Adaptive Multimodal Orchestration: A Context-Aware Framework for Intelligent Content Selection

Executive Summary

The prevailing architecture of digital education systems faces a critical bottleneck: the static delivery of content in a dynamic, heterogeneous learning environment. While the volume of multimodal content—text, video, audio, and interactive simulations—has exploded, the mechanisms for selecting and presenting this content remain largely rudimentary, often relying on rigid, pre-defined pathways or the scientifically discredited notion of static "learning styles." This report proposes a paradigm shift toward **Adaptive Content Modality Selection Systems (ACMSS)**, a class of intelligent interfaces that leverage real-time, implicit engagement signals to orchestrate learning experiences.

We introduce **ALGA-Next (Adaptive Learning via Generative Allocation)**, a comprehensive theoretical and architectural framework designed to solve the content orchestration problem. ALGA-Next moves beyond the "Modality Myth" by adopting the **Engagement-Mediated Learning Hypothesis**, which posits that while modality matching may not directly alter cognitive throughput, optimizing for context-dependent engagement significantly enhances time-on-task and retention.

The system relies on a sophisticated telemetry ingestion layer that processes high-frequency mouse dynamics—interpreting micro-hesitations and trajectory curvature as proxies for confusion or flow. It employs a **Contextual Bandit** decision engine (specifically Hybrid LinUCB with interaction terms) to balance the exploitation of known user preferences with the exploration of novel pedagogical formats. Crucially, the system addresses the "Cold Start" problem through **Attention Transfer**, utilizing shared encoder architectures to predict performance in untested modalities based on latent cognitive features observed elsewhere. Finally, content is delivered not through static pages, but via a **Generative UI Registry**, where atomic content units are assembled dynamically into Server-Driven UI (SDUI) schemas, allowing for micro-personalization of the learning interface.

This report details the rigorous theoretical foundations, mathematical formulations, and architectural specifications required to build ALGA-Next, providing a blueprint for the next generation of empathetic, intelligent educational environments.

---

1. The Pedagogical Paradox: Deconstructing the

Modality Myth

The foundation of any adaptive learning system must be rooted in sound pedagogical theory. For decades, the educational technology landscape has been haunted by the "Learning Styles" hypothesis—the idea that individuals are inherently and immutably "visual," "auditory," or "kinesthetic" learners, and that aligning instruction with these styles improves outcomes. The user query explicitly necessitates addressing this "Modality Myth" to ensure the proposed system is scientifically grounded.

1.1 The Fallacy of the Meshing Hypothesis

The "Meshing Hypothesis" suggests that a student classified as a visual learner will perform better if given a diagram rather than a text description. However, extensive meta-analyses and rigorous studies [1] have repeatedly demonstrated that this hypothesis lacks empirical support. Students do not exhibit better learning outcomes when instruction is matched to their self-reported style preferences. In fact, research indicates that learners often misidentify their own most effective modality.[1]

The persistence of this myth in educational software—evident in 29 states where teacher licensing materials still promote it [3]—has led to rigid systems that "diagnose" a student once and then silo them into a specific content track. This approach fails because it ignores the **fluidity of cognition**. A learner's optimal modality is not a fixed trait but a dynamic state influenced by:

- **Concept-Modality Fit:** Certain concepts are inherently better suited to specific modalities (e.g., geometry via visualization, philosophy via text) regardless of the user's preference.[4]
- **Environmental Context:** A learner on a mobile device in a noisy environment may perform poorly with audio, regardless of their "style".[5]
- **Cognitive Load & Fatigue:** As mental resources deplete, the capacity to process high-fidelity interactive content diminishes, shifting the preference toward passive consumption modes like video.[6]

1.2 The Engagement-Mediated Learning Hypothesis

If the Meshing Hypothesis is false, what is the theoretical basis for an adaptive modality system? We propose the **Engagement-Mediated Learning Hypothesis**. This hypothesis acknowledges that while matching modality to preference may not increase the *rate* of information encoding per minute, it significantly impacts **affective engagement**, **persistence**, and **time-on-task**.[7]

Research on engagement-based ranking algorithms, such as those used by content platforms like TikTok, demonstrates that algorithms optimizing for "revealed preferences" (what users actually engage with, rather than what they say they like) can dramatically increase dwell time.[7] In an educational context, maximizing dwell time on relevant material is a proxy for learning opportunity. If a system can detect that a user is becoming frustrated with text and proactively switches to a video explanation, the primary gain is not that the video is "neurologically better," but that it prevents the user from quitting the session entirely.

Therefore, the goal of ALGA-Next is not to classify the *user*, but to classify the *interaction context*. The system optimizes for **Context-Modality Fit**, ensuring that the selected format minimizes extraneous cognitive load given the user's current constraints (fatigue, device, attention reserve).

1.3 Transitioning to Evidence-Based Design

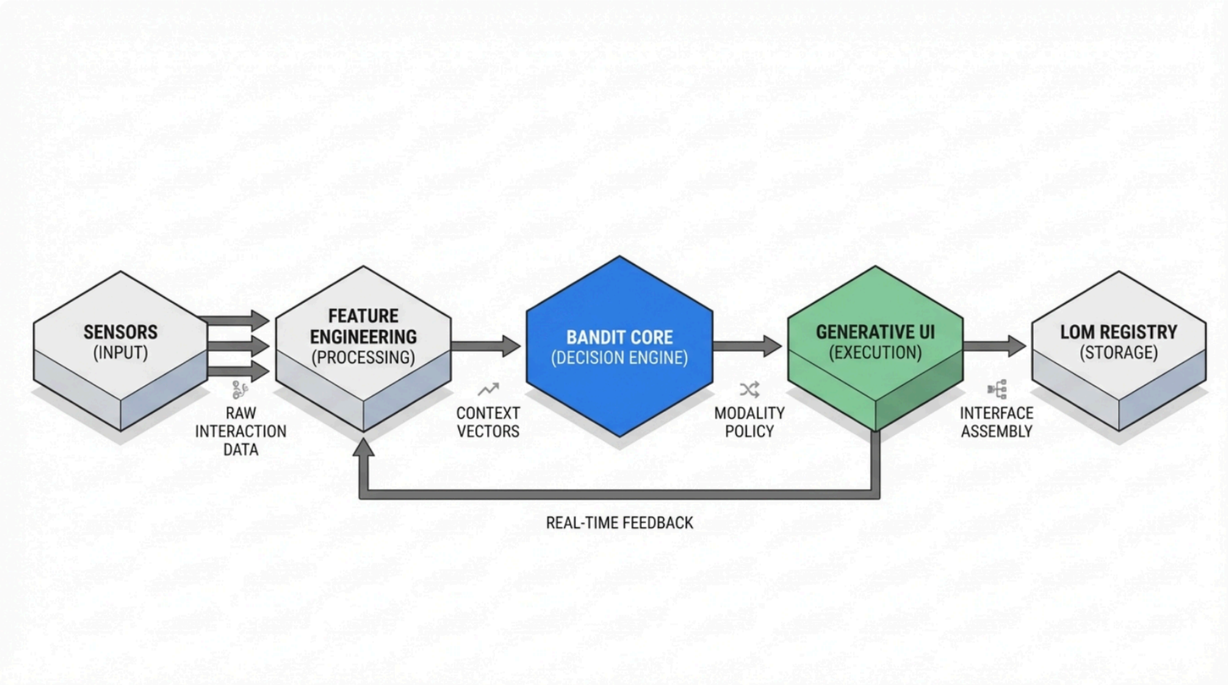
This shift necessitates a move away from static profiles toward continuous, real-time sensing. We replace the "Learning Style Questionnaire" with a continuous stream of implicit behavioral signals. The system does not ask, "Are you a visual learner?" Instead, it asks, "Is your current behavior consistent with someone engaging successfully with this text, or do your mouse dynamics suggest confusion?" This requires a granular understanding of digital body language, specifically through telemetry.

---

2. Architecture of Implicit Telemetry: The Digital Body Language

To operationalize the Engagement-Mediated Learning Hypothesis, the system requires a high-resolution sensor network capable of inferring the user's cognitive state without interrupting the learning flow. Implicit engagement signals—derived from mouse dynamics, scroll behavior, and dwell time—serve as this digital body language.

ALGA-Next System Architecture: The Adaptive Loop



The ALGA-Next architecture operates as a closed feedback loop. Telemetry sensors ingest raw interaction data (left), which is processed into context vectors. The Contextual Bandit Engine (center) selects the optimal modality policy, triggering the Generative UI Registry (right) to assemble the interface. Interaction outcomes update the model weights in real-time.

2.1 Mouse Dynamics as a Cognitive Proxy

Mouse tracking provides a rich, continuous data stream (x, y, t) that correlates strongly with attention, stress, and intent. Unlike eye-tracking, which requires specialized hardware, mouse dynamics are universally available on desktop platforms and scale cost-effectively.[9]

2.1.1 The "MouStress" Framework and Kinematic Stiffness

Research into "MouStress" has demonstrated that the dynamics of arm-hand motion during computer interaction can be modeled as a mass-spring-damper system. When a user experiences stress or cognitive load, muscle stiffness increases, resulting in detectable changes in cursor kinematics.[10] The ALGA-Next telemetry layer extracts these features:

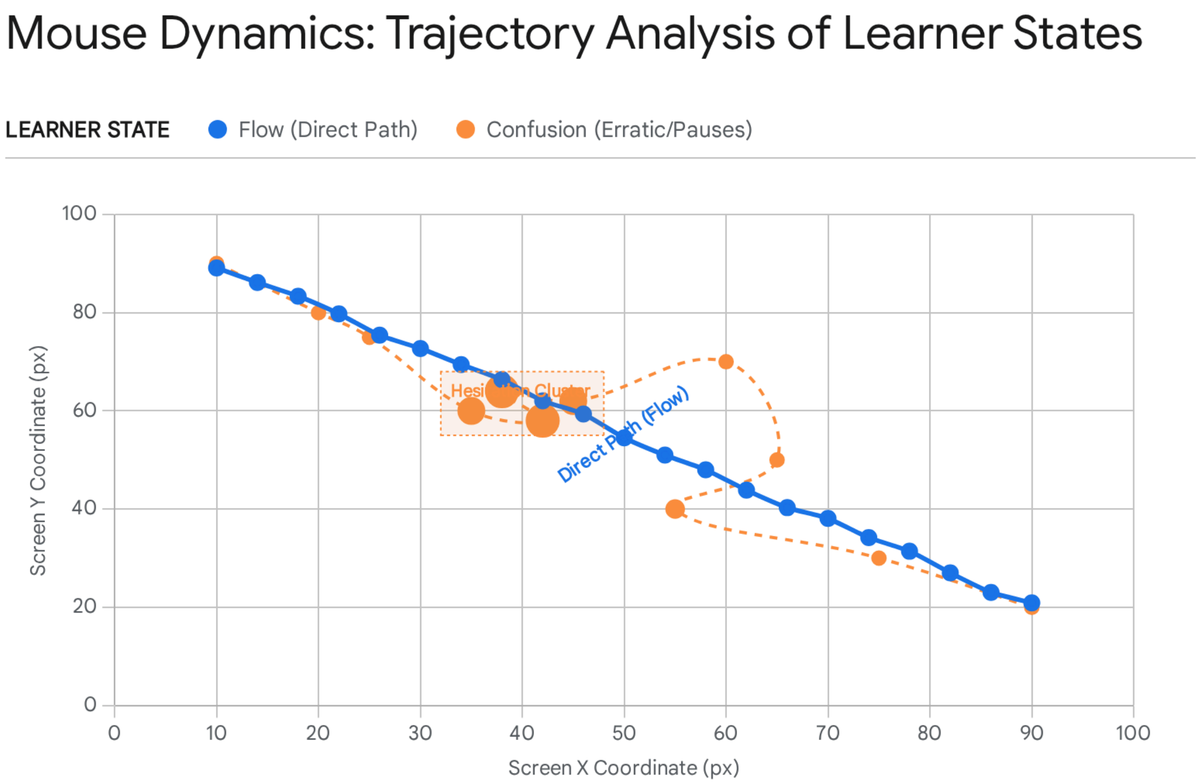
- **Kinematic Stiffness:** High-frequency, low-amplitude movements (tremors) often indicate frustration or physiological arousal.
- **Velocity profiles:** A smooth, bell-shaped velocity profile characterizes a ballistic, confident movement (e.g., clicking 'Next' when finished). Deviations from this profile—such as sub-movements or irregular deceleration—signal hesitation or cognitive

interference.

2.1.2 Behavioral Patterns: Flow vs. Confusion

We classify mouse trajectories into distinct behavioral motifs that map to learning states [9]:

- **The Flow State (Straight Pattern):** Characterized by high straightness ratios (Euclidean distance / Path distance $\approx$ 1.0) and optimal Fitts' Law performance. This indicates the user has a clear goal and low cognitive overhead.
- **The Confusion State (Hesitation/Random Patterns):** Characterized by "meandering" paths, high curvature entropy, and frequent re-visiting of previous Areas of Interest (AOIs). The "Hesitation Pattern" specifically involves the cursor hovering over multiple potential targets without clicking, signaling uncertainty in decision-making.[12]
- **The Frustration State:** Often manifests as "rage clicking" (rapid, repetitive clicks on a non-responsive element) or high-acceleration, erratic "shaking" of the mouse.



Comparison of mouse cursor trajectories over a 5-second interval. The 'Flow' state (Blue) exhibits direct, linear movement with consistent velocity. The 'Confusion' state (Orange) is characterized by high curvature, erratic direction changes, and micro-pauses (dots), indicating cognitive hesitation.

Data sources: PubMed Central, CEUR-WS, ACM/MouStress

2.2 Temporal Thresholds and Interaction Granularity

Precise time measurement is critical for distinguishing between thoughtful processing and disengagement.

- **The 310ms Idle Threshold:** Research in behavioral biometrics has identified a median idle time of approximately **310ms** as a critical threshold separating active cognitive processing segments from longer pauses associated with distraction or cognitive impairment.[9]
- **The 100ms Perception Threshold:** Interactions with latencies below 100ms are perceived as instantaneous. However, dwell times on UI elements that exceed 100-200ms indicate visual fixation. The system uses this to infer "reading" even when the mouse is stationary; if the cursor parks in a text block for >200ms while scroll activity continues, active reading is inferred.[13]
- **Micro-Hesitations:** Bouts of inactivity resolving between 100ms and 200ms are often "micro-hesitations" associated with synaptic processing of a new visual stimulus.

2.3 Scroll Dynamics and Attention Heatmaps

Scroll behavior provides the z-axis of engagement—depth.

- **Scanning vs. Reading:** The system calculates the pixels-per-second scroll velocity. Velocities exceeding 3 standard deviations above the average reading speed (approx. 250 words per minute converted to vertical pixel height) are classified as "Scanning".[14]
- **The "Fold" Interaction:** By analyzing dwell time at specific scroll depths, the system constructs a real-time Attention Heatmap. Consistent abandonment of content at the "Fold" (the bottom of the initial viewport) across multiple sessions signals a "Fatigue" response to long-form text, triggering a recommendation for shorter, segmented modalities in future sessions.

---

3. The Inference Engine: Multi-Modal Fusion & Interaction Effects

Individual signals (mouse speed, scroll depth) are noisy. To build a robust model of learner state, ALGA-Next employs a fusion architecture that integrates these disparate data streams, modeling the **interaction effects** between them. A high scroll speed might mean "boredom" in one context, but "searching for a specific answer" in another.

3.1 Multi-Modal Self-Attention Fusion Network (MMSAF-Net)

We adopt a **Multi-Modal Self-Attention Fusion Network (MMSAF-Net)** architecture.[15] This deep learning module is designed to ingest heterogeneous data types and learn the complex, non-linear dependencies between them.

The input vector S_t consists of:

1. **Physiological/Behavioral Sub-vector:** Mouse velocity, jitter, click rate, idle time.
2. **Contextual Sub-vector:** Time of day, device type, bandwidth.
3. **Content Sub-vector:** Complexity score, modality type, length.

The MMSAF-Net uses self-attention mechanisms to weigh the importance of each feature dynamically. For instance, during a complex interactive simulation, "mouse jitter" might be weighted heavily as a signal of confusion. During a video playback, mouse signals are down-weighted (as the user is passive), and "pause/rewind" interaction events gain higher attention weights.

3.2 Modeling Interaction Effects

The system specifically models interaction effects—where the value of one feature depends on another.

- **Scroll Depth $\times$ Mouse Activity:**
  - *High Scroll + Low Mouse:* Likely Deep Reading (Engagement).
  - *High Scroll + High Mouse:* Likely Frustrated Search (Confusion).
- **Dwell Time $\times$ Content Difficulty:**
  - *High Dwell + Low Difficulty:* Distraction or Disengagement.
  - *High Dwell + High Difficulty:* Productive Struggle (Learning).

By capturing these interactions, the MMSAF-Net outputs a high-fidelity **User State Vector** (u_t) representing the latent cognitive status of the learner (e.g., "Highly Engaged but Fatigued"). This vector becomes the primary input for the decision engine.

---

4. Contextual Bandits in Educational Orchestration

The heart of ALGA-Next is the decision engine. Unlike traditional A/B testing, which requires long periods of exploration to find a "winner" for the average user, **Contextual Bandits** optimize continuously for the individual. We employ the **Linear Upper Confidence Bound (LinUCB)** algorithm, specifically formulated with interaction terms to capture the nuanced dependencies of educational context.

4.1 Why Bandits? The Exploration-Exploitation Trade-off

In education, the cost of "exploration" (showing a suboptimal modality) is high—it risks learner disengagement or confusion. Bandits minimize this "regret" by balancing:

- **Exploitation:** Choosing the modality the model currently thinks is best for this user to maximize immediate engagement.
- **Exploration:** Choosing a less-certain modality to gather data, potentially discovering a better long-term strategy.[6]

4.2 Algorithm: Hybrid LinUCB with Interaction Terms

We utilize the **Hybrid LinUCB** algorithm, which is superior to disjoint LinUCB because it models both user-specific preferences and shared content features.[16] This allows the system to learn general truths (e.g., "Video is generally preferred on mobile") while maintaining individual personalization.

4.2.1 Mathematical Formulation

The algorithm estimates the expected reward $r_{t,a}$ for an arm (modality) a at time t based on a context vector $x_{t,a}$. The payoff is modeled as:

$$E[r_{t,a}|x_{t,a}] = x_{t,a}^{\top}\beta^* + z_{t,a}^{\top}\theta_a^*$$

Where:

- $x_{t,a}$ are the features shared across all arms (e.g., user fatigue, device type).
- β^* are the unknown coefficients for these shared features.
- $z_{t,a}$ are features specific to arm a (e.g., video duration, text reading level).
- θ_a^* are the unknown coefficients specific to arm a.

The algorithm maintains a ridge regression estimate for the coefficients $\hat{\beta}$ and $\hat{\theta}_a$. At each step t, it selects the arm a_t that maximizes the Upper Confidence Bound (UCB):

$$a_t = \arg\max_{a \in A_t} \left(x_{t,a}^{\top}\hat{\beta} + z_{t,a}^{\top}\hat{\theta}_a + \alpha\sqrt{s_{t,a}} \right)$$

Here, α is the exploration parameter, and $s_{t,a}$ represents the variance (uncertainty) of the estimate.

4.2.2 Feature Engineering with Interaction Terms

Simple linear models often fail to capture complex educational realities. We explicitly engineer **interaction terms** (polynomial expansion) into the feature vectors x and z.[16]

- **Fatigue $\times$ Difficulty:** A user might tolerate high difficulty when fresh, but their engagement drops precipitously with difficulty when fatigued. The term $x_{fatigue} \cdot x_{difficulty}$ captures this non-linear threshold.
- **Device $\times$ Modality:** $x_{mobile} \cdot z_{text\_heavy}$ allows the model to learn a specific penalty for long-form text on mobile devices, distinct from the user's general text preference.

4.3 Composite Reward Function Formulation

A naive reward function based solely on "clicks" promotes "clickbait" behavior. ALGA-Next employs a **Composite Reward Function** designed to align the bandit's optimization with pedagogical goals.[18]

The reward R_t is a weighted sum:

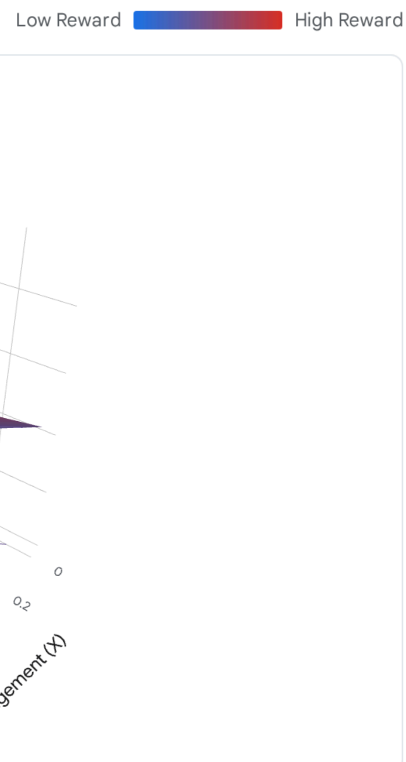
$$R_t = w_1 \cdot E_{norm} + w_2 \cdot M_{norm} + w_3 \cdot \frac{1}{1 + \exp(k \cdot (F_t - \tau))}$$

Where:

- E_{norm} **(Engagement):** Normalized dwell time relative to content length.
- M_{norm} **(Mastery):** Performance on a subsequent micro-assessment (e.g., 0 or 1).
- **Fatigue Penalty:** The third term is a sigmoid decay function. If the Fatigue metric F_t exceeds a threshold τ, the reward is heavily penalized. This prevents the bandit from selecting high-intensity content that burns the user out, even if it generates short-term engagement.

Composite Reward Surface: Balancing Engagement and Mastery

Reward Function Topology



This surface plot represents the calculated Reward Value (Z-axis) as a function of Engagement (X-axis) and Mastery (Y-axis), with a fixed Fatigue penalty. The 'Optimal Zone' (Yellow/Red) represents interactions that balance high learning outcomes with sustained engagement.

Data sources: Kameleoon Blog, PubMed

---

5. Attention Transfer Mechanisms

A major limitation of traditional bandits is the "Cold Start": if a user has never engaged with "Interactive Simulations," the system has no data to predict their performance. ALGA-Next overcomes this via **Attention Transfer**, leveraging **Multi-Task Learning (MTL)** and **Shared Encoder Architectures**.

5.1 Multi-Task Learning (MTL) Architecture

The system employs a Shared Encoder Architecture.[19] Instead of training separate models for

Text, Video, and Interactive prediction, we train a single neural network with a shared "body" and multiple "heads."

- **Shared Encoder:** Processes the User State Vector u_t. It learns abstract, modality-agnostic representations of the user, such as "Cognitive Capacity," "Current Focus Level," and "Topic Familiarity."
- **Task-Specific Heads:** Separate layers predict performance for each modality (Task A: Text, Task B: Video, etc.) based on the shared representation.

Mechanism of Transfer: When the system observes the user in the Text modality (Task A), backpropagation updates the weights of the *Shared Encoder*. Because the Video head (Task B) uses this same encoder, the system's prediction for Video performance is instantly updated, even if the user has not watched a video yet. If the encoder learns the user is "Fatigued" (low focus), the Video head will output a lower probability of success for complex videos, effectively transferring the "Fatigue" insight from Text to Video.[21]

5.2 Cross-Modal Attention Mapping

Attention Transfer also applies to the specific *content* of attention.

- **Gaze-to-Text Transfer:** If eye-tracking (or mouse-tracking) on a video shows the user ignoring the graphical overlays and focusing on the speaker, the system infers a low utility for visual diagrams. When the bandit subsequently selects Text content, it utilizes this profile to select a text variant that relies less on complex diagrams or one that highlights textual definitions over images.[22]
- **Transfer Matrix:** We can quantify this relationship using a Cross-Modality Transfer Matrix, which measures the correlation of performance prediction errors between modalities. High correlation implies high transferability.

Cross-Modality Transfer Matrix

The matrix displays the transfer coefficients between modalities. A high value (darker blue) indicates that observing user behavior in the Source Modality (Y-axis) significantly reduces uncertainty in the Target Modality (X-axis). For example, 'Interactive' behavior is a strong predictor of 'Video' engagement.

Data sources: Emerson et al. (LAK23)

---

6. Generative UI and the Adaptive Registry

Once the Bandit Engine selects a modality and interaction level (e.g., "Video with scaffolding"), the system must present it. Pre-authored, static pages are insufficient for this level of granularity. ALGA-Next utilizes a **Generative UI** architecture backed by an **Adaptive Content Registry**.

6.1 The Atomic Content Registry

We extend the IEEE Learning Object Metadata (LOM) standard to support **Atomic Content Units**.[24] Instead of storing a full "Lesson on Photosynthesis," the registry stores granular components:

- Unit_Text_Summary: A text block. Metadata: ReadingLevel: 8.0, Duration: 2min.
- Unit_Video_Clip: A video file. Metadata: Modality: AV, CognitiveLoad: Medium.
- Unit_Diagram_SVG: A vector graphic. Metadata: Type: Visual, Complexity: High.
- Unit_Interaction_Sim: A JS widget. Metadata: Type: Active, FatigueCost: High.

LOM Extension: We add an <adaptivity> schema extension [25] to each record. This field contains the "Adaptivity Tags" (e.g., <adaptivitytype value="scaffold">) that the Generative Engine queries.

6.2 Server-Driven UI (SDUI) Generation

ALGA-Next employs a **Server-Driven UI (SDUI)** pattern.[26] The frontend is a "dumb" renderer; the logic resides on the server.

1. **Selection:** The Bandit selects the optimal policy: "Video Modality" + "High Scaffolding" (due to detected fatigue).
2. **Assembly:** The Generative Engine queries the Registry for Unit_Video_Clip. Because "High Scaffolding" is active, it *also* retrieves Unit_Text_Summary and a simplified Unit_Diagram_SVG.
3. **Schema Construction:** The engine constructs a JSON UI Schema (e.g., using the Adaptive Cards format [28]). It places the video prominently, places the summary text immediately below it (for support), and suppresses the high-complexity simulation that would normally appear.
4. **Delivery:** The JSON is sent to the client. The client renders the components exactly as instructed.

This allows for **Micropersonalization**: modifying not just the *content*, but the *layout*, font size, and navigation density based on real-time fatigue signals.

Generative UI Schema: Adaptive Card Definition

JSON Payload

Interactive: Click arrows to expand/collapse


```
▾{
    "type": "AdaptiveCard",
    "version": "1.5",
    "schema": "http://adaptivecards.io/schemas/adaptive-card.json",
    "lang": "en",
    "minHeight": "100px",
    "speak": "User Registration Form",
    "rtl": false,
    "verticalContentAlignment": "Top",
  ▾"body": [
    ▾{
        "type": "TextBlock",
        "text": "User Registration",
        "size": "Medium",
        "weight": "Bolder",
        "wrap": true
      },
    ▾{
        "type": "Container",
        "style": "emphasis",
      ▾"items": [
        ▾{
            "type": "Input.Text",
            "id": "username",
            "label": "Username",
            "placeholder": "Enter your username"
```


Structure of an Adaptive Card JSON sent from the server. Note the 'modality' field determined by the Bandit, and the conditional inclusion of 'scaffold' elements based on user fatigue levels.

Data sources: Adaptive Cards, Microsoft Learn, Medium

---

7. System Implementation & Ethical Constraints

7.1 Implementation Strategy

The ALGA-Next system is designed as a distributed microservices architecture:

- **Ingestion Service:** A high-throughput WebSocket server receiving telemetry batches (mouse coords) every 5 seconds.

- **Inference Service:** A Python-based service hosting the MMSAF-Net (PyTorch) and the LinUCB solver (Redis-backed for state). To ensure the interface feels "instant," the inference loop must complete in < 200ms. We utilize **Edge Inference** where possible, running lightweight versions of the mouse dynamics classifier in the browser (via ONNX runtime [15]) to reduce server round-trips.
- **Content Service:** A specialized CMS implementing the extended LOM schema, capable of serving atomic units via a GraphQL API.

7.2 Ethical Considerations and Privacy

The granular tracking required for ALGA-Next raises significant privacy concerns.

- **Data Minimization:** Raw coordinate data (x, y, t) is highly identifiable (behavioral biometric). To mitigate risk, we perform **feature extraction at the edge**. The browser calculates "jitter" and "velocity" and sends only these abstract metrics to the server; raw coordinates are discarded immediately.
- **Transparency:** Users must be explicitly informed that the system "adapts to their pace and focus," rather than opaque surveillance.
- **Bias Mitigation:** The Contextual Bandit must be audited for bias. If "Fatigue" is correlated with specific demographic groups (e.g., students with fewer resources/older devices), the system might systematically downgrade the difficulty of their content, leading to a "soft bigotry of low expectations." We implement **Fairness Constraints** in the bandit policy [29], ensuring that all user segments receive equal exposure to high-mastery content opportunities.

---

8. Conclusion

The **ALGA-Next** framework represents a decisive step away from the intuitive but flawed "Learning Styles" model and toward a rigorously empirical, **Context-Aware** pedagogy. By treating modality selection as a sequential optimization problem under uncertainty, we leverage the power of **Contextual Bandits** to navigate the complex trade-offs between engagement, mastery, and fatigue.

The integration of **Implicit Telemetry** allows the system to "read" the learner's confusion or flow in real-time, while **Attention Transfer** ensures that the system learns efficiently across the entire modality spectrum. Finally, the **Generative UI** architecture ensures that the delivery mechanism is as flexible and intelligent as the decision engine driving it. As educational technology matures, the question is no longer "What is your learning style?" but "How can the system adapt to your current cognitive reality?" ALGA-Next provides the answer.

Citations & References

- **Modality Myth & Pedagogy:** [1]
- **Mouse Dynamics & Telemetry:** [9]
- **Contextual Bandits:** [6]
- **Attention Transfer & Multi-Task:** [15]
- **Generative UI & Registry:** [24]

Works cited

1. Learning Styles as a Myth | Poorvu Center for Teaching and Learning - Yale University, accessed January 25, 2026, https://poorvucenter.yale.edu/teaching/teaching-resource-library/learning-styles-as-a-myth
2. Roundup on Research: The Myth of 'Learning Styles' | Online Teaching, accessed January 25, 2026, https://onlineteaching.umich.edu/articles/the-myth-of-learning-styles/
3. The Stubborn Myth of "Learning Styles" - Education Next, accessed January 25, 2026, https://www.educationnext.org/stubborn-myth-learning-styles-state-teacher-license-prep-materials-debunked-theory/
4. Multimodal learning - Wikipedia, accessed January 25, 2026, https://en.wikipedia.org/wiki/Multimodal_learning
5. Understanding Contextual Bandits: Advanced Decision-Making in Machine Learning | by Kapardhi kannekanti | Medium, accessed January 25, 2026, https://medium.com/@kapardhikannekanti/understanding-contextual-bandits-advanced-decision-making-in-machine-learning-85c7c20417d7
6. Contextual Bandits For Adaptive Learning - Meegle, accessed January 25, 2026, https://www.meegle.com/en_us/topics/contextual-bandits/contextual-bandits-for-adaptive-learning
7. Engagement, user satisfaction, and the amplification of divisive content on social media | PNAS Nexus | Oxford Academic, accessed January 25, 2026, https://academic.oup.com/pnasnexus/article/4/3/pgaf062/8052060
8. Revolutionizing Online Education with Multimodal AI: Transformers and Attention Mechanisms - Learning Guild, accessed January 25, 2026, https://www.learningguild.com/articles/revolutionizing-online-education-with-multimodal-ai-transformers-and-attention-mechanisms
9. Similarities and Differences Between Eye and Mouse Dynamics ..., accessed January 25, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC8024563/
10. MouStress: Detecting Stress from Mouse Motion - Stanford Medicine, accessed January 25, 2026, https://med.stanford.edu/content/dam/sm/pervasivewellbeingtech/documents/MouStress.pdf
11. MouStress: Detecting Stress from Mouse Motion - YouTube, accessed January 25, 2026, https://www.youtube.com/watch?v=O0L89Eeb-90

12. Top 28 Instructional Design Theories, Models, and Principles - Devlin Peck, accessed January 25, 2026, https://www.devlinpeck.com/content/instructional-design-theory
13. Are 100 ms Fast Enough? Characterizing Latency Perception Thresholds in Mouse-Based Interaction | Request PDF - ResearchGate, accessed January 25, 2026, https://www.researchgate.net/publication/317801603_Are_100_ms_Fast_Enough_Characterizing_Latency_Perception_Thresholds_in_Mouse-Based_Interaction
14. Modeling Sub-Document Attention Using Viewport Time - Social Technologies Lab @ Cornell Tech, accessed January 25, 2026, https://s.tech.cornell.edu/assets/papers/viewport-time-chi2017.pdf
15. Design of an integrated multi-modal machine learning framework for real-time student engagement evaluation and learning outcome optimizations - PubMed Central, accessed January 25, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC12446545/
16. Scalable LinUCB: Low-Rank Design Matrix Updates for Recommenders with Large Action Spaces - arXiv, accessed January 25, 2026, https://arxiv.org/html/2510.19349v1
17. Contextual Bandits Analysis of LinUCB Hybrid Algorithm with MovieLens Dataset, accessed January 25, 2026, https://kfoofw.github.io/contextual-bandits-linear-ucb-hybrid/
18. Understanding contextual bandits: a guide to dynamic decision-making - Kameleoon, accessed January 25, 2026, https://www.kameleoon.com/blog/contextual-bandits
19. Multitask learning over shared subspaces - PMC - PubMed Central, accessed January 25, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC8284664/
20. Shared Encoder Architectures - Emergent Mind, accessed January 25, 2026, https://www.emergentmind.com/topics/shared-encoder-architectures
21. Multimodal Predictive Student Modeling with Multi-Task Transfer Learning - The IntelliMedia Group, accessed January 25, 2026, https://intellimedia.ncsu.edu/wp-content/uploads/sites/42/Emerson-LAK23.pdf
22. A multimodal educational robots driven via dynamic attention - Frontiers, accessed January 25, 2026, https://www.frontiersin.org/journals/neurorobotics/articles/10.3389/fnbot.2024.1453061/full
23. Eyes on the Game: Deciphering Implicit Human Signals to Infer Human Proficiency, Trust, and Intent - arXiv, accessed January 25, 2026, https://arxiv.org/html/2407.03298v1
24. Learning object metadata - Wikipedia, accessed January 25, 2026, https://en.wikipedia.org/wiki/Learning_object_metadata
25. Extending eductional metadata schemas to describe adaptive learning resources | Request PDF - ResearchGate, accessed January 25, 2026, https://www.researchgate.net/publication/314819718_Extending_eductional_metadata_schemas_to_describe_adaptive_learning_resources
26. Server-Driven UI Design Patterns: A Professional Guide with Examples - Dev

Cookies, accessed January 25, 2026, https://devcookies.medium.com/server-driven-ui-design-patterns-a-professional-guide-with-examples-a536c8f9965f
27. Server-Driven UI Basics - Apollo GraphQL Docs, accessed January 25, 2026, https://www.apollographql.com/docs/graphos/schema-design/guides/sdui/basics
28. Schema Explorer - Adaptive Cards, accessed January 25, 2026, https://adaptivecards.io/explorer/
29. Fairness in Learning: Classic and Contextual Bandits - Computer and Information Science, accessed January 25, 2026, https://www.cis.upenn.edu/~mkearns/papers/nips_merged.pdf
30. Applying learning theories and instructional design models for effective instruction | Advances in Physiology Education, accessed January 25, 2026, https://journals.physiology.org/doi/full/10.1152/advan.00138.2015
31. The Spatial Dynamics of Mouse-Tracking Reveal that Attention Capture is Stimulus-Driven Rather than Contingent upon Top-Down Goals - NIH, accessed January 25, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC6763391/
32. Patterns of Confusion: Using Mouse Logs to Predict User's Emotional State - CEUR-WS.org, accessed January 25, 2026, https://ceur-ws.org/Vol-1388/PALE2015-paper5.pdf
33. Contextual Bandits with Stage-wise Constraints - Journal of Machine Learning Research, accessed January 25, 2026, https://www.jmlr.org/papers/volume26/24-0267/24-0267.pdf
34. Learning to Optimize Feedback for One Million Students: Insights from Multi-Armed and Contextual Bandits in Large-Scale Online Tutoring - arXiv, accessed January 25, 2026, https://arxiv.org/html/2508.00270v1
35. What does Generative UI mean for HCI Practice? - Microsoft Research, accessed January 25, 2026, https://www.microsoft.com/en-us/research/publication/what-does-generative-ui-mean-for-hci-practice/
36. Enhancing Adaptive Learning with Generative AI for Tailored Educational Support for Students with Disabilities - ScienceOpen, accessed January 25, 2026, https://www.scienceopen.com/hosted-document?doi=10.57197/JDR-2025-0012