Sam Weese

Dr. Annexstein

CS 5001

September 10th, 2023

Assignment 3, Individual Capstone Assessment

The project I would like to complete is using a large language model (LLM) to effectively generate Dungeons and Dragons content with visual components using stable diffusion, which is informed and handled wholly by the LLM. This is quite an unbounded feat, as the nature of LLM's being new makes this project much much closer to research than development. To nuance my goals a bit, I would like the final project to be able to generate simple sprites as described by the LLM (simple meaning a common object, such as an apple or tree, with a simple modifier, such as changing color). As this relates to what it means to me academically, this is simply an excuse to work on a project I find interesting, and to be forced to actually make headway on it due to the accountability provided in a teammate and class structure. In relation to the field of academia, this project should be fairly new, as nowhere on the open internet does anyone mention hooking up LLMs to stable diffusion for intuitive user experience. If this project were to succeed, (which I suspect will be very hard in the short time frame of the class), it would set the groundwork for LLM being able to display simple scenes and ideas visually.

In regards to the usefulness of classes within UC in respect to this project, the most useful course is Software Engineering (EECE 3093C). This course is extremely applicable, as it covered the nature of managing a long term software project, and showed me the pitfalls of poor

project management. Another class which is applicable is Database Design and Development. This class is an odd one in its applicability, but the general applicability comes from the mile high view of multiple different databases. At its core, the stable diffusion LLM interface problem is a storage problem, as being able to read from generated content and modify it. As a whole, however, most of the classes I have taken have been largely in applicable, as the AI classes are not covered until Senior Year, and useful LLMs (ChatGPT 3 and beyond) public accessibility is less than a year old.

When it comes to COOPs, the only useful one will be my current one at Cryptic Vector. While non-technical skill are always useful, especially in respect to human interaction, these I have largely picked up in highschool, not college. The reason I find Cryptic Vector so useful is my mentor there, Steve Kunath. Kunath has a PhD in linguistics and economics, and a Master's in computer science. The bulk of his work is in natural language processing, and the resources at Stanford and connections to topics such as Resource Description Frameworks and Web Ontology Languages have allowed myself to skip months worth of trial and error to simply know of a resource which already exists, and is heavily applicable to my area of study. Additionally, Cryptic Vector has a number of machine learning scientists, of whom I am in contact with and learning on a weekly basis. While I would like to have the accessibility to Dr. Minai that other students at UC do, I have that accessibility to AI researchers at Cryptic Vector on a daily basis, as I am currently on COOP there.

My motivation for this project is three fold. First, I want to work with a project on a topic I already know. I play Dungeons and Dragons on a weekly basis, and building and designing maps is an extremely time intensive process. Being able to have an application which does this for me would both make my life easier, and knowing where my pain points are in map design

and creation makes this project easier to jump into. Secondly, I want a project which is in a cutting edge field, and able to eventually be converted to a product. As I have mentioned previously, LLMs are relatively new to public consumption. This coupled with the popular appeal of DnD (over 50k monthly subscribers to DnD Beyond, a character manager for players) shows a fairly large market for the success of other projects. Finally, I want a project which is extensible into another project. If we can get this project working, this opens up the possibility of using LLMs to generate a fully unique, plot oriented world in video games. While character dialogues in some video games are already in development (Convai), the wholistic creation of both graphics to dialogue is unrealized at this current time.

My preliminary solution to creating sprites from LLM input is to simply have LLMs take human input, and convert it to the commands entered in a stable diffusion model. Eventually, we would like to include back propagation and some form of AI lens to be able to read the data back to the LLM. This is much harder, however there are currently in development extensions and models which accept image input to the AI. These, by themselves, are not sufficient, as evaluating the image without context provided by the original generation creates a disconnect between what the LLM evaluates and believes itself to generate. A solution to this is including the prompt with the image for back propagation, and this is where the Resource Description Framework (RDF) comes into play. The RDF allows us to carry both the image, text used to generate the image, and the important points (such as a well in a generation of a castle, or something similar). However, we (Prateek and myself) are starting smaller, and attempting to build human and LLM usable tools, so that even if we do not get a solution to our satisfaction, we can still have a human usable tool developed out of the experience.