

AirCanvas: A Multimodal Framework for Spatial Mathematical Reasoning

16 January 2026

Independent Research Group
Quantum Tiger

Technical brief for educators & public sector stakeholders

Executive Summary

The integration of physical intuition with digital computation remains a central challenge in Human-Computer Interaction (HCI). While digital tools are ubiquitous, traditional input modalities (keyboards/mice) impose a cognitive load that separates mathematical formulation from representation.

AirCanvas introduces a framework leveraging real-time computer vision and Multimodal Large Language Models (MLLMs) to solve mathematical problems via free-space hand gestures. By utilizing the MediaPipe framework for skeletal tracking and Gemini models for reasoning, we demonstrate a pipeline that lowers the barrier to complex mathematical exploration. This paper analyzes the system's geometric heuristics, kinematic smoothing, and thermodynamic efficiency, proposing a new paradigm for AI-assisted education.

1 The Input Bottleneck: Context & Challenge

The history of computing—from punched cards to GUIs—has sought to reduce friction between human thought and machine execution. However, a significant “Input Bottleneck” persists in STEM domains.

- **The Translation Cost:** Writing complex integrals or chemical structures in digital formats requires syntax knowledge, interrupting mathematical flow.
- **The Symbol Grounding Problem:** Typing $x^2 + y^2 = r^2$ involves unrelated keystrokes. “Drawing” a circle in the air restores the motor component of learning (embodied cognition), creating a feedback loop between physical action and abstract reasoning.

Policy Implication: To democratize STEM education, we must move beyond symbolic input to **Spatial Computing**, where the user's physical environment becomes the primary input modality.

2 System Architecture & Sovereign Compute

AirCanvas operates as a hybrid edge-cloud system, balancing responsiveness with reasoning depth.

2.1 The Local Vision Cortex (Edge)

- **Privacy & Speed:** Runs at 30 FPS on local hardware to preserve the illusion of direct manipulation.
- **Technology:** Uses MediaPipe for 21-point skeletal landmarks.
- **Gesture State Machine:** A robust Finite State Machine (FSM) distinguishes between Drawing, Hovering, and Clearing with > 95% precision.
- **Kinematic Smoothing:** Implements Exponential Moving Average (EMA) and adaptive filtering to handle physiological tremors (8–12 Hz) without haptic feedback.

2.2 The Cloud Reasoning Engine

- **Visual Chain-of-Thought:** Bypasses brittle OCR. Raw pixel data of the “drawn” problem is sent to multimodal models (Gemini).
- **Zero-Shot Prompting:** The system uses structured prompting to Transcribe → Identify Branch → Solve Step-by-Step → Provide Answer.

3 Strategic Analysis: Thermodynamics & Efficiency

As AI scales, energy consumption becomes a critical national constraint. AirCanvas demonstrates a thermodynamic optimal via a split-compute architecture.

- **Local Efficiency:** Tracking on local CPU consumes \approx 5–10 Watts.
- **Cloud Cost:** Heavy inference queries cost significantly more (\approx 3600 Joules).
- **The Hybrid Advantage:** Offloading reasoning while keeping perception local minimizes total entropy. Running VQA models locally would spike inference time to 10–20s and cause thermal throttling.

Sustainability Insight: This hybrid model aligns with sustainable AI goals, ensuring high-performance educational tools do not require prohibitive local hardware energy budgets.

4 Pedagogical Implications & Recommendations

4.1 Constructivism & Spatial Reasoning

Aligned with Seymour Papert’s Constructionism, AirCanvas turns math into a constructive act. Students “build” equations in space, fostering intuitive grasp of geometric properties that static text cannot convey.

4.2 Reducing Cognitive Load

By mapping directly to the user’s mental model, AirCanvas reduces “extraneous load” (the difficulty of the tool), freeing working memory for “intrinsic load” (the difficulty of the subject).

4.3 Accessibility & Inclusion

For students with fine motor impairments (e.g., dysgraphia), large gross-motor arm movements offer a viable alternative to precise pen/keyboard control.

4.4 Future Roadmap

- **Volumetric Drawing (3D):** Leveraging depth/LiDAR for 3D structures.
 - **Collaborative Spaces:** Networked “air whiteboards” for remote teacher-student problem solving.
 - **Haptic Integration:** Ultrasonic feedback to simulate “touching” the virtual canvas.
-

Authors

Chirag Nahata, Snigdha Ghosh, Somyadip Ghosh

Independent Research Group

Quantum Tiger

Research Partner

Quantum Tiger

A Morning Bay Technology Pvt Ltd Company

P-175, (A-1/1), Diamond Park Joka, Amgachi, South 24 Parganas, Kolkata, West Bengal, India, 700104

Email: hello@quantumtiger.in | Web: <https://quantumtiger.in>