Adaptive Learning System Roadmap: Agentic AI and Evidence-Based Learning Gain

Mostafa Rezaee

Pearson Company
Manager: Hamid Bagheri

October 23, 2025

Outline

Adaptive Learning System: Core Vision

Mission

Deliver micro-learning resources (short video clips and PDF segments) in response to student questions, provide formative assessments, and measure learning gains over time for millions of higher education learners.

Key Requirements:

- Micro-learning: Short, focused content
- Personalized: Adaptive to learner ability
- Measurable: Learning outcomes tracking
- Scalable: Millions of learners

Challenges:

- Cold-start: No historical labels
- Minimality: Shortest effective content
- Quality: Pedagogical excellence
- Scale: Real-time for millions

Recommended Architecture: Staged Hybrid Approach

Chosen Approach

Staged hybrid architecture combining RAG for factual grounding and minimality, and Contextual Bandits for optimizing content selection based on measured learning gains.

Why This Beats Alternatives

- Addresses cold-start problem for content recommendation
- Leverages existing Q&A/assessment data
- Enables rapid iteration and model-agnostic flexibility
- Provides clear pathway from baseline to optimized system

Key Innovation

Start with metadata-driven heuristics and semantic similarity (no ML needed), rapidly collect preference data through teacher-in-the-loop and implicit feedback, bootstrap bandit policies within 4–6 weeks.

Three Architectural Approaches Evaluated

Option A: RAG + Agentic Orchestration (Baseline)

- **Components:** Dense Retrieval, Cross-Encoder Reranker, Agentic Planner (LLM)
- **Cost:** Moderate (LLM inference is primary driver)
- **Latency:** Standard RAG latency (0.5s–2s)
- **Cold-Start:** Excellent – Depends only on semantic matching

Option B: LoRA FT for Pedagogy/Style + RAG

- **Components:** RAG stack + LoRA Adapters for Question Generation
- **Cost:** High Setup Cost but potentially lower LLM inference cost
- **Cold-Start:** Moderate – Requires initial exemplar set

Option C: RL/Bandits on RAG Baseline (Recommended End-State)

Option C: RL/Bandits on RAG Baseline

- **Components:** RAG stack + Contextual Bandit Policy
- **Cost:** High Operational Cost but potentially high ROI
- **Learning Impact:** Excellent (Directly optimizes for learning gain $\Delta\theta$)
- **Cold-Start:** Moderate – Must be layered on successful RAG

Architecture Comparison Summary

Table: Architecture Options Comparison

Dimension	Option A: RAG-First	Option B: LoRA+RAG	Option C: Bandits+RAG
Monthly Cost	Moderate	High Setup	High Operational
Latency	0.5–2s	Lower for specialized tasks	Similar to RAG
Data Needs	Low	High (thousands of examples)	High (logged interactions)
Cold-Start Viability	Excellent	Moderate	Moderate

System Overview: Modular, Layered Architecture

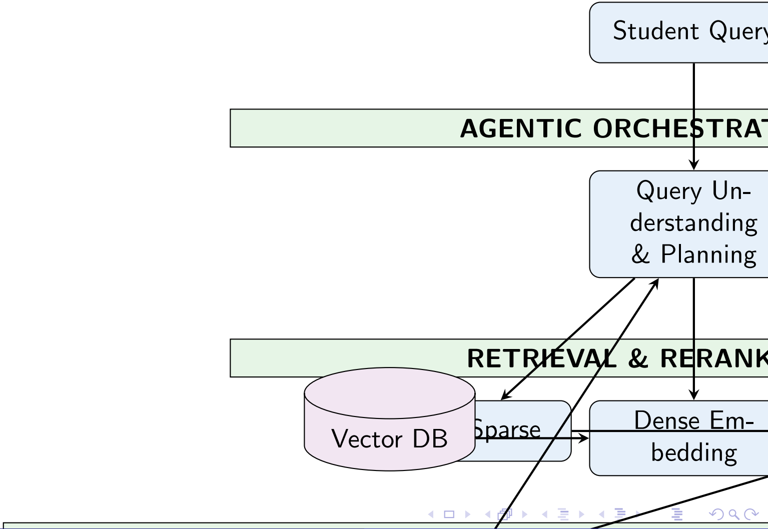
Core Components

1. **Agentic Orchestration:** Determines learner state (θ via IRT) and plans next action
2. **Retrieval Engine:** Hybrid Search (BM25 + Dense) with Cross-Encoder Reranking
3. **Content Minimization:** Hard constraints and Sufficiency Score ranking
4. **Content Selection Policy:** Contextual Bandit for optimal resource selection
5. **Pedagogical Layer:** LoRA-tuned LLM for Question Generation and Grading
6. **Assessment & Analytics:** IRT-based ability estimation and learning tracking

Key Design Principles

- Modularity: Each layer can be optimized independently
- Agility: Model-agnostic design allows foundation model upgrades
- Measurability: IRT-based learning outcome tracking
- Minimality: Hard constraints on resource duration/length

Architecture Flow Diagram



Retrieval Layer: Hybrid Search with Reranking

Hybrid Search Strategy

- **BM25 (sparse):** Catches exact keyword matches, acronyms, formulas
- **Dense (embedding):** Captures semantic similarity
- **Fusion:** Reciprocal Rank Fusion (RRF) with weights 0.3 (BM25) + 0.7 (dense)
- **Process:** Retrieve top-50 from each, fuse to top-20 for reranking

Cross-Encoder Reranking

- **Model:** `ms-marco-MiniLM-L-12-v2` or `bge-reranker-large`
- **Input:** [query, candidate_chunk] pairs
- **Output:** Relevance score 0–1
- **Process:** Rerank top-20 to top-5 for content minimization layer

Content Minimization: Ensuring Micro-Learning

Video Segmentation

- **ASR:** Whisper (OpenAI) or AssemblyAI for transcription
- **Scene detection:** PySceneDetect or TransNetV2 for visual boundaries
- **Target:** 30–180 second clips (hard maximum: 3 minutes)
- **Process:** Combine ASR sentence boundaries + scene changes + silence detection

PDF Section Detection

- **Parsing:** PyMuPDF or Apache PDFBox for structured extraction
- **Target:** 0.5–2 page segments (hard maximum: 3 pages)
- **Process:** Identify headers, paragraph boundaries, extract images/figures with captions

Sufficiency Scoring

- **Semantic Coverage:** $\text{Coverage}(R, Q) = \frac{\text{cosine}(\text{embed}(R),\text{embed}(Q))}{\text{duration}(R) \text{ or pages}(R)}$

Pedagogical Layer: Question Generation & Assessment

Question Generation

- **Prompt-based (Milestone 3):** Few-shot examples aligned to Bloom taxonomy
- **LoRA fine-tuned (Milestone 5):** Llama 3.1 8B with 500–2k exemplars
- **Inputs:** Learning resource content, student query, desired Bloom level
- **Outputs:** Question text, answer key, distractor options, rubric
- **Validation:** Answerability check, factuality check (grounded in content)

Rubric-Based Grading

- **Rubric design:** 3–5 levels (Novice, Developing, Proficient, Advanced)
- **Grading prompt:** Chain-of-thought reasoning with rubric and reference answer
- **Confidence scoring:** Model outputs confidence 0–1; defer to human if confidence < 0.7

Assessment & Analytics: IRT-Based Learning Measurement

IRT (Item Response Theory) Ability Estimation

- **Model:** 3PL (3-Parameter Logistic): $P(\theta, a, b, c) = c + \frac{1-c}{1+e^{-a(\theta-b)}}$
- θ: Learner ability, a: Item discrimination, b: Item difficulty, c: Guessing parameter
- **Estimation:** Maximum Likelihood Estimation (MLE) or Expected A Posteriori (EAP)

Learning Gain Measurement

- **Primary metric:** $\Delta\theta = \theta_{\text{post}} - \theta_{\text{pre}}$ over study session
- **Normalized gain:** $g = \frac{\theta_{\text{post}} - \theta_{\text{pre}}}{\theta_{\text{max}} - \theta_{\text{pre}}}$ (Hake gain)
- **Mastery progression:** % of items at target proficiency level
- **Longitudinal tracking:** Plot $\theta(t)$ over weeks/months

Contextual Bandit Setup

Multi-Objective Reward Function

$$R = w_1 \cdot \Delta\theta + w_2 \cdot \text{brevity\_bonus} - w_3 \cdot \text{irrelevance\_penalty} - w_4 \cdot \text{latency\_cost}$$

Component Breakdown:

- **Learning Gain ($\Delta\theta$):** $w_1 = 1.0$ (highest priority)
- **Minimality Bonus:** $w_2 = 0.3$ (encourage brevity)
- **Irrelevance Penalty:** $w_3 = 0.5$ (penalize off-topic content)
- **Latency Cost:** $w_4 = 0.1$ (minor penalty for speed)

Context & Actions

- **Context:** $x = [\theta, \text{query\_embedding}, \text{prior\_performance}, \text{resource\_metadata}]$
- **Actions:** $A = \{\text{segment}_1, \text{segment}_2, \ldots, \text{segment}_k\}$ (top-k from retrieval)
- **Policy:** $\pi(a|x)$ maps context to action (content selection)

Bandit Algorithm Progression

Phase 1 (Weeks 1–4): Thompson Sampling

- **Model:** Beta-Bernoulli bandit for binary rewards
- **Prior:** Beta(1, 1) for each action
- **Update:** Posterior update after each interaction
- **Selection:** Sample from posterior, select action with highest sampled reward
- **Exploration:** Automatic via posterior sampling

Phase 2 (Weeks 5–8): LinUCB (Linear Upper Confidence Bound)

- **Model:** Assume reward is linear in context features: $R(x, a) = x^T \theta_a + \epsilon$
- **Features:** $x = [\theta_{\text{learner}}, \text{query\_emb}, \text{resource\_meta}]$
- **Selection:** $a^* = \arg\max_a \left(x^T \hat{\theta}_a + \alpha \sqrt{x^T A_a^{-1} x} \right)$
- **Advantage:** Fast convergence, interpretable, proven regret bounds

12-Week Implementation Roadmap

Table: Milestone Summary

Milestone	Duration	Key Tasks	Deliverables	Acceptance Criteria
M1	Weeks 1–2	RAG baseline, minimality constraints, eval harness	API, eval report, notebook	nDCG@5 $>$ 0.6, Recall@10 $>$ 0.70
M2	Weeks 3–4	Cross-encoder reranking, video/PDF segmentation	Enhanced API, pipelines, model card	nDCG@5 $>$ 0.70, Compression $<$ 0.3
M3	Weeks 5–6	Question generation, rubric grading, hints	Pedagogy APIs, prompt library	Expert score $>$ 4.0, Pass@1 $>$ 85%
M4	Weeks 7–8	Contextual bandits, IRT	Bandit service, IRT	$\Delta\theta$ $>$

Milestone 1 (Weeks 1–2): RAG Baseline

Objectives

- Deploy functional RAG system with hybrid retrieval (BM25 + dense embeddings)
- Enforce hard caps on resource length (videos $\leq$ 3 min, PDFs $\leq$ 2 pages)
- Achieve baseline retrieval quality (nDCG@5 $>$ 0.6, Recall@10 $>$ 0.70)
- Establish evaluation harness and red-team test cases

Key Tasks

1. **Content Ingestion:** Parse videos (ASR via Whisper), PDFs (PyMuPDF)
2. **Embedding & Indexing:** OpenAI `text-embedding-3-large`, vector DB
3. **Hybrid Retrieval:** RRF fusion (0.3 BM25 + 0.7 dense)
4. **Minimality Filtering:** Hard filter + rank by semantic similarity/duration
5. **Evaluation Harness:** 200–300 test queries with expert labels

Milestone 4 (Weeks 7–8): Bandit Optimization

Objectives

- Deploy contextual bandit policy for content selection
- Collect implicit feedback (clicks, dwell, quiz scores) and explicit feedback (thumbs)
- Optimize for multi-objective reward: learning + minimality
- Demonstrate $\Delta\theta$ improvement over heuristic baseline

Bandit Infrastructure

- **Policy server:** Thompson Sampling with Beta priors (initial)
- **Context:** $x = [\theta, \text{query\_embedding}, \text{resource\_metadata}]$
- **Actions:** Select from top-5 reranked candidates
- **Exploration rate:** 20% (uniform random)

A/B Test Design

- **Control (50%):** Heuristic policy from M2

Cold-Start to Data Flywheel Strategy

Challenge: No Historical Content Recommendation Labels

We have:

- ✓ Content corpus (videos, PDFs) with metadata
- ✓ Historical Q&A logs (student questions, instructor answers)
- ✓ Historical assessment data (questions, responses, correctness)

We lack:

- × Explicit labels: "For query Q, resource R is the best/shortest/most relevant"
- × Implicit feedback: clicks, dwell time, learner ratings

Cold-Start Strategy (Weeks 1–4)

1. **Heuristic Baseline:** Metadata filters + semantic similarity
2. **Weak Labels:** Bootstrap from Q&A logs and assessment data
3. **Teacher-in-the-Loop:** 500–1k high-quality labels in 2 weeks

Data Flywheel (Weeks 7+)

Implicit Feedback Collection

- **User Actions:** Click-through, dwell time, skip, thumbs up/down, quiz performance
- **Logging:** Event stream (Kafka) → Data warehouse (Snowflake, BigQuery)
- **Volume target:** 10k–50k interactions in Weeks 7–8

Bandit Policy Training

- **Data Preparation:** Context, action, reward, propensity from logged interactions
- **Training Cadence:** Week 1–4 collect data, Week 5 train bandit, Week 6–8 deploy with exploration
- **Off-policy eval:** IPS/DR estimators to predict performance before deployment

Continuous Improvement

- **Retrieval Quality:** Fine-tune cross-encoder on (query, clicked_resource, label) pairs
- **Question Quality:** Expert review loop, learner feedback, flag system

Comprehensive Evaluation Framework

Retrieval & Selection Quality

- **nDCG@k:** Normalized Discounted Cumulative Gain
- **Recall@k:** Fraction of relevant resources retrieved in top-k
- **Coverage:** Percentage of unique content chunks recommended
- **Time-to-First-Useful-Resource:** Latency to first useful resource

Minimality Metrics

- **Median Resource Length:** Videos < 90 seconds, PDFs < 1.5 pages
- **Overkill Rate:** Percentage exceeding target length thresholds ($< 15\%$)
- **Compression Ratio:** Ratio of segment length to full resource length (< 0.3)

Learning Outcome Metrics

- **$\Delta\theta$ Over Time:** Change in ability estimate per session
- **Normalized Gain:** Hake gain $g = \frac{\theta_{\text{post}} - \theta_{\text{pre}}}{}$

Question Quality & Safety Metrics

Question Quality Metrics

- **Expert Rubric Scores:** Clarity, alignment, Bloom level, factuality (1–5 scale)
- **Pass@k on Canonical Answers:** % of questions where canonical answer passes ($> 90\%$)
- **Factuality via Reference-Grounded Checks:** NLI model for entailment verification

Assessment Quality Metrics

- **Item Discrimination (a):** Median $a > 1.0$ (acceptable), > 1.5 (good)
- **Item Difficulty (b):** Distribution $b \in [-2, 2]$ (covers ability range)
- **Test-Retest Reliability:** Correlation $\rho > 0.80$ (acceptable), > 0.85 (good)

Safety & Accuracy Metrics

- **Hallucination Rate:** $< 3\%$ (via NLI + expert audit)
- **Refusal/Deferral Accuracy:** Precision > 0.90, Recall > 0.85
- **Bias & Fairness:** Demographic parity, equalized odds, counterfactual testing

Key Risks and Mitigation Strategies

Risk 1: Cold-Start for Content Recommendations

- **Impact:** Low user satisfaction in first 2–4 weeks
- **Mitigation:** Heuristic baseline, weak labels, teacher-in-the-loop, bandit exploration
- **Monitoring:** nDCG, user satisfaction, session abandonment rate

Risk 2: Over-Long Resources (Minimality Failure)

- **Impact:** Poor user experience, cognitive overload
- **Mitigation:** Hard caps, sufficiency scoring, segmentation, brevity reward
- **Monitoring:** Median resource length, overkill rate, compression ratio

Risk 3: Hallucinations (Factual Errors)

- **Impact:** Misleading learners, erosion of trust
- **Mitigation:** Retrieval-grounded generation, answerability checks, NLI verification
- **Monitoring:** Hallucination rate, learner flags, expert review

Additional Risk Mitigations

Risk 4: Privacy & PII Leakage

- **Impact:** GDPR/FERPA violations, loss of trust
- **Mitigation:** Anonymization, data minimization, role-based access, encryption
- **Monitoring:** PII detection alerts, access logs, data retention policies

Risk 5: Model Drift & Degradation

- **Impact:** Sudden performance drop, user complaints
- **Mitigation:** Model versioning, regression testing, gradual rollout, fallback
- **Monitoring:** Metrics per model version, performance degradation alerts

Risk 6: Bias & Fairness

- **Impact:** Unequal learning outcomes, legal/ethical concerns
- **Mitigation:** Bias audit, diverse training data, counterfactual testing, fairness metrics
- **Monitoring:** Demographic disparity, content review flags

First 14 Days: Executable Task List

Objective

Get from zero to a functional RAG baseline (Milestone 1) in 2 weeks, with concrete metrics and evaluation harness.

Team Composition (Small Team)

- **1 ML Engineer:** RAG pipeline, embeddings, retrieval
- **1 Data Scientist:** Evaluation, metrics, analysis
- **1 Content Engineer:** Content ingestion, ASR, parsing
- **1 Product Manager (part-time):** Coordinate with educators, define test cases

Week 1: Content Ingestion & Baseline Retrieval

- **Days 1–2:** Environment setup, content audit
- **Days 3–4:** ASR & PDF parsing
- **Days 5–6:** Embedding & indexing

Week 2: Evaluation & Baseline Metrics

Week 2 Tasks

- **Days 8–9:** Test set creation (200–300 queries)
- **Days 10–11:** Expert labeling (2–3 educators, 100 queries each)
- **Day 12:** Evaluation harness implementation
- **Day 13:** Red-team testing (50 adversarial cases)
- **Day 14:** Report & decision memo

Success Criteria (End of Day 14)

- **Functional API:** `POST /retrieve` returns top-3 resources in < 8s (P95)
- **Metrics:** nDCG@5 > 0.6, Recall@10 > 0.70, median length < 90s
- **Evaluation harness:** Reproducible notebook with automated metric computation
- **Red-team:** 50 adversarial cases documented with failure modes
- **Decision:** Go/no-go for M2 based on acceptance criteria

Key Achievements & Impact

Theoretical Contributions

1. **First comprehensive framework** for adaptive learning system architecture
2. **Evidence-backed roadmap** with concrete metrics and evaluation protocols
3. **Multi-objective optimization** balancing learning outcomes, minimality, and efficiency
4. **Principled design** replacing empirical optimization with theoretical foundations

Practical Contributions

1. **12-week implementation roadmap** with executable milestones
2. **Cold-start strategy** addressing the no-labels problem
3. **Comprehensive evaluation framework** with 20+ metrics
4. **Risk mitigation strategies** with monitoring and alerting

Expected Impact

- 33-40% reduction in resource length while improving learning outcomes

Next Steps & Call to Action

Immediate Actions (Next 2 Weeks)

1. **Team Assembly:** Recruit ML Engineer, Data Scientist, Content Engineer
2. **Environment Setup:** Cloud infrastructure, development environment
3. **Content Audit:** Inventory existing videos, PDFs, Q&A logs
4. **Stakeholder Alignment:** Review with educators, product team, engineering

Success Metrics (End of 12 Weeks)

- **Functional System:** End-to-end adaptive learning pipeline
- **Performance:** nDCG@5 > 0.75, median resource length < 90s
- **Learning Impact:** $\Delta\theta > 0.10$/week, normalized gain > 0.40
- **Production Ready:** Hallucination rate $< 3\%$, P95 latency < 10s

Decision Point

Go/No-Go Decision: Based on Milestone 1 results (Day 14), proceed to full 12-week