

Thesis Project Portfolio

Benchmarking AI Agents' Creativity in Dynamic Virtual Worlds

(Technical Report)

Competing Visions: Narratives of AI's Role in Early Childhood Education

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Yifan Zhou

Spring, 2025

Department of Computer Science

Table of Contents

1. Sociotechnical Synthesis
2. Benchmarking AI Agents' Creativity in Dynamic Virtual Worlds
3. Competing Visions: Narratives of AI's Role in Early Childhood Education
4. Prospectus

Sociotechnical Synthesis

Introduction

My academic interest lies in the application layer of artificial intelligence, particularly its implications for education. Within early childhood education, creativity serves as a foundational component of cognitive development. While my capstone project, *Benchmarking AI Agents' Creativity in Dynamic Virtual Worlds*, focused on the technical capabilities of generative AI models embedded in Minecraft as a testbed, my STS thesis, *Competing Visions: Narratives of AI's Role in Early Childhood Education*, analyzed the sociotechnical negotiations among educators, parents, tech companies, and policymakers attempting to define AI's place in learning. Both the capstone project and the STS thesis allowed me to explore the evolving role of artificial intelligence in education from two distinct but complementary angles: one experimental and system-based, the other social and discursive. Insights from the technical project offer empirical grounding and technical possibilities for social debate, while sociotechnical discourse analysis enriches our understanding of how AI products are received, challenged, and reshaped in society. Through this dual lens, I aim to contribute to a more integrative and socially responsive approach to AI development in education. AI is not merely a tool; it is a co-constructed and negotiated presence whose meaning and impact are shaped by both technical design and discourse.

Technical Project

My capstone project benchmarks the creativity of AI agents powered by large language models (LLMs) in the dynamic virtual world of Minecraft. Inspired by projects like Mindcraft and Mineflayer, it explores how four LLMs—GPT-4o, Claude, Gemini, and LLaMA—perform in three structured creative tasks: building a house, decorating a house, and designing a garden.

The models received three types of prompts—basic, instructive, and chain-of-thought—to stimulate varied outputs. These tasks were framed using psychological frameworks from cognitive science, namely the Torrance Tests of Creative Thinking (TTCT) and the Consensual Assessment Technique (CAT), and evaluated along three key metrics: originality, appropriateness, and aesthetic appeal. Human participants ranked each model’s outputs to assess perceived creativity across tasks. Results revealed that GPT-4o demonstrated strong task alignment and practicality, Claude produced the most original and visually appealing designs, while Gemini and LLaMA showed lower consistency and creative coherence. The study also introduced a sentiment analysis experiment, where AI agents and humans collaboratively selected building blocks for different architectural styles. A high degree of overlap in choices demonstrated the AI’s alignment with human aesthetic judgment and strength in sentimental matching. Most participants expressed surprise and strong interest in the human-AI collaboration within the Minecraft creative setting. While limitations remain in contextual creativity and action execution, the findings suggest a promising role for AI in supporting creative expression and co-design in education and interactive environments.

STS Project

My STS research paper explores how different stakeholders—educators, parents, tech companies, and policymakers—compete to define artificial intelligence (AI)’s role in early childhood education. Using discourse analysis and the Social Construction of Technology (SCOT) framework, following the AI development timeline, the study traces how these groups construct competing narratives around AI’s purpose, risks, and legitimacy. Drawing on a wide range of sources (including policy documents, media articles, marketing content, and online forums), the research reveals: how educators emphasize developmental appropriateness and

ethical concerns; parents focus on safety and data privacy; tech firms highlight innovation and personalization; and policymakers seek to balance innovation with accountability. Historically, AI in education began as a technocratic theme driven by researchers and scholars, but with the rise of generative models like ChatGPT, it has become a contested and widely visible topic for everyone. Today, AI is no longer a neutral tool but a socio-technical artifact whose meaning is constantly negotiated across sectors and redefined along the technology timeline. This paper concludes that AI's role is never a fixed entity. While AI offers powerful possibilities, its successful integration in early education depends on transparent policies, cross-sector dialogue, and a shared commitment to supporting children's cognitive and emotional development.

Conclusion

Working on these two projects concurrently allowed me to engage with AI from both a builder's and a critic's perspective. In my capstone, I witnessed participants forming emotional bonds with the AI agent—calling it “Steve” and expressing trust, curiosity, and reduced loneliness. This revealed the public's hopeful imagination of AI companionship in learning. Yet, while developing the system—coding, calling APIs, and structuring prompts—I encountered technical barriers that made me realize how inaccessible such experiences still are to everyday users. This gap between public expectation and technical reality highlighted the need for product designers to act as bridges, making complex systems truly user-friendly and inclusive. At the same time, my STS research revealed that concerns around AI in education extend far beyond technical performance. Parents worry about privacy, screen addiction, and emotional well-being—issues not easily captured by technical metrics. I also came to see how commercial forces quickly absorb and reshape educational AI: I discovered a startup offering a product nearly identical to my capstone, but monetized for entertainment, with little attention to

educational or creative alignment. These experiences reminded me that AI tools are shaped not only by innovation but also by ethics, market pressures, and social feedback. By doing both projects together, I developed a deeper understanding that building responsible and trustworthy AI means not only optimizing what it can do, but also continuously reflecting on what it should do—and for whom.