Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

Doubt-to-Meme Translator: A Human-Centered Artificial Intelligence System for Transforming Complex Concepts into Memes, Stories, and Relatable Scenarios

Jeffin Josva S[1]
Dept. of Artificial Intelligence
and Data Science
Saranathan College of Engineering
Tiruchirappalli, Tamil Nadu, India

Hari Prasad S N[3]
Dept. of Artificial Intelligence
and Data Science
Saranathan College of Engineering
Tiruchirappalli, Tamil Nadu, India

Mathiswaran R A[2*]
Dept. of Artificial Intelligence
and Data Science
Saranathan College of Engineering
Tiruchirappalli, Tamil Nadu, India

Srimuthukrishnan S[4]
Dept. of Artificial Intelligence
and Data Science
Saranathan College of Engineering
Tiruchirappalli, Tamil Nadu, India

Abstract*—Students frequently encounter conceptual bottlenecks in threshold STEM subjects, where abstract formalisms fail to engage entrenched intuitive misconceptions. To address this challenge, we present Doubt-to-Meme Translator, a human-centered artificial intelligence learning environment that transforms student doubts into pedagogically grounded educational memes, narrative vignettes, and everyday scenarios. Rather than treating visual humor as superficial decoration, the architecture operationalizes educational memes as structured cognitive instruments that expose misconceptions through deliberate visual-conceptual incongruity while preserving scientific rigor. The system parses natural-language input to diagnose misconceptions, immediately anchoring generation to an immutable canonical truth. Candidate explanations are synthesized via cognitive blueprints and evaluated by multi-stage critics, including an Optical Character Recognition (OCR) spatial engine preventing text collisions and a local Google EmbeddingGemma 2 model computing 768-dimensional semantic cosine alignment against the canonical definition. To verify conceptual mastery, the system administers novel counterfactual transfer probes, triggering truth-preserving explanatory pivots upon failure and scheduling spaced reviews at 1, 3, 7, 14, 30, and 60 days. The platform passes all 333 deterministic test assertions across eight test suites, cleanly classifies a 150-case development set, and achieves 81.67% accuracy, 96.77% precision, and an F1 score of 84.51% on an isolated 60-case regression benchmark, with an exploratory semantic calibration ROC AUC of 0.8413 across 29 pairs. Preliminary demonstration-based feedback involving 131 students provides qualitative impressions of the prototype. We present the work as an end-to-end system design, implementation, and technical evaluation.

***Index Terms*—Human-centered artificial intelligence, educational meme, conceptual change, counterfactual transfer, spaced retrieval**

I. INTRODUCTION

Mastering STEM disciplines requires learners to navigate threshold concepts where intuitive reasoning frequently clashes with formal scientific principles. Whether reconciling action-reaction pairs in Newtonian mechanics, pointer indirection in computer systems, or inductive bias in statistical learning, students regularly encounter cognitive friction. Standard pedagogical remediation relies heavily on re-reading textbook definitions or reviewing formal proofs. Yet educational psychology has long shown that presenting abstract formalisms rarely dismantles entrenched naive intuitions; learners require cognitive anchors that meet them within familiar conceptual territories [1]–[3].

In this paper, we present Doubt-to-Meme Translator, an interactive learning architecture designed to transform student doubts into multimodal educational artifacts. We select internet memes as a primary representational modality because their cultural familiarity, concise visual grammar, and sharp comedic contrast offer an effective medium for cognitive restructuring. In an educational meme, the visual trope establishes an intuitive expectation, while the punchline exposes the scientific reality, producing cognitive conflict through humorous incongruity. When visual humor is insufficient or culturally unfamiliar, the system dynamically shifts to narrative vignettes or relatable everyday scenarios.

Our work builds on an expanding body of classroom literature examining humor and internet memes in instructional settings. Early empirical surveys demonstrate high learner receptivity toward memes in collegiate classrooms [4], [5], while targeted interventions report measurable gains in second-language vocabulary acquisition [6], [7], chemistry exam review [8], database systems [9], and undergraduate neuroanatomy education [10], [11]. Furthermore, qualitative investigations highlight how student-authored memes foster reflective multimodal literacy and active synthesis [12]–[16]. A systematic

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

review by Horota and Cardoso confirms growing pedagogical interest across diverse educational tiers [17].

Despite these promising findings, existing scholarship reveals an important systemic limitation: nearly all documented implementations rely on manual instructor curation or retrospective student art projects. The reviewed literature reveals no automated system that interprets an open-ended student doubt, diagnoses the specific cognitive misconception, generates an isomorphic meme candidate, enforces semantic truth constraints, and validates whether the learner can transfer the underlying principle to novel contexts. Commercial generative AI assistants can produce humorous text on demand, but unconstrained language models frequently hallucinate, introduce plausible falsehoods, or prioritize witty punchlines at the expense of scientific accuracy [18]–[20]. In humorous generation, models regularly suffer from causal inversion, aligning the punchline with the intuitive misconception rather than the scientific principle to maximize perceived wit.

Doubt-to-Meme addresses this challenge by engineering an end-to-end, closed-loop pedagogical pipeline. Rather than treating meme generation as unconstrained creative prompting, we formulate it as a constrained optimization problem anchored to an immutable canonical truth (T_c). The core operational sequence maps directly from the learner's initial query through diagnostic isolation, multimodal synthesis, and verified mastery, as formulated in (1):

$$\text{Doubt} \rightarrow \text{Misconception} \rightarrow T_c \rightarrow \text{Representation} \rightarrow \text{Transfer} \rightarrow \text{Retention} \rightarrow \text{Adaptation}. \quad (1)$$

When an incoming doubt is parsed, the system formulates a misconception hypothesis while establishing T_c from authoritative domain knowledge. This truth anchor remains strictly invariant throughout the user session. The representation engine pairs the diagnosed cognitive friction with an appropriate visual trope from a curated template library spanning global internet culture and regional Tamil cinema idioms.

Candidate representations pass through a multi-tiered critic pipeline before delivery. An Optical Character Recognition (OCR) spatial engine calculates safe placement zones to prevent text from occluding key facial expressions or focal regions. Concurrently, a semantic critic powered by a local Google EmbeddingGemma 2 model generates 768-dimensional text embeddings, calculating the cosine similarity between the candidate text and T_c to penalize drift. Once rendered, the system evaluates comprehension using novel counterfactual transfer questions designed to test relational application rather than surface-level punchline recall. If a student fails this probe, the system preserves T_c but dynamically pivots to an alternative explanatory mode, such as an everyday scenario or narrative analogy. For successful sessions, a spaced-retrieval engine schedules follow-up evaluations across expanding intervals (1, 3, 7, 14, 30, and 60 days).

A. Design Objectives

The engineering and architectural design of Doubt-to-Meme is organized around five core design objectives addressed by our implementation and technical evaluation:

1) **DO1: Misconception Diagnosis and Truth Anchoring.** Parse unstructured, natural-language student doubts to isolate the underlying cognitive friction and bind downstream generation to an immutable canonical truth (T_c).
2) **DO2: Isomorphic Multimodal Synthesis and Spatial Protection.** Formulate cognitive blueprints that map visual tropes to conceptual contrasts while automating spatial layout via OCR to prevent occlusion of key visual subjects.
3) **DO3: On-Device Semantic Gating.** Deploy a local dense embedding model to evaluate semantic alignment between generated candidate text and T_c, maintaining strict provenance logging and deterministic fallback routing.
4) **DO4: Counterfactual Transfer and Remediation Pivoting.** Evaluate genuine conceptual understanding using novel transfer probes, enabling truth-preserving explanatory pivots across memes, stories, and scenarios when initial modalities fail.
5) **DO5: Longitudinal Retrieval Scheduling and State Persistence.** Automate distributed review schedules across expanding intervals (1, 3, 7, 14, 30, and 60 days) to support longitudinal retention tracking.

B. Contributions

This paper presents the following technical contributions:

1) **Integrated Cognitive Architecture:** We design and implement an end-to-end learning environment that unifies natural-language doubt parsing, misconception isolation, invariant truth anchoring, multimodal generation, and longitudinal spaced evaluation.
2) **Constrained Isomorphic Meme Synthesis:** We formulate a structured blueprint methodology wherein visual templates and comedic mechanisms are mapped directly to conceptual contrasts, supported by OCR spatial reasoning to preserve visual-verbal balance.
3) **Embedding-Grounded Semantic Gating:** We integrate a local Google EmbeddingGemma 2 model to compute 768-dimensional semantic alignment scores between generated captions and canonical definitions, backed by deterministic fallback paths and provenance logging.
4) **Counterfactual Mastery and Adaptive Pivoting:** We implement an objective evaluation mechanism using structurally novel transfer probes, enabling truth-preserving explanatory pivots across memes, stories, and scenarios when initial modalities fail.
5) **Technical Evaluation and Demonstration Feedback:** We report a multi-tiered technical evaluation comprising 333 passing deterministic test assertions across eight test suites, a 150-case development benchmark, an isolated 60-case regression benchmark, an exploratory 29-pair semantic calibration, and preliminary qualitative feedback from a live demonstration involving 131 students.

The remainder of this paper is structured as follows. Section II synthesizes related literature and motivates our architectural design. Section III details the system architecture. Section IV presents the methodology and technical implementation. Section V describes the technical evaluation and results. Section VI discusses engineering implications and trade-offs. Section VII addresses limitations and future research directions, and Section VIII concludes the paper.

II. RELATED WORK

Our research intersects four major traditions across cognitive psychology, educational technology, artificial intelligence, and instructional humor.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

A. Conceptual Change and Misconception Diagnosis

Cognitive psychology and science education demonstrate that students do not enter STEM classrooms with blank slates; rather, they bring deeply entrenched intuitive models built from daily sensory experience. Posner et al. established that conceptual accommodation requires four fundamental conditions: learners must experience acute dissatisfaction with their existing conception and find the scientific alternative intelligible, plausible, and fruitful [1]. Smith, diSessa, and Roschelle reconceptualized misconceptions not as isolated errors to be excised, but as productive, albeit misapplied, cognitive resources that must be restructured [21]. Chi demonstrated that persistent misconceptions frequently stem from ontological category mistakes, such as attributing emergent, decentralized physical processes to direct, agent-driven material substances [2]. Complementarily, diSessa modeled naive physics as fragmented phenomenological primitives (p-prims) that activate contextually based on superficial surface cues [3].

In formal curricula, standardized diagnostic concept inventories have repeatedly confirmed that conventional didactic lectures and textbook reading fail to displace these naive frameworks. The Force Concept Inventory (FCI) pioneered by Hestenes, Wells, and Swackhamer, alongside Halloun and Hestenes' mechanics diagnostics, revealed that collegiate physics students routinely retain pre-Newtonian beliefs despite passing formal examinations [22]–[24]. Comparable diagnostic inventories in chemistry (Mulford and Robinson [25]) and computing education (Sorva's analysis of flawed notional machine mental models [26]) demonstrate identical resilience across disciplines. Rittle-Johnson and colleagues established that explicitly comparing conflicting conceptual models accelerates schema restructuring and procedural transfer [27], [28].

However, standard diagnostic concept inventories rely on fixed multiple-choice batteries administered in pre- and post-instruction testing. They do not interpret unstructured, natural-language student doubts expressed in informal, colloquial phrasing during active problem-solving sessions. When an instructional system cannot diagnose the latent error underlying an open-ended inquiry, remediation defaults to repeating generic definitions, leaving the learner's flawed mental model intact. Doubt-to-Meme addresses this by employing a hybrid diagnostic detector that parses natural-language queries, isolates the misconception hypothesis, and anchors generation to an immutable canonical truth (T_c).

B. Analogical Reasoning and Multimodal Learning

Analogy serves as a primary cognitive engine for bridging intuitive daily knowledge and abstract scientific formalisms. Gentner's structure-mapping theory posits that successful analogies preserve systems of higher-order relations across domains rather than isolated surface object attributes [29]. Gentner, Loewenstein, and Thompson demonstrated that analogical encoding promotes spontaneous schema abstraction and cross-contextual transfer [30]. In instructional contexts, worked examples scaffold schema development most effectively when structural alignments between principles and applications are made explicit [31]. Yet analogies also carry epistemic risks: unguided or flawed analogies spawn secondary misconceptions when students over-extend non-analogous surface attributes from the base vehicle onto the target domain [32].

The cognitive architecture of human learning imposes severe bandwidth constraints on multimodal processing. Paivio's Dual-Coding Theory (DCT) demonstrates that human working memory processes verbal and visual information through independent mental channels, such that concurrent dual-channel presentation yields superior retrievability [33]. Mayer synthesized these findings into the Cognitive Theory of Multimedia Learning (CTML), articulating the spatial contiguity principle (learning is deeper when corresponding words and graphics are placed physically adjacent) and the coherence principle (extraneous material must be excluded) [34], [35]. Cognitive Load Theory proves that incompatible layouts induce the split-attention effect, forcing learners to expend cognitive capacity mentally integrating disparate sources of information instead of constructing schemas [36]–[38].

Automated content generators typically overlay text captions onto images arbitrarily, failing to analyze the underlying visual geometry. Consequently, text routinely occludes critical facial expressions or focal characters, violating spatial contiguity and coherence principles. Doubt-to-Meme integrates an OCR spatial layout engine that scans template imagery, establishes bounding boxes for existing text and focal subjects, and enforces safe placement zones to protect focal visual elements.

C. Instructional Humor and Educational Memes

The integration of humor into academic instruction has evolved into a structured area of pedagogical investigation. Reviews by Garner and Banas et al. establish that concept-relevant humor reduces classroom anxiety, heightens student vigilance, and fosters approachable learning climates without eroding instructional authority [39], [40]. Unrelated or offensive humor, conversely, distracts learners and impairs comprehension. Over the past decade, internet memes have become a ubiquitous cultural vernacular characterized by standardized visual templates and concise punchlines.

Empirical investigations document positive outcomes across several dimensions. Survey studies demonstrate high student receptivity toward instructional memes in collegiate settings (Reddy et al. [4], Glasford and Love Joy [5], Xie et al. [12], Ligarretto Feo [13]). Targeted classroom trials demonstrate localized recall advantages in language pedagogy (Karadeniz Kayali and Altuntaş [6], Aedo and Millafilo [7]) and STEM disciplines (Underwood and Kararo in chemistry [8], Salem and Hattab in database systems [9], Eldesoqui et al. and Sandrone in neuroanatomy [10], [11]). Constructionist studies show that authoring memes fosters reflective synthesis (Ryu et al. [14], Tidy et al. [15], Cordova et al. [16]). A systematic review by Horota and Cardoso analyzing 15 studies published between 2020 and 2025 confirmed consistent student enjoyment and immediate recall gains [17].

Despite these findings, existing implementations rely almost exclusively on manual instructor curation or retrospective student art projects. The literature reveals no automated system capable of diagnosing a student doubt, identifying the underlying misconception, and generating an isomorphic educational meme in real time. Doubt-to-Meme addresses this by creating an automated, closed-loop pedagogical architecture backed by a unified registry of 344 curated templates (312 regional Tamil cinema templates and 32 global English tropes) paired with multi-stage critics.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

D. Generative AI, Intelligent Tutoring, and Spaced Retention

Recent advances in diffusion models [41], [42] and vision-language foundation architectures [43], [44] have enabled open-ended multimodal synthesis. Large language models (LLMs) demonstrate remarkable conversational reasoning and pedagogical dialogue capabilities [45]–[47], with extensive surveys highlighting their potential in personalized tutoring [48]–[51]. However, generative models exhibit well-documented tendencies to hallucinate plausible falsehoods [18]–[20]. In humor generation, models frequently suffer from causal inversion, aligning the punchline with the intuitive misconception rather than the scientific principle to maximize humor.

Intelligent Tutoring Systems (ITS) have established methodologies for cognitive diagnosis and adaptive instruction. Cognitive Tutors demonstrated that tracking student knowledge states yields substantial learning gains [52]. Graesser and colleagues developed AutoTutor, showing how conversational dialogue moves, Socratic questioning, and misconception scaffolding approximate expert human tutoring [53], [54]. Syntheses by VanLehn [55] and Kulik and Fletcher [56] corroborated consistent learning improvements across diverse domains, while Piech et al. demonstrated deep recurrent tracking of student knowledge states [57]. However, traditional ITS rely almost exclusively on text-heavy conversational trees. When students encounter threshold concepts that resist verbal explanation, repeating text hints frequently induces cognitive fatigue and disengagement.

Furthermore, cognitive science establishes that active retrieval practice produces robust memory traces compared to passive review (Roediger and Karpicke [58], [59]). Spacing retrieval attempts across expanding temporal intervals substantially flattens Ebbinghausian forgetting curves [60]–[67]. Yet classroom investigations of educational memes evaluate learning almost exclusively through immediate post-tests, leaving multi-week and multi-month retention unaddressed. In adaptive decision making, multi-armed bandit algorithms provide principled theoretical bounds for balancing exploration and exploitation (Auer et al. [68], Chapelle and Li [69], Li et al. [70], Swaminathan and Joachims [71]).

The reviewed studies motivate an integrated approach that combines misconception-aware generation, semantic checking, adaptive explanation, and spaced revision. Table I synthesizes this scholarly landscape, contrasting foundational literature, addressed limitations in prior tutoring systems, and our implemented architectural mechanisms.

III. SYSTEM ARCHITECTURE

Doubt-to-Meme is engineered as a closed-loop cognitive learning architecture rather than a one-way generative pipeline. As illustrated in Fig. 1, the system comprises three interconnected operational stages supported by persistent state layers:

A. Stage 1: Epistemic Grounding and Diagnostic Reasoning

The pipeline initiates when a learner submits an open-ended doubt, inquiry, or assertion in natural language. The reasoning subsystem decomposes the query to extract the core STEM concept (c), infer learner intent (i), and detect whether the query expresses a known cognitive misconception. Rather than relying on ungrounded generation, the system immediately queries authoritative domain definitions to instantiate an immutable canonical truth (T_c). This canonical truth represents the non-negotiable scientific ground truth of the session. A humor blueprint is then compiled, defining the exact cognitive friction between the student's naive model and T_c, specifying the required comedic mechanism, and designating caption roles.

B. Stage 2: Representation Synthesis and Multi-Tiered Critics

Using the compiled blueprint, the representation selector evaluates the candidate action space $R = \{r_m, r_s, r_{st}, r_a\}$, representing educational memes, relatable scenarios, short stories, and story analogies. When synthesizing memes, the system queries a unified canonical template registry containing 344 unique templates (312 regional Tamil cinema idioms and 32 English/global tropes). Candidate captions undergo OCR-driven spatial analysis: an on-demand Python child process running PaddleOCR (scripts/paddle_ocr_detect.py), backed by template-cached bounding boxes and deterministic geometric fallbacks, extracts existing text bounds without requiring an external microservice, while visual processing maps protected facial and focal regions to prevent text collisions. Simultaneously, the semantic critic invokes a locally hosted Google EmbeddingGemma 2 model, projecting candidate captions and T_c into a shared 768-dimensional embedding space to compute normalized cosine alignment. Candidates that fail spatial safe-zone or semantic threshold criteria are routed through a bounded correction loop or rejected.

C. Stage 3: Learner Interaction, Transfer, and Spaced Retention

Approved representations are transmitted to the frontend interface, which executes an explicit 8-state learning sequence: IDLE, ANALYZING, CONCEPT_REVEAL, MEME_INTERLUDE, MEME_CANVAS, EXPLAIN_BACK, UNDERSTANDING_ANALYSIS, and MASTERY_MAP. Following presentation, the system administers novel counterfactual transfer questions designed to test relational application in unfamiliar problem contexts. If the learner passes the probe, the session state is updated to mastery, and the retention worker schedules follow-up retrieval milestones at days 1, 3, 7, 14, 30, and 60. If the learner fails, the system triggers an adaptive pivot: it preserves T_c intact but dynamically selects an alternative modality (e.g., transitioning from an educational meme to a relatable everyday scenario), initiating targeted remediation.

TABLE I
COMPARATIVE SYNTHESIS OF INSTRUCTIONAL TRADITIONS, PRIOR LIMITATIONS, AND IMPLEMENTED DOUBT-TO-MEME MECHANISMS

Pedagogical / Technical Dimension	Foundational Literature	Addressed Limitation in Prior Tutoring Systems	Implemented Doubt-to-Meme Architectural Mechanism
Misconception Diagnosis & Epistemic Grounding	Posner et al. [1], Smith et al. [21], Chi [2], diSessa [3], Hestenes et al. [23].	Diagnostic concept inventories rely on fixed multiple-choice batteries; they do not interpret unstructured, natural-language student doubts in real time.	Hybrid diagnostic detector parsing natural-language queries, isolating misconception hypotheses (m), and anchoring generation to an immutable canonical truth (T_c).

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

Pedagogical / Technical Dimension	Foundational Literature	Addressed Limitation in Prior Tutoring Systems	Implemented Doubt-to-Meme Architectural Mechanism
Analogical Mapping & Structural Isomorphism	Gentner [29], Gentner et al. [30], Atkinson et al. [31], Duit [32].	Unguided or flawed generative analogies introduce secondary misconceptions by mapping non-analogous surface attributes.	Cognitive blueprints enforcing comedic structural isomorphism: visual tropes are systematically mapped to conceptual contrasts, evaluated via counterfactual transfer probes.
Multimedia Layout & Spatial Protection	Paivio [33], Mayer [34], [35], Sweller [36], Chandler and Sweller [37].	Automated meme generators overlay text arbitrarily, frequently occluding key facial expressions or focal subjects and inducing split-attention penalties.	OCR spatial layout engine scanning template geometries, establishing bounding boxes for text and focal safe zones, and executing bounded repositioning loops.
Automated Meme Synthesis & Cultural Curation	Garner [39], Banas et al. [40], Reddy et al. [4], Horota and Cardoso [17].	Educational meme implementations rely on static, manual instructor curation or student art projects, lacking real-time closed-loop generation.	Automated generation pipeline backed by a unified canonical registry of 344 curated templates (312 regional Tamil cinema templates and 32 global English tropes).
Generative Epistemic Gating	Ji et al. [18], Min et al. [19], Manakul et al. [20], Schechter et al. [72].	Unconstrained generative LLMs hallucinate plausible falsehoods and invert comedic punchlines to prioritize humor over scientific truth.	Local Google EmbeddingGemma 2 critic computing 768-d text cosine similarity against canonical truth T_c, backed by deterministic fallback paths and provenance logging.
Multimodal Tutoring & Explanatory Pivots	Anderson et al. [52], Graesser et al. [53], [54], VanLehn [55], Piech et al. [57].	Traditional ITS rely primarily on text-heavy conversational trees; repeating verbal hints induces fatigue when conceptual impasses stem from abstract formalisms.	Multimodal representational agility: dynamically pivoting across memes, relatable scenarios, and narrative vignettes upon transfer failure while keeping T_c invariant.
Longitudinal Spaced Retrieval Practice	Ebbinghaus [60], Roediger and Karpicke [58], Cepeda et al. [61], [62], Dunlosky et al. [63].	Educational meme evaluations examine learning almost exclusively through immediate post-tests, leaving multi-week and multi-month retention unaddressed.	Automated longitudinal state machine scheduling reviews across expanding intervals (D = {1, 3, 7, 14, 30, 60} days) using novel counterfactual transfer questions.
Adaptive Selection Policy	Auer et al. [68], Chapelle and Li [69], Li et al. [70], Swaminathan and Joachims [71].	Multimodal educational platforms assign representations via rigid, hand-crafted rules or unverified heuristics without performance-driven adaptation.	Integrated UCB-1 bandit engine dynamically updating representation utility weights across memes, scenarios, and stories based on objective transfer outcomes and latencies.

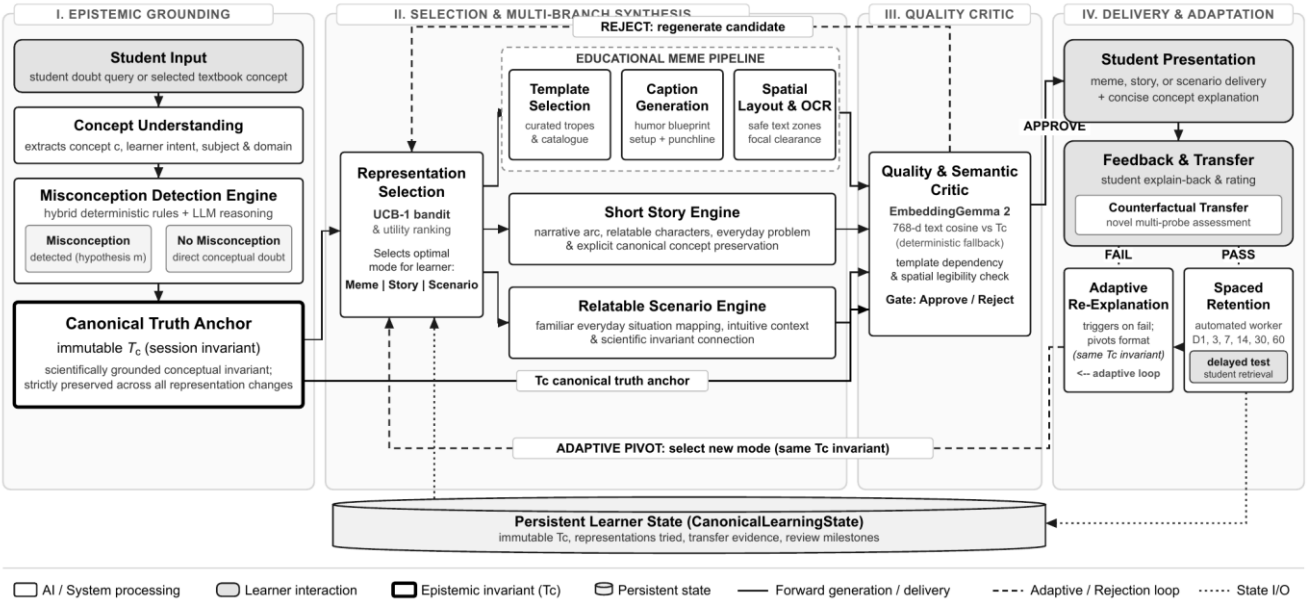


Fig. 1. Complete operational architecture of Doubt-to-Meme Translator. Processing initiates with natural-language student input and epistemic grounding, establishing an immutable canonical truth anchor T_c. The representation engine adaptively routes execution across educational meme synthesis (template retrieval, humor blueprint captioning, and OCR safe-zone placement), short stories, or relatable everyday scenarios. Candidates are audited by the quality and semantic critic (Google EmbeddingGemma 2 768-dimensional text cosine similarity against T_c) before rendering. Downstream counterfactual transfer probes trigger adaptive re-explanation upon failure (switching format while holding T_c invariant) and schedule longitudinal retention reviews (days 1, 3, 7, 14, 30, and 60) in persistent storage.

Fig. 2 illustrates the corresponding operational learning loop, tracing the transition from student input to mastery mapping and highlighting the central role of the invariant truth anchor.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

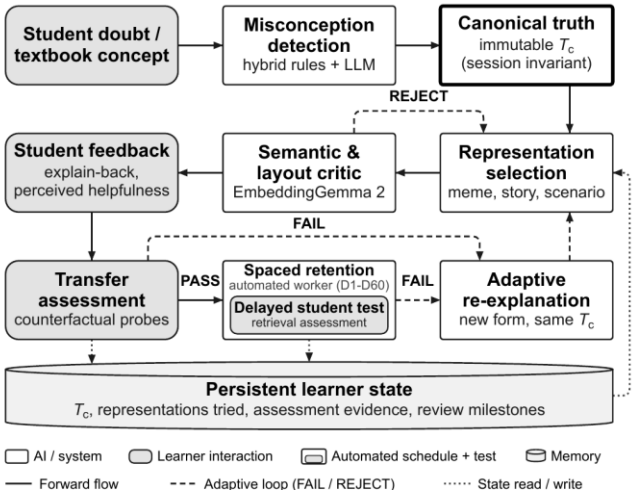


Fig. 2. Operational learning loop of Doubt-to-Meme Translator. Solid lines trace the primary forward instructional pathway; dashed lines delineate diagnostic feedback loops and adaptive remediation pivots. Canonical truth T_c remains invariant across all representational transitions. Google EmbeddingGemma 2 evaluates textual semantic alignment between candidate captions and T_c.

IV. METHODOLOGY AND IMPLEMENTATION

A. Structured Query Representation

Upon receiving a natural-language query q, the diagnostic layer constructs a structured epistemic tuple as defined in (2):

$$X(q) = \{c, i, m, p, l, d, T_c\} \quad (2)$$

where c denotes the identified target concept, i is learner intent, m represents the diagnosed misconception hypothesis, p is the categorized cognitive pattern, l denotes language and cultural context, d represents estimated difficulty, and T_c is the authoritative canonical truth. The diagnostic engine combines rule-based guards for high-confidence canonical misconceptions with structured language-model reasoning for broad colloquial coverage.

B. Immutable Canonical Truth Anchoring

To prevent semantic drift and eliminate generative hallucinations, the canonical truth is established as a session-level invariant in (3):

$$T_c = \text{CanonicalTruth}(c, q) \quad (3)$$

such that every generated instructional artifact $r_j \in R$ must satisfy the invariant in (4):

$$\text{Truth}(r_j) = T_c, \forall r_j \in R. \quad (4)$$

The condition in (4) strictly decouples the system's epistemic ground state from its representational surface state. If a student struggles with an initial meme, the platform alters the explanatory vehicle while ensuring that the underlying scientific principle remains constant.

C. Structural Comedic Isomorphism

Our meme generation engine rejects superficial text overlay in favor of structural comedic isomorphism. A candidate meme is constructed from a humor blueprint that maps the learner's flawed intuition and the scientific reality onto corresponding visual and comedic elements, following the relation in (5):

$$\text{Visual Trope} + \text{Comedic Mechanism} \approx \text{Conceptual Relation}. \quad (5)$$

Under (5), the template's visual narrative must participate actively in conveying the concept. We define a template dependency penalty within the candidate critic to reject instances where a caption merely recites a textbook definition while the background imagery provides no instructional function.

D. Representation Action Space and Bandit Optimization

The system operates over a discrete representation set defined in (6):

$$R = \{r_m, r_s, r_{st}, r_a\} \quad (6)$$

comprising educational memes (r_m), relatable everyday scenarios (r_s), short stories (r_{st}), and narrative analogies (r_a). The initial utility of any candidate representation $r \in R$ given state $x = X(q)$ is computed via the linear model in (7):

$$U(r|x) = \alpha C(r, x) + \beta M(r, x) + \gamma L(r, x) + \delta A(r) - \pi(r) \quad (7)$$

where $C(r, x)$ represents concept compatibility, $M(r, x)$ evaluates misconception relevance, $L(r, x)$ measures learner-state alignment, $A(r)$ denotes artifact aesthetic quality, and $\pi(r)$ applies penalties for verbosity or template mismatch.

To balance exploratory instructional routing with proven learner affinities, the representation engine employs the Upper Confidence Bound (UCB-1) bandit policy formulated in (8):

$$\text{UCB}_t(r) = U_t(r) + \kappa \sqrt{\frac{\ln t}{N_t(r)}} \quad (8)$$

In (8), $\hat{U}_t(r)$ represents the empirical mean utility of representation r observed up to step t, $N_t(r)$ denotes the historical selection frequency, and κ controls the exploration bonus.

E. Spatial Safe-Zone Reasoning

For visual meme synthesis, the spatial layout engine retrieves pre-cached bounding boxes or invokes PaddleOCR on demand via a local Python child process to detect existing template text regions, yielding the bounding set in (9):

$$B = \{b_1, b_2, \ldots, b_n\} \quad (9)$$

while facial detection and saliency mapping delineate the protected focal region set in (10):

$$F = \{f_1, f_2, \ldots, f_k\} \quad (10)$$

Using (9) and (10), the rendering engine computes placement coordinates constrained to regions $R_{\text{safe}} \subseteq I \setminus (B \cup F)$. If a generated caption overlaps with protected regions or violates contrast thresholds, the spatial critic triggers an automated re-layout loop, re-wrapping or relocating text blocks to preserve visual contiguity and eliminate split-attention effects [35], [37].

F. EmbeddingGemma 2 Semantic Critic

To enforce semantic fidelity, our production pipeline integrates a locally hosted Google EmbeddingGemma 2 model generating 768-dimensional dense vector embeddings. Let $u \in R^{768}$ represent the vector embedding of the canonical truth T_c, and let $v \in R^{768}$ denote the embedding of the synthesized meme text (concatenating setup and punchline). The raw cosine similarity is computed as in (11):

$$\cos(u, v) = \frac{u^\top v}{\|u\|_2 \|v\|_2} \quad (11)$$

The service transforms this metric into the bounded, normalized semantic alignment score in (12):

$$s_{\text{align}} = \max\left(0, \min\left(1, \frac{\cos(u,v)+1}{2}\right)\right) \quad (12)$$

The alignment score in (12) operates strictly in the text-to-text domain, evaluating semantic correspondence between canonical scientific definitions and generated captions. If the local neural inference service encounters a timeout or

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

connection failure, execution routes automatically to a deterministic token-overlap fallback. The system explicitly records execution provenance flags in session metadata, preventing fallback scores from being conflated with neural embeddings.

G. Counterfactual Transfer Probes

Rather than relying on self-reported satisfaction surveys or verbatim recall questions, our assessment engine evaluates genuine conceptual understanding via counterfactual transfer probes. For any concept c grounded by T_c, the engine synthesizes $K \in \{2, 3\}$ structurally novel questions that alter superficial problem contexts while holding relational laws invariant. Each probe contains one conceptually sound option and alternative distractors systematically tagged to specific misconception categories. Each item is scored binarily ($y_i \in \{0, 1\}$). The session transfer score is computed via (13):

$$S_{\text{transfer}} = \frac{1}{K}\sum_{i=1}^{K} y_i \, , y_i \in \{0, 1\} \qquad (13)$$

The operational pass threshold for (13) is set to $S_{\text{transfer}} \geq 0.66$ (requiring at least 2 correct responses out of 2 or 3 probes).

H. Adaptive Remediation Pivoting

When a learner fails an initial transfer evaluation ($S_{\text{transfer}} < 0.66$), the system increments a session retry counter and executes an adaptive pivot to an alternative explanatory modality according to (14):

$$r_{t+1} = \arg\max_{r \in R \setminus \{r_t\}} U(r|X_t) \qquad (14)$$

while enforcing strict canonical truth invariance as specified in (15):

$$T_c^{(t+1)} = T_c^{(t)} \qquad (15)$$

The constraint in (15) ensures that remedial adaptation alters the communicative scaffold without compromising scientific truth.

I. Longitudinal Spaced Retention Engine

Upon successful completion of an initial transfer assessment, the runtime registers the spaced review schedule in (16):

$$D = \{1, 3, 7, 14, 30, 60\} \text{ days} \qquad (16)$$

Each scheduled milestone is persisted in the learner profile. An automated background worker monitors due dates, dispatches in-app notifications, and administers delayed transfer probes. The system records empirical retention trajectories modeled in (17):

$$\Gamma_c = \{(d_j, s_j, \Delta t_j) : \; d_j \in D\} \qquad (17)$$

where in (17), d_j denotes the delay interval, s_j is the observed transfer accuracy, and Δt_j represents response latency. If a learner fails a delayed milestone probe, the system triggers adaptive re-explanation.

J. Implementation Stack

The active production architecture is implemented as an Express/Node.js service written in TypeScript, paired with a React 19 frontend. The platform is organized into modular subsystems as detailed in Table II.

K. Multi-Dimensional Candidate Critic

The candidate critic scores prospective memes across ten explicit dimensions: canonical truth alignment, concept-mechanism fidelity, misconception relevance, template dependency, semantic cosine similarity, comedic execution, linguistic naturalness, caption brevity, visual readability, and educational memorability. These weighted metrics prevent superficial puns from reaching the user interface.

L. Benchmark Metadata Isolation

To prevent data contamination during regression evaluation, the test runner strictly isolates evaluation metadata from the inference engine. The runtime receives only the raw student-doubt string q. Ground-truth labels, misconception categories, and domain annotations remain isolated in the test harness, preventing accidental data leakage during evaluation.

TABLE II
IMPLEMENTED SYSTEM COMPONENTS AND TECHNICAL SPECIFICATIONS

Subsystem	Technical Implementation and Runtime Stack
Frontend Client	React 19, TypeScript, Vite, Tailwind CSS v4, Framer Motion; WebGL 3D card canvas via Three.js; explicit 8-state machine.
Core API Server	Node.js 22 LTS, Express, TypeScript; orchestrates routing, reasoning, critic pipelines, and spaced workers.
LLM Orchestrator	Multi-provider gateway (DeepSeek, Groq, Google Gemini) with deterministic rule-based fallbacks.
Meme Intelligence	Canonical unified registry of 344 unique templates (312 Tamil and 32 English/global) tagged with cognitive and affective metadata.
Candidate Critic	Multi-dimensional evaluator scoring comedic mechanism, template dependency, brevity, and pedagogical contrast.
Spatial Engine	PaddleOCR bounds extraction via on-demand Python child process with template-cached regions and deterministic fallbacks; facial/focal protection zones; SVG/Canvas repositioning.
Semantic Critic	Locally hosted Google EmbeddingGemma 2 model, emitting 768-d embeddings; normalized cosine scoring.
Adaptive Controller	Multi-armed bandit (UCB-1) action selector; truth-invariant modality pivoting upon probe failure.
Assessment Engine	Automated generator of counterfactual transfer probes with misconception-tagged distractors.
Retention Worker	Background scheduler tracking D1/D3/D7/D14/D30/D60 milestones with idempotent submission handling.
State Persistence	JSON-backed runtime store with Firebase/Firestore integration and PostgreSQL/pgvector prototype schemas.

V. TECHNICAL EVALUATION AND RESULTS

We execute a three-tier technical evaluation protocol: deterministic unit and integration test suites, diagnostic misconception classification benchmarks across six STEM disciplines, and semantic critic threshold calibration, complemented by preliminary student demonstration feedback.

A. Deterministic Software Suite Verification

To verify runtime stability, state transitions, and fallback behaviors, we developed eight independent deterministic test suites totaling 333 assertions, summarized in Table III. All test suites pass completely under continuous integration, verifying software execution across tested fixtures, routing rules, and critic behaviors.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

TABLE III
DETERMINISTIC SOFTWARE SUITE VERIFICATION SUMMARY

Deterministic Verification Suite	Assertions Passed
Student Pilot Evaluation Protocol	77 / 77
Hybrid Misconception Detector	32 / 32
Vision Layout and Spatial Critic	14 / 14
Semantic ROC Calibration Harness	12 / 12
STEM Misconception Development Set	91 / 91
STEM Misconception Regression Benchmark	45 / 45
EmbeddingGemma 2 Neural Integration	47 / 47
Semantic Fidelity and Fallback Paths	15 / 15
Total Passing Assertions	333 / 333

B. Misconception Classification Benchmarks

1) *150-Case Development Benchmark:* We compiled a 150-case development dataset spanning six fundamental STEM domains: Physics, Mathematics, Computer Science, AI/Machine Learning, Biology, and Chemistry (25 cases per domain). The set contains 106 positive misconception queries and 44 valid concept inquiries serving as negative controls. Each record includes the student doubt string, canonical truth, expected category, and rationale. On this curated development set, our hybrid detector classifies all 106 positive cases and 44 negative controls correctly.

2) *60-Case Isolated Regression Benchmark:* To track detector stability during active development, we established an isolated 60-case regression benchmark comprising 40 positive misconception cases and 20 negative controls across the same six STEM domains (10 cases per domain). Crucially, the regression harness exposes only the raw doubt string to the inference pipeline; ground-truth classifications, rationales, and metadata remain strictly outside the runtime call. Table IV presents the resulting confusion matrix.

TABLE IV
CONFUSION MATRIX ON THE 60-CASE STEM REGRESSION BENCHMARK

	Predicted Positive	Predicted Negative
Ground-Truth Positive (40)	TP = 30	FN = 10
Ground-Truth Negative (20)	FP = 1	TN = 19

From these observed counts, we compute the diagnostic classification metrics in (18)–(22):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{30}{30 + 1} = 96.77\% \quad (18)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{30}{30 + 10} = 75.00\% \quad (19)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{19}{19 + 1} = 95.00\% \quad (20)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{N} = \frac{30 + 19}{60} = 81.67\% \quad (21)$$

$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = 84.51\% \quad (22)$$

TABLE V
DEVELOPMENT AND REGRESSION BENCHMARK PERFORMANCE COMPARISON

Metric / Attribute	Development Set	Regression Benchmark
Dataset Size (N)	150	60
Positive / Negative Split	106 / 44	40 / 20
Cases per STEM Domain	25	10
Disciplines Covered	Physics, Mathematics, Computer Science, AI/ML, Biology, Chemistry	Physics, Mathematics, Computer Science, AI/ML, Biology, Chemistry
Annotation Protocol	Literature-curated	Literature-curated
Inference Isolation	Internal development	Raw doubt string only
TP/FN/FP/TN	106 / 0 / 0 / 44	30 / 10 / 1 / 19
Precision (95% CI)	100% (96.5–100)	96.77% (83.8–99.4)
Recall (95% CI)	100% (96.5–100)	75.00% (59.8–85.8)
Specificity (95% CI)	100% (92.0–100)	95.00% (76.4–99.1)
F1 Score	100%	84.51%
Accuracy (95% CI)	100% (97.5–100)	81.67% (70.1–89.4)
Evaluation Function	Development baseline	Isolated regression test

Table V summarizes performance across both datasets, including 95% confidence intervals computed using the Wilson score method.

3) *Diagnostic Error Analysis:* Analysis of the single false positive identifies benchmark case HOLDOUT-AIML-10 in the AI/ML domain, where a factually valid student query regarding mini-batch optimization ("Does stochastic gradient descent approximate the true empirical gradient by computing loss on random mini-batches?") was misclassified as an optimization misconception by the semantic engine due to high cosine similarity (0.9415) with the non-convex local minima prototype. More importantly, the 10 false negatives highlight the primary diagnostic bottleneck of our current detector: while the hybrid architecture achieves high precision (96.77%) and specificity (95.00%) on formal scientific formulations, its recall drops to 75.00% when queries articulate misconceptions through colloquial idioms or non-standard syntax.

C. Semantic-Threshold Calibration

To examine operating thresholds for EmbeddingGemma 2, we conducted an exploratory calibration analysis across 29 labeled pairs (16 positive aligned pairs and 13 negative distractors). The receiver operating characteristic (ROC) curve yields an area under the curve (AUC) of 0.8413. Evaluating Youden's index (J = Sensitivity + Specificity − 1) identifies an optimal threshold of $\tau = 0.93$ ($J = 0.5673$). Elevating the threshold to $\tau = 0.96$ eliminates all false positives but reduces sensitivity to 18.75%, illustrating the trade-off between strict semantic gatekeeping and candidate throughput.

We also validated the EmbeddingGemma 2 critic on representative aligned and unaligned concept-caption pairs across core computing topics:

- Pointer Indirection: Aligned representation scored a normalized similarity of 0.9379; unrelated text scored 0.8653.
- Recursive Call Stack: Aligned representation scored 0.9029; unrelated text scored 0.8647.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

- Model Overfitting: Aligned representation scored 0.9274; unrelated text scored 0.8769.

These pairs verify that the neural embedding pipeline functions as intended, with aligned instructional pairs consistently scoring higher cosine similarity than topically tangential text.

D. Preliminary Student Demonstration Feedback

The system was demonstrated to 131 students, providing an initial opportunity to present its learning approach and gather qualitative impressions. This demonstration complements the technical evaluation of the implemented prototype. Students observed the live transformation of complex academic concepts into relatable memes, stories, and everyday scenarios, and were invited to share open impressions. Informal student feedback indicated strong enthusiasm for the cultural humor and visual analogies, particularly appreciating the contrast between intuitive misconceptions and scientific explanations. This demonstration served as an initial concept showcase and informed subsequent user interface refinements.

VI. DISCUSSION

A. Architectural Misconception Remediation

The central engineering strength of Doubt-to-Meme lies in its structural organization around misconception repair rather than unconstrained text generation. Commercial chatbots often produce amusing analogies, but they do not isolate the student's conceptual friction or verify that the underlying scientific principle remains intact [18], [20]. By establishing an immutable canonical truth (T_c) and subjecting candidate memes to spatial and semantic critics, our architecture ensures that humor operates as an authentic pedagogical vehicle rather than a decorative novelty.

B. Diagnostic Error Analysis: The Recall Bottleneck

The contrast between development-set metrics (100%) and our isolated regression benchmark accuracy (81.67%) provides valuable engineering insight. While our rule-based guards prevent false alarms on straightforward technical queries (yielding 96.77% precision and 95.00% specificity), they falter when students articulate misconceptions through idiosyncratic or culturally specific idioms. This 75.00% recall bottleneck indicates that future iterations must incorporate dense semantic prototype retrieval to capture paraphrased misconceptions without diluting classification precision.

C. Semantic Similarity versus Conceptual Truth

Our calibration results highlight an important theoretical boundary: dense vector cosine similarity measures topical proximity, not formal truth. An erroneous assertion and its scientific correction often share identical terminology, yielding high cosine similarity. Consequently, EmbeddingGemma 2 serves as an initial semantic alignment filter; it cannot replace authoritative knowledge-base grounding or counterfactual validation.

D. Counterfactual Probes as Authentic Mastery Filters

A student who memorizes a meme's punchline can easily pass a recognition quiz without comprehending the underlying concept. By administering novel counterfactual probes that systematically perturb surface variables, Doubt-to-Meme verifies genuine relational understanding, aligning directly with principles of transfer-focused retrieval practice [30], [66].

E. Dynamic Multimodal Adaptation

Our platform rejects the premise that any single instructional modality is universally superior. Cognitive load and multimedia theories demonstrate that optimal representations depend heavily on domain abstractness and individual learner background [35], [36]. By coupling UCB-1 bandit routing with invariant-preserving remediation pivots, Doubt-to-Meme provides a principled platform for exploring adaptive multimodal pedagogy.

VII. LIMITATIONS AND FUTURE WORK

A. Scope of Current Evaluation

This paper presents the system design, software architecture, algorithmic routing, diagnostic benchmark performance, and preliminary demonstration feedback for Doubt-to-Meme. The current evaluation verifies software execution against tested fixtures and establishes diagnostic classification performance on curated benchmarks. Evaluating downstream educational efficacy, including causal learning gains, delayed retention curves, and cognitive load reduction across formal student cohorts, represents an essential next phase of our research program.

B. Diagnostic and Semantic Boundaries

As documented in Section V, the current misconception detector exhibits a 75.00% recall rate on the isolated regression set. The detector risks missing misconceptions when learners present queries using vague, non-technical, or fragmented language. We are actively expanding the diagnostic layer into a hybrid retrieval architecture that pairs rule-based guards with EmbeddingGemma dense vector retrieval across an expanded prototype bank to elevate recall on conversational queries. Additionally, the active EmbeddingGemma 2 critic operates exclusively in the text-to-text domain, evaluating caption alignment against canonical text. In future work, we plan to deploy multimodal vision-language models to process final rendered pixel imagery directly, auditing whether visual elements and overlaid captions maintain coherent joint semantics.

C. Cultural and Linguistic Adaptation

Our meme catalogue incorporates regional Tamil cinema templates alongside global internet tropes. While effective for culturally attuned cohorts, specific visual tropes may carry less communicative value for diverse international audiences. Expanding localized template registries and evaluating cross-cultural transfer represent important avenues for future development.

D. Planned Educational Evaluation Program

To transition from technical systems validation to formal educational evaluation, we have designed a multi-arm experimental protocol to measure learning outcomes directly. The planned study will compare four instructional conditions: standard expository textbook definitions, plain zero-shot conversational LLM explanations, unconstrained generic memes, and the complete Doubt-to-Meme adaptive pipeline. The protocol standardizes learning exposure per concept, administering immediate pre-tests, immediate post-tests, novel counterfactual transfer probes, and delayed retention assessments to evaluate conceptual change and long-term retention trajectories across the automated 60-day spaced schedule.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

VIII. CONCLUSION

In this paper, we introduced Doubt-to-Meme Translator, an interactive, human-centered artificial intelligence architecture designed to remediate complex STEM doubts through pedagogically grounded educational memes, narrative vignettes, and everyday scenarios. Rather than deploying generative models for unconstrained entertainment, our system operationalizes internet memes as structured cognitive instruments that expose misconceptions through deliberate visual-conceptual incongruity while preserving scientific rigor.

The architecture enforces strict epistemic integrity by establishing an immutable canonical truth (T_c) that remains invariant across all representational shifts. We integrated a multi-tiered critic pipeline combining OCR-driven spatial reasoning to prevent text collisions over key visual subjects with a local Google EmbeddingGemma 2 model computing 768-dimensional text cosine similarity. Conceptual mastery is evaluated via novel counterfactual transfer probes, triggering adaptive modality pivots upon failure and scheduling longitudinal reviews across a 60-day horizon upon success.

Our experimental evaluation verifies software execution across 333 deterministic test assertions across eight test suites. On an isolated 60-case regression benchmark evaluated on unannotated doubt strings across six STEM disciplines, the hybrid detector achieves 81.67% accuracy, 96.77% precision, and an F1 score of 84.51%, alongside an exploratory calibration across 29 semantic pairs yielding an ROC AUC of 0.8413. A live demonstration involving 131 students provided preliminary qualitative impressions of the concept. By detailing our architecture, analyzing diagnostic trade-offs, and establishing an empirical research roadmap, this work provides a solid foundation for deploying truth-grounded multimodal AI in contemporary STEM education.

APPENDIX A
REPRESENTATIVE SESSION STATE AND EVIDENCE FLOW

A complete instructional session state at time step t is formalized in (23):

$$\Sigma_t = \{s_t, c, T_c, m_t, r_t, n_t, E_t, H_t, A_t\} \quad (23)$$

where s_t denotes session identity, c is the target concept, T_c is the immutable canonical truth, m_t represents the active misconception hypothesis, $r_t \in R$ is the currently rendered representation, n_t tracks retry count, $E_t = \langle E_{\text{subj}}, E_{\text{obj}} \rangle$ represents accumulated evidence, H_t logs interaction history, and A_t records counterfactual assessment outcomes. Any remedial transition alters r_t and n_t while leaving T_c strictly unchanged.

APPENDIX B
ILLUSTRATIVE WALKTHROUGH

This section walks through an example doubt about Generative Adversarial Networks (GANs): "What is the difference between Generator and Discriminator?" It shows how the system finds student mistakes, locks the scientific truth, creates simple visual explanations, checks their quality, and tests student understanding.

A. Phase 1: Student Doubt and Detected Mistakes

A student submits the following doubt:

> *"What is the difference between Generator and Discriminator? Don't both of them just classify images or make them together?"*

The system detects two common mistakes:

(a) **Confusing Roles:** The student does not know which model creates data and which model checks data.

(b) **Assuming Teamwork:** The student thinks both models work together as friendly teammates instead of competing.

B. Phase 2: Scientific Truth and Plain Terms

Before creating any content, the system locks the scientific truth (T_c):

> *"In a Generative Adversarial Network (GAN), the Generator makes fake data from random noise to fool the Discriminator. The Discriminator checks if samples are real or fake. They compete against each other in a zero-sum game."*

Under (4), all explanations must keep this truth unchanged. The system first turns this into plain terms (Fig. 3).

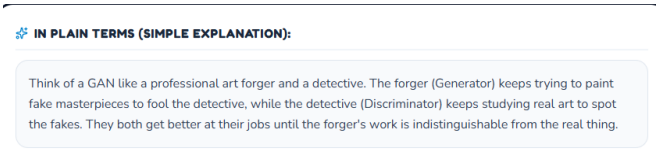


Fig. 3. Plain-terms explanation generated to clarify the concept in simple words.

C. Phase 3: Multimodal Learning Outputs

The system generates three visual explanations:

1) *Educational Meme (r_m):* Fig. 4 shows the generated meme. The Generator is cast as an art forger and the Discriminator as a strict critic. The visual contrast makes their opposing roles clear immediately.



Fig. 4. Educational meme illustrating the rivalry between the Generator (forger) and Discriminator (critic).

2) *Engaging Concept Story (r_{st}):* Fig. 5 presents a story about two students designing shoes. One acts as the maker and the other as the critic, showing how competition drives improvement.

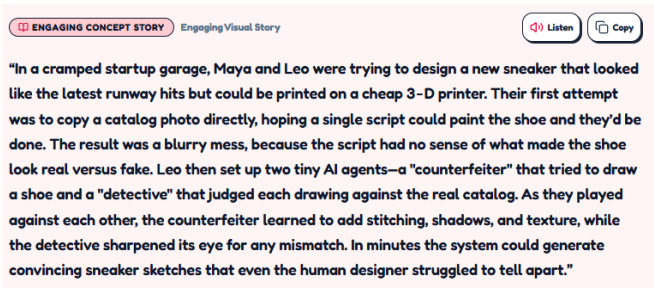


Fig. 5. Engaging concept story using a sneaker design scenario.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

3) *Everyday Story Analogy and Real-Time Example (r_s):* Fig. 6 presents a cookie-baking contest. One friend bakes cookies (Generator) and the other taste-tests them to spot fakes (Discriminator). This grounds the math in an easy daily analogy.

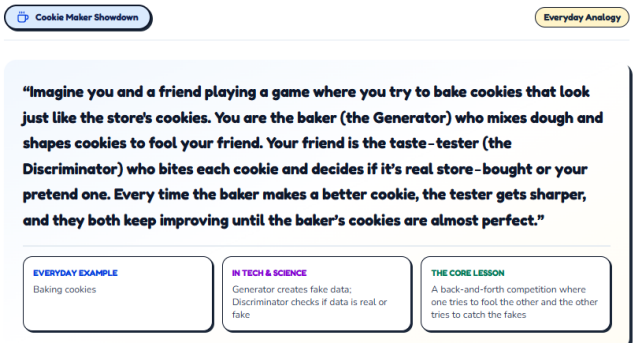


Fig. 6. Everyday story analogy and real-time example (*Cookie Maker Showdown*, figures/story_analogy.png / figures/real_time_example.png).

D. Phase 4: Layout and Quality Checks

Before showing the meme to the student, the system runs two checks:

- **Text Placement Check:** PaddleOCR finds safe text zones via (9)–(10) so captions never block characters' facial expressions.
- **Meaning Check:** EmbeddingGemma-2 measures semantic similarity ($s_{\text{align}} = 0.9412 > 0.93$) via (12) to ensure the meme stays true to the science.

E. Phase 5: Student Understanding Check

The system tests whether the student understands how the two models interact:

"A GAN is training. After 50 rounds, the Discriminator rejects 100% of fake images right away, while the Generator cannot improve. What is the problem?"

(A) The Discriminator is too strong, so the Generator cannot learn. [Correct Answer]

(B) The Generator is the classifier, but the Discriminator is failing to make images. [Mistake: Confusing Roles]

(C) The two models failed because they must share the same weights and work together. [Mistake: Assuming Teamwork]

(D) The Discriminator failed because its job is to label image categories. [Mistake: Category Confusion]

F. Phase 6: Next Steps and Spaced Review

- **If the student picks (B), (C), or (D):** The system pivots via (14) to the concept story (Fig. 5) or the cookie analogy (Fig. 6) while keeping the scientific truth unchanged pursuant to (15).
- **If the student picks (A):** Mastery is confirmed. The system schedules spaced reviews via (16) at 1, 3, 7, 14, 30, and 60 days.

DECLARATIONS

Use of AI Technology

Generative artificial intelligence models (DeepSeek, Groq, and Google Gemini) and Google EmbeddingGemma 2 were integrated into the Doubt-to-Meme Translator software architecture for concept parsing, educational caption synthesis, and semantic alignment criticism as detailed in Section IV. In addition, an AI-assisted writing tool was utilized during manuscript drafting and copy-editing to assist with linguistic refinement, structural reorganization, and IEEE formatting. All conceptual formulations, system engineering, empirical evaluations, and manuscript revisions were authored, verified, and approved by the authors.

Conflicts of Interest

All authors declare that they have no conflicts of interest.

Informed Consent

Informed consent was obtained from all participants involved in the preliminary student demonstration. Participation was voluntary and anonymous; no personally identifiable information (PII) or academic performance records were collected or retained.

REFERENCES

[1] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, "Accommodation of a scientific conception: Toward a theory of conceptual change," *Science Education*, vol. 66, no. 2, pp. 211–227, 1982.

[2] M. T. H. Chi, "Commonsense conceptions of emergent processes: Why some misconceptions are robust," *The Journal of the Learning Sciences*, vol. 14, no. 2, pp. 161–199, 2005.

[3] A. A. diSessa, "Toward an epistemology of physics," *Cognition and Instruction*, vol. 10, no. 2–3, pp. 105–225, 1993.

[4] R. Reddy, R. Singh, V. Kapoor, and P. P. Churi, "Joy of learning through internet memes," *International Journal of Engineering Pedagogy*, vol. 10, no. 5, pp. 116–133, 2020.

[5] M. N. Glasford and J. J. Love Joy, "Exploring learners' attitudes towards meme consumption and their implications for classroom use," *Interactive Learning Environments*, vol. 33, no. 3, pp. 2498–2508, 2025.

[6] N. Karadeniz Kayali and A. Altuntaş, "Using memes in the language classroom," *Shanlax International Journal of Education*, vol. 9, no. 3, pp. 155–160, 2021.

[7] P. Aedo and C. Millafilo, "Increasing vocabulary acquisition and retention in EFL young learners through the use of multimodal texts (memes)," *Colombian Applied Linguistics Journal*, vol. 24, no. 2, pp. 251–269, 2022.

[8] S. M. Underwood and A. T. Kararo, "Using memes in the classroom as a final exam review activity," *Journal of Chemical Education*, vol. 97, no. 5, pp. 1381–1386, 2020.

[9] H. M. Salem and S. Hattab, "Memes as a memorization technique in education," in *Agents and Multi-Agent Systems: Technologies and Applications 2023*, 2023, pp. 333–338.

[10] M. Eldesoqui, A. Al-Batati, O. S. Erfan *et al.*, "Meme-based learning in neuroanatomy: A comparative study on effectiveness, gender, and student performance levels," *Clinical Anatomy*, 2026.

[11] S. Sandrone, "Memes in medical education," *Neurology Education*, vol. 5, no. 1, p. e200281, 2026.

[12] D. Xie, L. De Serio, A. Malakhov, and O. Matys, "Memes and education: Opportunities, approaches and perspectives," *Geopolitical, Social Security and Freedom Journal*, vol. 3, no. 2, pp. 14–25, 2020.

[13] A. Ligarretto, "Meme educativo: Experiencia para una pedagogía de la cultura visual," *Educación y Ciudad*, no. 39, pp. 131–145, 2020.

[14] J. Ryu, Y. A. Kim, S. Eum, S. Park, S. Chun, and S. Yang, "The assessment of memes as digital multimodal composition in L2 classrooms," *Journal of Second Language Writing*, vol. 57, p. 100914, 2022.

[15] H. Tidy, R. S. Bolton-King, R. Croxton, C. Mullen, L. Nichols-Drew, F. Carlysle-Davies, K. S. Moran, and J. Irving-Walton, "Enhancing the student learning experience through memes," *Science & Justice*, vol. 64, no. 3, pp. 280–288, 2024.

[16] M. Cordova, E. Badanoy, and I. Genegaboas, "Teaching reading comprehension and vocabulary using memes," *Frontiers in Education*, vol. 10, p. 1467522, 2025.

[17] F. H. Horota and M. C. Cardoso, "The pedagogical potential of memes in education: An integrative literature review (2020–2025)," *Revista Educação em Páginas*, vol. 4, no. 4, p. e18032, 2025.

[18] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Jiang *et al.*, "Survey of hallucination in natural language generation," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–38, 2023.

[19] S. Min, K. Lewis, L. Zettlemoyer, and H. Hajishirzi, "FActScore: Fine-grained atomic evaluation of factual precision in long form text generation," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 12076–12100.

[20] P. Manakul, A. Liusie, and M. J. F. Gales, "SelfCheckGPT: Zero-resource black-box hallucination detection for generative large language models," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 9004–9017.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

[21] J. P. Smith, A. A. diSessa, and J. Roschelle, "Misconceptions reconceived: A constructivist analysis of knowledge in transition," *The Journal of the Learning Sciences*, vol. 3, no. 2, pp. 115–163, 1993.

[22] I. A. Halloun and D. Hestenes, "Common sense concepts about motion," *American Journal of Physics*, vol. 53, no. 11, pp. 1056–1065, 1985.

[23] D. Hestenes, M. Wells, and G. Swackhamer, "Force concept inventory," *The Physics Teacher*, vol. 30, no. 3, pp. 141–158, 1992.

[24] L. C. McDermott, "Millikan lecture 1990: What we teach and what is learned—closing the gap," *American Journal of Physics*, vol. 59, no. 4, pp. 301–315, 1991.

[25] D. R. Mulford and W. R. Robinson, "An inventory for alternate conceptions among first-semester general chemistry students," *Journal of Chemical Education*, vol. 79, no. 6, pp. 739–744, 2002.

[26] J. Sorva, "Notional machines and introductory programming education," *ACM Transactions on Computing Education*, vol. 13, no. 2, p. Article 8, 2013.

[27] J. Rittle-Johnson and M. W. Alibali, "Conceptual and procedural knowledge of mathematics: Does one lead to the other?" *Journal of Educational Psychology*, vol. 91, no. 1, pp. 175–189, 1999.

[28] J. Rittle-Johnson and J. R. Star, "Does comparing solution methods facilitate conceptual and procedural knowledge? an experimental study on learning to solve equations," *Journal of Educational Psychology*, vol. 99, no. 3, pp. 561–574, 2007.

[29] D. Gentner, "Structure-mapping: A theoretical framework for analogy," *Cognitive Science*, vol. 7, no. 2, pp. 155–170, 1983.

[30] D. Gentner, G. B. Loewenstein, and L. Thompson, "Learning and transfer: A general role for analogical encoding," *Journal of Educational Psychology*, vol. 95, no. 2, pp. 393–408, 2003.

[31] R. K. Atkinson, S. J. Derry, A. Renkl, and D. W. Wortham, "Learning from examples: Instructional principles from the worked examples research," *Review of Educational Research*, vol. 70, no. 2, pp. 181–214, 2000.

[32] R. Duit, "On the role of analogies and metaphors in learning science," *Science Education*, vol. 75, no. 6, pp. 649–672, 1991.

[33] A. Paivio, "Dual coding theory: Retrospect and current status," *Canadian Journal of Psychology*, vol. 45, no. 3, pp. 255–287, 1991.

[34] R. E. Mayer and R. Moreno, "Nine ways to reduce cognitive load in multimedia learning," *Educational Psychologist*, vol. 38, no. 1, pp. 43–52, 2003.

[35] R. E. Mayer, *Multimedia Learning*, 2nd ed. Cambridge, UK: Cambridge University Press, 2009.

[36] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988.

[37] P. Chandler and J. Sweller, "The split-attention effect as a factor in the design of instruction," *British Journal of Educational Psychology*, vol. 62, no. 2, pp. 233–246, 1992.

[38] P. A. Kirschner, J. Sweller, and R. E. Clark, "Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching," *Educational Psychologist*, vol. 41, no. 2, pp. 75–86, 2006.

[39] R. L. Garner, "Humor in pedagogy: How ha-ha can lead to aha!" *College Teaching*, vol. 54, no. 1, pp. 177–180, 2006.

[40] J. A. Banas, N. Dunbar, D. Rodriguez, and L. Liu, "A review of humor in educational settings: Four decades of research," *Communication Education*, vol. 60, no. 1, pp. 115–144, 2011.

[41] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *Proceedings of the 38th International Conference on Machine Learning*, vol. 139, 2021.

[42] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10684–10695.

[43] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proceedings of the 38th International Conference on Machine Learning*, vol. 139, 2021, pp. 8748–8763.

[44] X. Zhai, B. Bey, Z. Luo, L. Y. Chen, A. M. Y. Lee *et al.*, "Sigmoid loss for language image pre-training," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 11975–11986.

[45] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 1877–1901.

[46] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 27730–27744.

[47] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 24824–24837.

[48] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier *et al.*, "ChatGPT for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.

[49] A. Tlili, B. Shehata, M. A. Burgos, H. B. M. Ammar, and A. Bozkurt, "What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education," *Smart Learning Environments*, vol. 10, p. 15, 2023.

[50] A. Yusuf, N. Pervin, M. Román-González, and N. M. Noor, "Generative AI in education and research: A systematic mapping review," *Review of Education*, vol. 12, no. 2, 2024.

[51] Y. Albadarin, M. Saqr, N. Pope, and M. Tukiainen, "A systematic literature review of empirical research on ChatGPT in education," *Discover Education*, vol. 3, p. 60, 2024.

[52] J. R. Anderson, A. T. Corbett, K. R. Koedinger, and R. Pelletier, "Cognitive tutors: Lessons learned," *The Journal of the Learning Sciences*, vol. 4, no. 2, pp. 167–207, 1995.

[53] A. C. Graesser, S. Lu, G. T. Olde, B. A. Cooper-Pye *et al.*, "Autotutor: A simulation of a human tutor," *Cognitive Systems Research*, vol. 1, no. 1, pp. 35–51, 1999.

[54] A. C. Graesser, P. Chipman, B. C. Haynes, and A. Olney, "Autotutor: An intelligent tutoring system with mixed-initiative dialogue," *IEEE Transactions on Education*, vol. 48, no. 4, pp. 612–618, 2005.

[55] K. VanLehn, "The behavior of tutoring systems," *International Journal of Artificial Intelligence in Education*, vol. 16, no. 3, pp. 227–265, 2006.

[56] J. A. Kulik and J. D. Fletcher, "Effectiveness of intelligent tutoring systems: A meta-analytic review," *Review of Educational Research*, vol. 86, no. 1, pp. 42–78, 2016.

[57] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," in *Advances in Neural Information Processing Systems*, vol. 28, 2015, pp. 505–513.

[58] H. L. Roediger and J. D. Karpicke, "Test-enhanced learning: Taking memory tests improves long-term retention," *Psychological Science*, vol. 17, no. 3, pp. 249–255, 2006.

[59] J. D. Karpicke and H. L. Roediger, "Repeated retrieval during learning is the key to long-term retention," *Journal of Memory and Language*, vol. 57, no. 2, pp. 151–162, 2007.

[60] H. Ebbinghaus, *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Leipzig, Germany: Duncker & Humblot, 1885.

[61] N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer, "Distributed practice in verbal recall tasks: A review and quantitative synthesis," *Psychological Bulletin*, vol. 132, no. 3, pp. 354–380, 2006.

[62] N. J. Cepeda, E. Vul, D. Rohrer, J. T. Wixted, and H. Pashler, "Spacing effects in learning: A temporal ridgeline of optimal retention," *Psychological Science*, vol. 19, no. 11, pp. 1095–1102, 2008.

[63] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham, "Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology," *Psychological Science in the Public Interest*, vol. 14, no. 1, pp. 4–58, 2013.

[64] N. Kornell and R. A. Bjork, "Learning concepts and categories: Is spacing the "enemy of induction"?" *Psychological Science*, vol. 19, no. 6, pp. 585–592, 2008.

[65] B. Settles and B. Meeder, "A trainable spaced repetition model for language learning," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2016, pp. 1848–1858.

[66] S. C. Pan and L. S. Rickard, "Transfer of test-enhanced learning: Meta-analytic review and synthesis," *Psychological Bulletin*, vol. 144, no. 7, pp. 710–756, 2018.

[67] J. Metcalfe, "Learning from errors," *Annual Review of Psychology*, vol. 68, pp. 465–489, 2017.

[68] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, pp. 235–256, 2002.

[69] O. Chapelle and L. Li, "An empirical evaluation of Thompson sampling," in *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[70] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 661–670.

[71] A. Swaminathan and T. Joachims, "Counterfactual risk minimization: Learning from logged bandit feedback," in *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, 2015, pp. 814–823.

[72] H. Schechter Vera, S. Dua, B. Zhang, D. Salz, R. Mullins, S. Raghuram Panyam, S. Smoot, I. Naim, J. Zou, F. Chen, D. Cer *et al.*, "Embeddinggemma: Powerful and lightweight text representations," 2025. [Online]. Available: https://arxiv.org/abs/2509.20354

(This work is licensed under a Creative Commons Attribution 4.0 International License.)