Bess Hagan

2025 SURF Reflection Essay

This summer, I worked with Dr. Jacob Schrum on his *Procedural Content Generation (PCG) with Generative AI* research project, exploring how modern AI models can be used for PCG. My work was split across two codebases: *MarioDiffusion*, a PyTorch-based repository for training models on Super Mario Bros (SMB) levels, and *LLMTinker*, a Flask web app for a multiplayer storytelling game powered by Large Language Models (LLMs). In *MarioDiffusion*, I trained and evaluated both conditional and unconditional diffusion models with UNet architectures, implemented a cross-entropy loss function and early stopping, developed methods to track, visualize, and compare each model's A* solvability, and created batch scripts for data preparation, model training, and metrics visualization. I also added new GUI features for mixed-initiative systems that allow designers to build larger levels from diffusion-generated scenes. In *LLMTinker*, I built and tested a browser-based multiplayer game where users collaborate on stories guided by LLMs with diffusion-generated illustrations based on user input. I implemented features such as secret player goals and a win condition based on narrative coherence. Throughout SURF, I tracked progress through daily reports and GitHub Issues, contributed to an AIIDE-25 paper submission, created a video demo for one of our mixed-initiative GUI tools, presented our results through slides at MAD MONDAY, and prepared a research poster for a future presentation this fall.

My SURF research built directly on the foundation I developed through coursework at Southwestern University and prior research experience through the Distributed Research Experience for Undergraduates (DREU) program with Dr. Luc Paquette at the University of Illinois Urbana-Champaign and the INVITE Institute. I used Python extensively in both my DREU research and a machine learning course I took last fall, which gave me the confidence to dive into large, unfamiliar codebases and contribute meaningfully to this collaborative project. While the collaboration during SURF echoed my experiences with paired programming projects in my computer science courses, this program challenged me to navigate version control, shared repositories, and debugging in a much more complex manner. My work in *MarioDiffusion* expanded my understanding of AI and machine learning concepts and reinforced my prior experience visualizing data using Matplotlib. I also drew on skills I developed while building my personal portfolio on GitHub when designing and debugging the web interface in *LLMTinker*.

The research we conducted has the potential to support ongoing work in PCG, particularly as it relates to applying generative AI techniques in game design. Our team explored a wide range of conditional and unconditional methods for generating content

and compared them against MarioGPT, a recent model for generating SMB levels. While our models were competitive, we found that none clearly outperformed MarioGPT, and this outcome is itself a meaningful contribution. To make our work accessible to others, we documented our findings in a paper submitted to the AIIDE-25 conference and shared a preprint on arXiv.org. We also released our codebase publicly on GitHub. These resources may support future researchers in evaluating or extending PCG methods and help developers experiment with diffusion models, large language models, or mixed-initiative tools for content creation.

Through SURF, I gained both collaborative and technical skills that will be valuable in future academic and professional settings. Working in a shared codebase helped me strengthen my communication skills, especially when coordinating through GitHub, resolving merge conflicts, and contributing to open issues. I became more comfortable using the command line, writing batch files, and working in Anaconda-based coding environments. I also deepened my understanding of software development practices such as version control and issue tracking. On the technical side, I developed a much stronger grasp of diffusion models, UNet architecture, and masked language modeling, all of which are increasingly relevant in generative AI research. I also gained experience building a local area network game using Flask, Socket.IO, and LLMs, skills applicable to both web development and interactive AI systems. Perhaps most importantly, I learned how to use AI tools to navigate and contribute to large, unfamiliar codebases efficiently, which is a valuable skill in any software engineering or data science role.

I plan to graduate this December and pursue a career in the tech industry as a data analyst, data scientist, or software engineer. This research experience has strengthened my interest in working with AI systems and data-driven tools, and it has given me a clearer picture of how I can contribute to real-world projects in those areas. It also made me more confident in my ability to learn new technologies quickly, collaborate on complex codebases, and communicate technical ideas to others. While I am currently focused on entering the workforce, this experience has also made me more open to the possibility of graduate school in the future, especially if I find a program that builds on the kind of work I enjoyed this summer.