:** Updated pipeline with reranker & segmenter, JSON outputs, new evaluation report.
- **M3:** Pedagogy module code, prompt library, initial IRT calibration, expert rubric results.
- **M4:** Bandit module + off-policy eval notebook, policy performance report.
- **M5:** Fine-tuned adapters, A/B test report, model cards for tuned modules.
- **M6:** Compliance report, dashboards, continuous-learning plan, final evaluation.

First 14 Days Task List

1. Content ingestion and chunking (video + PDF).
2. Implement hybrid retrieval and baseline LLM prompt.
3. Build test query set and compute nDCG/Recall metrics.
4. Prototype end-to-end RAG QA flow.
5. Document M1 findings and prepare for M2 integration.