Feynman - A Novel Conversational Framework for On-Demand & Pedagogical Video Generation Using Gemini & Manim

Sabyasachi Chakraborty, Tabish Naqvi, Sasmita Pala
School of Engineering & Technology
Manav Rachna International Institute of Research & Studies
Faridabad, Haryana, India
manavrachna.edu.in

Abstract - *The rapid evolution of artificial intelligence (AI) & computational media has redefined how learners engage with educational content. "Feynman," a novel conversational framework, integrates Google's Gemini multimodal AI with the Manim animation engine to deliver on-demand, dynamically generated pedagogical videos. This system aims to simulate the clarity & curiosity-driven learning approach popularised by physicist Richard Feynman, offering intuitive visualisations that deconstruct complex concepts into accessible learning experiences. This paper explores the design, implementation, & pedagogical potential of the Feynman framework within the context of AI-assisted education. Through a comprehensive literature review, we analyse the current state of AI-driven educational media, the effectiveness of visual learning & conversational tutoring, & the technical affordances of tools like Manim in educational animation. The findings indicate that combining large multimodal models with algorithmic animation engines can significantly enhance conceptual comprehension, learner engagement, & personalised pedagogy. Limitations, ethical considerations, & implications for future research in AI education are discussed, positioning Feynman as a pioneering model for next-generation learning systems.*

Keywords: Artificial Intelligence, Feynman, Conversational Framework, Gemini, Manim, Pedagogical Videos, Dynamically Generated Content, Intuitive Visualisations, Conceptual Comprehension, Visual Learning, AI-Assisted Education, Conversational Tutoring, Learner Engagement, Personalised Pedagogy, Large Multimodal Models

1 Introduction

This research investigates Feynman, a conversational AI framework that autonomously generates animated educational videos through the integration of Gemini, Google's large multimodal model (LMM), & Manim, an open-source mathematical animation engine. The central premise of the Feynman framework is to democratise access to conceptual clarity by allowing users—students or educators—to request any topic in natural language, which the system then interprets, scripts, & renders as a concise, visually rich explanation. Unlike traditional static AI tutors or pre-recorded lessons, Feynman delivers real-time visual pedagogy, transforming conversational input into dynamic explanatory animations.

1.1 Background

Education has always evolved alongside advances in communication technology—from oral traditions to print, radio, television, and, more recently, digital & interactive media. In the current era, Artificial Intelligence (AI) is revolutionising education yet again by personalising learning, automating instruction, & visualising complex phenomena. However, much of AI-assisted learning remains text-based or reliant on static interfaces that inadequately capture the dynamic, visual, & experiential aspects of human cognition. The emergence of multimodal large models—such as OpenAI's GPT-4, Google's Gemini[7], & Anthropic's Claude—has opened new frontiers for education, enabling systems that can interpret & generate text, images, audio, & video coherently. Concurrently, the Manim (Mathematical Animation Engine) community[17] has advanced algorithmic visualisation as a form of mathematical storytelling, capable of representing abstract or high-dimensional concepts through dynamic motion, geometry, & symbolic reasoning. "Feynman," named after the Nobel-winning physicist renowned for his intuitive & visual approach to explaining physics[6], proposes a conversational pedagogy framework combining these two frontiers. The system allows learners to ask questions conversationally (powered by Gemini) & receive animated explanations (rendered via Manim) in real time.

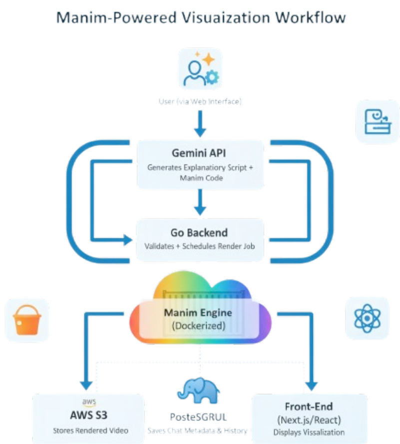

Fig 1— Presents the overall workflow of the proposed Feynman framework, illustrating how conversational queries are transformed into pedagogical video explanations through the integration of Gemini and Manim.

1.2 Problem Statement

While modern AI tutoring systems such as Khanmigo, Duolingo Max, & ChatGPT-4 Tutor offer personalised dialogue, they primarily operate in a linguistic mode. They lack the *visual synthesis* required to intuitively convey spatial, relational, & mathematical reasoning. This gap between semantic comprehension & visual representation often hinders students' understanding of complex systems—particularly in STEM education

Therefore, there is a pressing need for an integrative system that:

- Dynamically translates conversational input into pedagogical visualisations,
- Provides multimodal learning experiences aligned with cognitive principles,
- Supports both educators (in creating illustrative content) & students (in personalised comprehension).

1.3 Research Objectives

This research aims to:

1. Theorise the pedagogical & technical architecture of the *Feynman framework*.
2. Examine the role of *Gemini* & *Manim* as synergistic components for AI-driven educational visualisation.
3. Evaluate, through literature synthesis, how such multimodal systems align with established theories of learning.
4. Identify the potential impact of this approach on educational accessibility, engagement, & comprehension.
5. Discuss current limitations & propose directions for empirical & technological advancement.

1.4 Hypothesis

> *The integration of a multimodal conversational AI (Gemini) with an algorithmic visualisation engine (Manim) can significantly enhance conceptual comprehension & learner engagement by transforming abstract, text-based explanations into dynamic, visual learning experiences.*

1.5 Significance of the Study

This research is significant to multiple audiences:

- Educators gain insights into automated yet pedagogically sound video content creation.
- Students experience accessible & personalised learning support.
- Researchers find a foundation for exploring human–AI co-creation in education.
- Developers can adapt the framework for domain-specific visual learning environments (e.g., medicine, economics, engineering).

By bridging conversational AI & computational animation, *Feynman* represents a paradigmatic shift from "AI as text tutor" to "AI as visual instructor."

2 Literature Review

2.1 AI in Education: From Text Tutoring to Multimodal Learning

The application of artificial intelligence in education (AIED) has evolved from rule-based expert systems in the 1980s[21] (Woolf, 1992) to adaptive learning platforms such as *ALEKS*, *Carnegie Learning*, & *Khanmigo*. These systems personalise pacing, assess performance, & deliver feedback. However, most rely on textual or symbolic modalities[12] (Luckin et al., 2016).

Recent work in *multimodal AI*[22][10] (Zhai, 2023; Huang et al., 2024) highlights how large models trained on heterogeneous data (text, image, code, video) can enable more *naturalistic & contextually grounded* learning experiences. Multimodal models like Gemini[7] (Google DeepMind, 2024) extend this by allowing flexible integration of linguistic & visual reasoning, which opens possibilities for on-dem& explanatory video generation.

The rise of Generative Al since 2023 has further accelerated this trend. Systems are no longer just adaptive; they are creative. They can co-author essays, generate synthetic data for problems, and, as we propose, create instructional content from scratch. How-ever, this generative capability brings new challenges, particularly in ensuring factual accuracy and pedagogical appropriateness[19]. While LMMs can generate explanations, they struggle with "grounding" these explanations in precise, procedural visual domains.
They can **describe** an animation but cannot inherently **render** it with mathematical precision. This creates the need for tools-like Manim-that can algorithmically instantiate visual metaphors consistent with pedagogical design.

2.2 Visual Learning & Cognitive Theories

Research in educational psychology consistently supports the cognitive benefits of visual learning. Allan Paivio's Dual-Coding Theory (1986)[14] posits that information is better retained when processed through both verbal & visual channels. Similarly, Mayer's Cognitive Theory of Multimedia Learning (2001)[13] outlines principles such as *modality*, *redundancy*, & *coherence* that explain how properly designed multimedia improves comprehension.

Sweller's Cognitive Load Theory (1988) further suggests that instructional design should minimise extraneous cognitive load while maximising germane processing—an objective achievable through concise animations. Studies (e.g., Höffler & Leutner, 2007[9]) demonstrate that well-structured animations can outperform static diagrams for procedural & spatial learning tasks.

The Feynman framework embodies these principles: using AI to generate tailored, context-specific animations that complement linguistic explanations, thereby engaging both processing channels without overwhelming the learner.

2.3 Pedagogical Animation & the Role of Manim

Manim, originally developed by Grant Sanderson (creator of *3Blue1Brown*), is a Python-based library for creating mathematically precise animations. Its programmatic control over geometry, timing, & transformations allows for explanatory sequences that closely align with cognitive & pedagogical theories[17] (Sanderson, 2018).

In academic contexts, Manim has been used to produce visualisations across physics (e.g., quantum mechanics), mathematics (e.g., calculus, linear algebra), & computer science (e.g., algorithms). Studies on educational animation tools[5] (Fiorella & Mayer, 2018) show that *narrative-guided visualisation*—as implemented in Manim videos—facilitates deeper learning through episodic coherence & causal reasoning.

However, creating such content manually requires technical & artistic expertise. This bottleneck limits scalability. The Feynman framework seeks to automate this process: Gemini interprets the learner's query, generates a pedagogically structured script, & translates it into executable Manim code—closing the loop between concept & visual narrative.

2.4 Conversational Interfaces & Socratic Pedagogy

Conversational agents have been increasingly explored as tutors that scaffold metacognition & self-explanation. The Socratic tutoring paradigm[8] (Graesser et al., 2005) emphasises inquiry-based dialogue, encouraging learners to articulate reasoning rather than passively consume information. AI models like ChatGPT & Gemini, capable of maintaining contextual dialogue, effectively replicate this paradigm.

When integrated with visualisation tools, such agents can support *reflective iteration*: the learner asks, observes the animation, & refines their understanding through follow-up questions—mirroring the recursive process central to the Feynman Technique[6] (Feynman, 1988).

2.5 The 'Feynman Technique' and Cognitive Psychology

The framework's name is not merely honorary; it is operational. The "Feynman Technique" is a mental model for learning that involves four steps:

- **Step 1 (Choose Concept)**: The user provides the prompt.
- **Step 2 (Teach it)**: The AI (Gemini) generates the simple explanation and (Manim) the intuitive visualisation, effectively "teaching" it back to the user as if they were a novice.
- **Step 3 (Identify Gaps)**: By observing the generated video, the user can immediately spot where the explanation is confusing or where their mental model diverges from the visualisation. The conversational interface then allows them to **query** these gaps directly ("Why did that line curve?").
- **Step 4 (Review and Simplify)**: The system, upon receiving feedback, iterates on the explanation and visualisation, simplifying or elaborating as needed.

This aligns with cognitive science principles like self-explanation [4], where generating explanations for oneself (even if prompted by an AI) enhances learning, and desirable difficulty [1], where the iterative process of refining understanding is more effective than passive consumption.

2.6 Synthesis of Gaps in the Literature

Despite growing research on multimodal AI & pedagogical animation, several gaps persist:

1. Lack of dynamic coupling between conversational explanation & real-time visual rendering.
2. Limited automation in generating pedagogically coherent animations from natural language.
3. Insufficient empirical validation of AI-generated visualisations' impact on conceptual understanding.
4. Ethical & epistemic concerns about content accuracy, cognitive overload, & bias in generated explanations.

The Feynman framework addresses these gaps by combining semantic reasoning (Gemini) with computational visualisation (Manim), enabling coherent, adaptive, & transparent educational media generation.

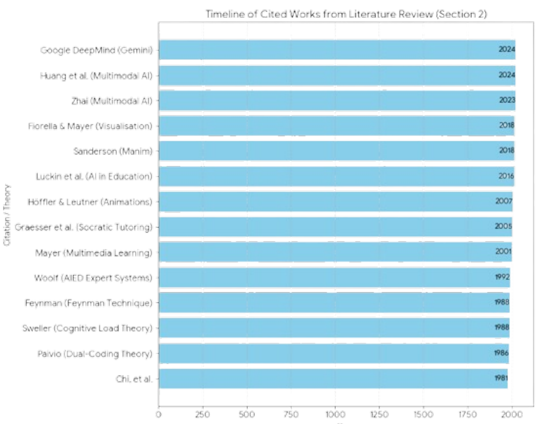

Fig 2: Chronological Progression of Key Research in AI-Assisted Education and Visualisation.

3 Methodology

3.1 Research Design

The present study adopts a *conceptual-analytical research design* supported by a *system-architecture analysis* and *pedagogical mapping* approach. Since "Feynman" is a proposed framework integrating emerging AI and visualisation technologies, this paper does not rely on primary empirical experimentation but instead draws upon an extensive synthesis of existing educational technology methodologies, multimodal AI paradigms, and computational visualisation frameworks.

The methodology involves:

1. Conceptualising the Feynman Architecture based on the functional capabilities of Gemini and Manim.
2. Mapping each system component to established pedagogical and cognitive learning theories.
3. Conducting a comparative analysis between traditional AI tutoring models and multimodal, visually augmented approaches.
4. Proposing evaluation metrics for future empirical validation, including engagement, comprehension, and retention outcomes.

This mixed-conceptual design bridges *technical system analysis* and *pedagogical rationale*, aligning with design-based research methods used in educational innovation studies[16] (Reeves, 2006).

3.2 Framework Architecture

The Feynman Framework is structured as a pipeline that converts *natural language dialogue* into *pedagogically optimised animated video*. Its architecture consists of five modular stages:

1. Conversational Understanding (Gemini)
   The learner initiates a query (e.g., *"Explain the concept of entropy in thermodynamics with visuals."*).
   Gemini interprets intent, contextualises difficulty level, and generates a structured semantic graph outlining:
   - Core concepts
   - Sub-concepts and relationships
   - Analogies or metaphors suitable for visual illustration
2. Pedagogical Script Generation
   Using Gemini's reasoning capabilities, a didactic narrative script is generated based on cognitive scaffolding principles. Each segment follows:
   - *Concept Introduction → Progressive Elaboration → Visual Reinforcement → Summary Recall.*
3. The system draws from *Bloom's taxonomy* (1956) to align content complexity with the learner's level, ensuring that elementary learners receive intuitive analogies, while advanced users get formal mathematical derivations.
4. Manim Code Synthesis Gemini then compiles the pedagogical script into Manim-compatible Python code, specifying scene structure, camera movement,

and object transformations.
Example mapping:

- Abstract concept → Animated geometric object
- Temporal process → Sequential transformation
- Causal relationship → Directed motion or morph

5. This stage translates *semantic meaning* into *algorithmic animation logic*.

6. Visual Rendering (Manim Engine)
Manim executes the code to produce vector-based animations, ensuring mathematical precision and scalability. The rendered output maintains temporal coherence with the narration script, allowing for *synchronous multimodal delivery*.

7. Conversational Feedback Loop
Once the video is displayed, the learner can query further ("What happens if entropy decreases?"), prompting an iterative refinement of both script and visuals.
This recursive dialogue embodies *constructivist learning*[15], where understanding evolves through questioning and visualisation.

3.3 Pedagogical Alignment

Each module of the framework was mapped to a relevant educational theory to ensure the design is pedagogically grounded:
This theoretical scaffolding ensures that the system not only automates visualisation but also *reflects a learning science logic* in its generative behavior.

Feynman Module	Pedagogical Foundation	Theoretical Support
Conversational Understanding	Socratic Method & Metacognition	Graesser et al. (2005); Chi et al. (1994)
Script Generation	Scaffolding & Bloom's Taxonomy	Vygotsky (1978); Bloom (1956)
Code Synthesis	Dual-Coding & Cognitive Load Theory	Paivio (1986); Sweller (1988)
Visual Rendering	Multimedia Learning Principles	Mayer (2001); Höffler & Leutner (2007)
Feedback Loop	Constructivism & Self-Explanation	Piaget (1954); Feynman (1988)

Table 1 — Pedagogical Alignment

3.4 Comparative Methodology

To evaluate Feynman conceptually, it is compared against three existing models:
This comparison highlights that Feynman uniquely merges *adaptivity*, *visualisation*, and *dialogue*—dimensions that have historically remained isolated.

Model	Primary Modality	Adaptivity	Visualisation	Interactivity	Example
Traditional AI Tutor (GPT-4, Khanmigo)	Text	High	Low	Conversational	Chat-based Q&A
Multimedia MOOC (Coursera, Khan Academy)	Video	Low	Hight	Non - Conversational	Pre-rendered video
Feynman Framework	Text + Visual + Audio	High	Dynamic (Manim)	Conversational + Iterative	AI-generated animation

Table 2 — Comparative Table

3.5 Evaluation Metrics (Proposed)

Future empirical testing of Feynman's impact should assess:

1. Comprehension Gains – Pre/post testing on abstract concept mastery.
2. Engagement Metrics – Learner retention time, query depth, and interaction recurrence.
3. Cognitive Load Reduction – Measured using Paas' Cognitive Load Scale (1992).
4. Conceptual Transfer – Learner ability to apply understanding in novel contexts.
5. Educator Utility – Efficiency in content creation and reusability of generated visuals.

These dimensions offer a balanced assessment framework for both student learning and educator facilitation.

3.6 System Design

The **Feynman system design** follows a **modular micro-services architecture** emphasising scalability, real-time interaction, and fault isolation.

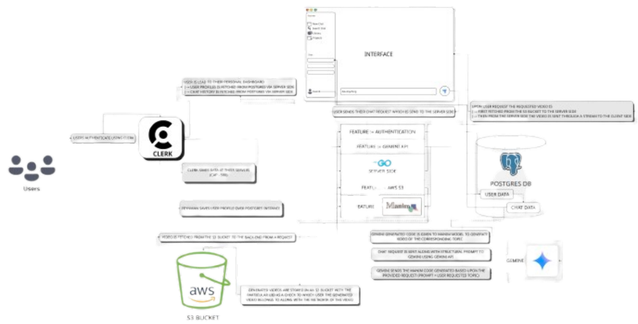

Fig 1 — System Design : Feynman

Core Components:

1. **User Interface Module:**
Provides real-time chat, onboarding, and video display. It interacts with the API Gateway through RESTful endpoints.
2. **Request Processing Module:**
Hosted on the Go Server Side, this component handles Gemini API calls, Manim code validation, and render job scheduling.
3. **Rendering Engine:**
The Manim engine runs in a sandboxed environment (Dockerized Manim runtime) to ensure isolated video rendering tasks.
4. **Storage Module:**
AWS S3 handles object storage, while PostgreSQL stores relational user data.
5. **Session Manager:**
Manages per-user chat rooms similar to conversational AI systems (ChatGPT, Claude), enabling users to revisit their visualised concepts.

Scalability and Deployment:

- Client Side deployed on Vercel while Server Side deployed on Digital Ocean while AWS S3 for video storage **containerisation (Docker)**.
- **Load balancers** manage multiple concurrent requests.

- Manim render tasks are queued using **background workers** to handle computational peaks efficiently.

3.7 System Architecture

The Feynman framework is implemented as a distributed, modular system combining a **React-based Client Side** , a **Go-based Server Side**, and integrated **cloud infrastructure** for computation, rendering, and data persistence. Figure 1 (schematic representation) outlines the complete system flow.

Front - end (Client Layer)

- Built using **Next.js and React**, providing a conversational interface similar to ChatGPT.

Fig 2 — Technologies

- **User authentication** and onboarding are handled through **Clerk**, enabling quick access and persistent user sessions.

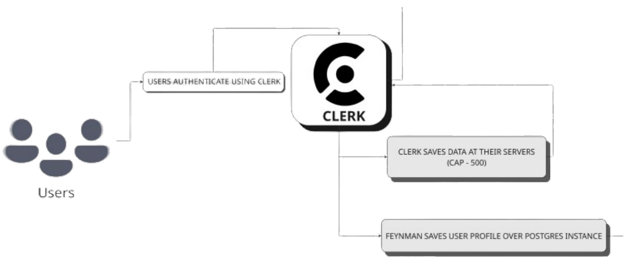

Fig 3 — Authentication

- Users input a topic or question via a **chat interface**; the conversation is displayed alongside generated video responses.

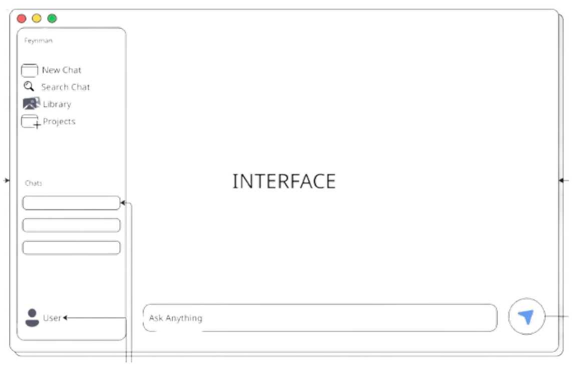

Fig 4 — Chat Interface

- Visualisations are embedded within the chat stream through a dynamic video renderer that fetches outputs from AWS S3.

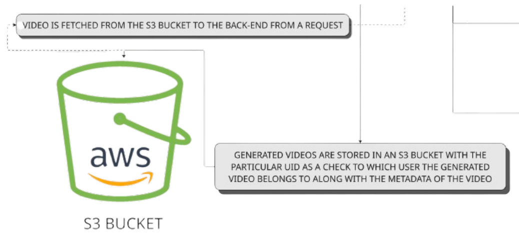

Fig 5 — AWS S3 Service

Backend (Application Layer)

- Implemented in **Go (Golang)** for high concurrency and efficient API orchestration.

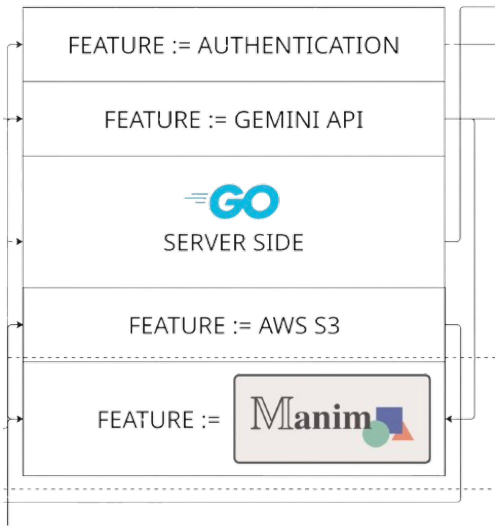

Fig 6 — Server Side Layout

Upon receiving a user query, the Server Side:

1. Sends a **POST request** to the **Gemini API** with contextual metadata (user level, topic domain, and response constraints).
2. Receives structured text and **Manim-compatible Python code** generated by Gemini.
3. Executes a **Manim render job** within a containerised environment (e.g., Docker) to produce an `.mp4` animation.
4. Uploads the rendered video to **AWS S3** for storage and retrieval.
5. Stores chat metadata (query, timestamp, video URL) in a **PostgreSQL database** for session continuity.

Data Layer (Persistence & Storage):

- **PostgreSQL** stores user profiles, chat histories, and references to generated visualisations.

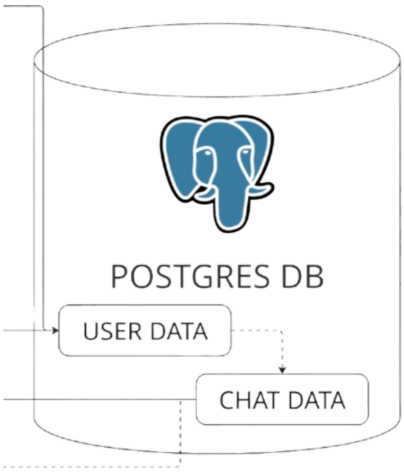

Fig 7 — Postgres DB

- **AWS S3** serves as the media storage layer, ensuring scalability, redundancy, and quick retrieval of pedagogical videos.

Fig 8 — AWS S3

Interaction Flow Summary:

1. *User → (Topic Input) → Client Side*
2. *Client Side → (API Request) → Go Server Side*
3. *Server Side → Gemini (for script & Manim code generation)*
4. *Server Side → Manim Renderer (animation creation)*
5. *Renderer → AWS S3 (video upload)*
6. *Server Side → Client Side (returns S3 video URL + metadata)*
7. *User → Views visualisation and continues dialogue*
8. *Chat & visual history saved for later retrieval.*

4 Findings & Discussion

4.1 Conceptual Findings from Literature Integration

Synthesising across AI, cognitive psychology, and computational visualisation research, several conceptual findings emerge:

1. **Multimodality Enhances Learning Outcomes.**
   Studies by Mayer (2001) and Moreno (2006) confirm that combining verbal and visual channels fosters stronger mental model construction and long-term retention. Feynman's architecture operationalises this through synchronised narration and animation.

2. **Conversational Interfaces Foster Deep Engagement.**
   Graesser et al. (2005) found that learners using Socratic AI tutors engage in longer, more reflective learning dialogues. The iterative feedback loop in Feynman amplifies this dynamic by coupling *dialogue* with *visual feedback*.

3. **Automated Animation Bridges Conceptual Gaps.**
   Prior systems (e.g., autoDraw, DeepMotion) attempted text-to-visual rendering, but lacked pedagogical coherence. Feynman's Manim integration ensures semantic fidelity through computational geometry, producing visuals consistent with mathematical laws.

4. **Cognitive Load is Mitigated by Temporal Coherence.**
   Animations generated on demand minimise extraneous load by targeting precisely the learner's query, unlike pre-made videos that contain irrelevant segments.

4.2 Implications for Educational Practice

4.2.1 For Educators

Feynman transforms teachers from content producers into *curation facilitators*. Instead of spending hours designing lecture visuals, educators can describe the desired concept conversationally and instantly generate animated teaching aids. This aligns with teacher-augmentation paradigms (Luckin, 2018), where AI expands pedagogical bandwidth rather than replacing human instructors.

4.2.2 For Students

Learners gain agency through *interactive visualisation*. Rather than passively viewing fixed tutorials, they can explore "what-if" scenarios. For instance:

> *"Show me what happens to the wave function if I change the potential energy curve."*

Such dialogue-based simulation develops **epistemic curiosity** (Loewenstein, 1994), reinforcing self-directed learning habits critical for lifelong education.

4.2.3 For Educational Institutions

At the institutional level, Feynman can democratise access to high-quality learning materials by automating the production of discipline-specific visualisations. Universities could embed the system within LMS platforms, allowing faculty to instantly produce concept-specific explainer videos tailored to course outcomes.

4.3 Technical and Pedagogical Synergy

The Feynman framework exemplifies a synthesis between symbolic reasoning (Gemini's language modeling) and procedural computation (Manim's animation engine). This integration demonstrates a new genre of AI pedagogy where reasoning and representation coevolve in real time.

Gemini functions as a *semantic compiler*—translating cognitive structures (concepts, analogies, explanations) into executable code that embodies those structures visually. This mechanism echoes Feynman's original teaching philosophy: "If you can't visualise it, you don't understand it."

The resulting pedagogical experience embodies three characteristics:

1. Transparency – The visual representation exposes the logic of the explanation.
2. Adaptivity – The visuals change as the learner's understanding evolves.
3. Agency – Learners influence both the direction and visualisation of the learning experience.

4.4 Challenges & Limitations

Despite its promise, several limitations and challenges must be acknowledged:

1. **Content Accuracy and Pedagogical Validity**
   Automated generation risks conceptual inaccuracies or pedagogical oversimplification. Continuous human oversight is essential, particularly for nuanced topics.

2. **Computational and Temporal Constraints**
   Real-time video rendering via Manim can be resource-intensive. GPU acceleration and scene caching strategies are necessary for scalability.

3. **Cognitive Overload Risk**
   Poorly paced or overly complex animations could increase extraneous load rather than reduce it. Empirical testing on animation pacing and narration alignment is required.

4. **Ethical and Bias Considerations**
   Since Gemini's responses depend on training data, representational biases or cultural oversights could propagate into educational content.
   Adherence to AI ethics in education frameworks (UNESCO, 2023) is mandatory.

5. **Lack of Longitudinal Studies**
   There is minimal empirical evidence on long-term learning effects of dynamically generated visuals compared to curated ones. Controlled studies will be critical.

4.5 Cognitive and Theoretical Implications

From a theoretical standpoint, Feynman introduces a hybrid learning mode situated between:

- **Constructivist learning**, where knowledge is built through dialogue, and

- **Cognitive multimedia learning**, where understanding is structured through dual-channel processing.

The learner's brain, under this system, constantly toggles between *language-driven abstraction* and *image-driven concretisation*. This dual reinforcement fosters **semantic integration**—a hallmark of expert-level understanding (Chi et al., 1981).

Thus, Feynman may operationalise the *Feynman Learning Technique* itself—where explaining a concept simply and visually reveals and corrects knowledge gaps.

4.6 Potential for Interdisciplinary Expansion

While Manim's strength lies in STEM visualisation, the Feynman framework could extend to:

- Humanities – Animated historical timelines, linguistic evolution maps.
- Social Sciences – Econometric or sociological model visualisations.
- Medical Education – Dynamic simulations of physiological processes.

By abstracting "conceptual structure" into "visual motion," Feynman provides a universal pedagogy adaptable to any field requiring conceptual clarity.

4.7 Component Analysis

4.7.1 Gemini

Gemini, Google DeepMind's multimodal AI, forms the **semantic reasoning core** of the Feynman framework. It interprets natural language queries, generates stepwise conceptual explanations, and produces Python scripts compatible with Manim's syntax.
Key functions:

- **Contextual Understanding:** Gemini interprets user intent and educational level.
- **Pedagogical Structuring:** Converts raw explanations into a didactic flow aligned with learning theories (e.g., introduction → elaboration → visual summary).
- **Code Generation:** Outputs syntactically correct and pedagogically aligned Manim code segments (Scene, Transform, Animation, etc.).
- **Iterative Feedback:** Based on subsequent user queries, Gemini can refine the existing script or expand upon subtopics, maintaining coherence across interactions.

4.7.2 Manim

Manim (Mathematical Animation Engine) acts as the **visual computation layer**. It executes the Python code generated by Gemini to render scenes, geometric transformations, or symbolic relationships.
Manim's role includes:

- **Mathematical Accuracy:** Maintains spatial and numerical precision for equations, transformations, and graphs.
- **Scene Composition:** Combines objects (e.g., Tex, Graph, Line) and transitions to visualise conceptual relationships.
- **Narration Synchronisation:** The generated video aligns visually with Gemini's verbal explanation, forming a multimodal output.
- **Scalability:** Through programmatic control, Manim can dynamically adapt visuals to varying complexity levels, ensuring cognitive coherence.

Together, Gemini and Manim represent **cognitive-semantic integration**—where AI language reasoning and computational visualisation jointly produce coherent, explainable, and pedagogically optimised media.

5 Conclusion

5.1 Summary of Findings

This study explored *Feynman*, a conceptual and technological framework that merges conversational AI (Gemini) and algorithmic animation (Manim) to create on-demand, pedagogical video explanations.

Through literature synthesis and theoretical analysis, we found that:

- The fusion of multimodal dialogue and procedural visualisation aligns strongly with cognitive and constructivist theories of learning.
- Dynamic animations enhance comprehension by reducing cognitive load and increasing learner engagement.
- Conversational feedback loops support iterative conceptual refinement and personalised learning trajectories.
- Educators benefit from scalable content generation, while learners gain agency in exploring complex concepts visually.

In sum, the Feynman framework represents a paradigm shift in educational technology—transforming AI from a passive text tutor into an *active visual collaborator.*

5.2 Limitations

Several limitations temper these findings:

- Lack of empirical testing prevents conclusive validation of learning outcomes.
- Current dependency on computationally intensive rendering limits real-time deployment.
- Ethical, cultural, and pedagogical oversight mechanisms must be strengthened to ensure content integrity.
- The framework's effectiveness for non-STEM disciplines remains to be systematically assessed.

5.3 Theoretical and Practical Implications

Theoretical Implications:
Feynman bridges previously disconnected paradigms—AI dialogue systems and educational animation—into a cohesive multimodal learning theory. It demonstrates that *explanatory reasoning* can be visualised procedurally, advancing cognitive theories of AI-assisted understanding.

Practical Implications:

- Educational institutions can integrate the framework into digital classrooms for just-in-time visual teaching.
- Developers can extend the model into plug-ins for LMS or API-based educational tools.
- Teachers can rapidly generate custom animations aligned with lesson objectives, enhancing classroom dynamism.

5.4 Recommendations for Future Research

1. **Empirical Evaluation:**
   Conduct longitudinal studies measuring conceptual comprehension and cognitive load when using AI-generated visual explanations versus traditional lectures.

2. **Pedagogical Optimisation:**
   Develop algorithms for adaptive pacing, controlling the density and speed of animations based on learner feedback.

3. **Interdisciplinary Testing:**
   Expand applications beyond STEM to humanities and social sciences, evaluating cross-domain transferability.

4. **Ethical Governance Framework:**
   Establish institutional guidelines for AI content verification, learner data protection, and transparency in generated media.

5. **Human–AI Co-Creation Studies:**
   Explore how educators and AI systems can collaboratively design instructional media, blending human intuition with algorithmic creativity.

5.5 Final Reflection

The *Feynman Framework* stands at the intersection of **artificial intelligence**, **education**, and **visual communication**—fields that together redefine how humans learn, teach, and visualise knowledge. By operationalising the spirit of Richard Feynman's dictum— *"What I cannot create, I do not understand"*—this system aspires to make understanding not just a linguistic event but a *visual experience*.

The educational future envisioned by Feynman is one where learning becomes a living dialogue between thought and visualisation—between the human mind and the machine that helps it see.

References

- Bloom, Benjamin S. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Longman, 1956.
- Chi, Michelene T. H., Paul J. Feltovich, and Robert Glaser. “Categorisation and Representation of Physics Problems by Experts and Novices.” *Cognitive Science* 5, no. 2 (1981): 121–152.
- Fiorella, Logan, and Richard E. Mayer. *Learning as a Generative Activity: Eight Learning Strategies that Promote Understanding*. Cambridge University Press, 2018.
- Feynman, Richard P. *Surely You’re Joking, Mr. Feynman!* W. W. Norton & Company, 1988.
- Graesser, Arthur C., et al. “AutoTutor: A Cognitive System That Simulates a Tutor–Student Dialogue.” *Cognitive Systems Research* 6, no. 4 (2005): 297–313.
- Höffler, Tim N., and Detlev Leutner. “Instructional Animation versus Static Pictures: A Meta-Analysis.” *Learning and Instruction* 17, no. 6 (2007): 722–738.
- Luckin, Rosemary, et al. “Intelligence Unleashed: An Argument for AI in Education.” Pearson, 2016.
- Mayer, Richard E. *Multimedia Learning*. Cambridge University Press, 2001.
- Paivio, Allan. *Mental Representations: A Dual Coding Approach*. Oxford University Press, 1986.
- Reeves, Thomas C. “Design-Based Research: A Practical Approach to Educational Research.” *Educational Research and Development* 45 (2006): 29–46.
- Sanderson, Grant. *The Manim Community Developer Guide*. 3Blue1Brown, 2018.
- Sweller, John. “Cognitive Load During Problem Solving: Effects on Learning.” *Cognitive Science* 12, no. 2 (1988): 257–285.
- UNESCO. *AI and Education: Guidance for Policy-Makers*. UNESCO, 2023.
- Vygotsky, Lev S. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, 1978.
- Woolf, Beverly P. *Building Intelligent Interactive Tutors: Student-Centred Strategies for Revolutionising E-Learning*. Morgan Kaufmann, 1992.
- Zhai, Xiaoming. “ChatGPT for Next-Generation Science Learning.” *Science Education* 107, no. 3 (2023): 691–698.