What is the question you are hoping to answer with your research?

Machine Learning (ML) is one of the most talked about fields in Computer Science today, and for good reason. The idea of having a computer derive patterns in pictures, natural languages, media streaming choices, consumer purchases and more, and then be able to make a "logical" prediction is very exciting. Right now, ML techniques are used across a wide variety of applications, yet most ML algorithms are inherently flawed because of input bias. A computer has no sense of morality, so it naturally won't be able to make the most moral, ethical, or diverse choice if that data being inputted is inherently human, and therefore biased. In my freshman Monroe research project, I realized just how dangerous biases can be in ML algorithms; when teaching a computer to play a simple, pre-built text adventure game, the computer learned in a biased way based on how I had written the game (for example, when I put the dragon on the left side of the cave, the computer started to associate leftness with badness). Now, with a deeper understanding of these biases, I'd like to see if there are measures I can implement in a ML algorithm that writes a simple story to counter biases present in traditional storytelling. Basically, what can we do, with full knowledge of biases present in ML algorithms, to counter or reduce the negative effects of these biases?

Discuss your methodology in detail. Layout the plan of exactly what you are going to do.

This research project poses quite a few inter-disciplinary questions. First, I would need to do research on bias in Machine Learning, which would require reading published academic papers about the subject. Then, I would need to investigate Neural Networks and Long-Short-Term Memory Recurrent Networks, which are the best ML algorithms for processing Natural Language (how I would make the computer "write" a comprehensible story). Then, depending on the text that I trained the algorithm on, I would need to research the kinds of biases found in those texts (ex. Children's books from different eras, classic literature, etc.). So, all in all, I'm going to do classic research and reading on bias in different forms of literature, bias in ML algorithms, Neural Networks and LSTM RNNs, and then learn how to code a ML model for generating coherent stories using the Keras ML frontend in Python.

What resources are available to you for this project (lab space, access to collections/archives, additional funding, etc.)?

As a CS major, I have remote access to the CS Lab system with lots more processing power, allowing me to train a model on vast datasets. I'd also have access to a wealth of online tutorials for Keras (the frontend that I'd use to develop my model) and online ML classes/tutorials to learn from. I would use W&M's Library portal to access research papers for reading about bias in ML and literature.

What new knowledge or interpretation will emerge from this proposed project?

Although I technically did my freshman project on ML, I ended up learning mostly about Natural Language Processing (NLP) and developing my basic coding skills, since I was still a relative beginner. Now, with tons more coding experience under my belt and a general understanding of NLP and ML, I can dig deeper into the various forms of Neural Networks and learn about one of the most pertinent subsets of CS (Machine Learning) in a way that relates to my other passions, which are helping people (promoting diversity by countering biases) and telling a story. With this exciting foray into ML, my career possibilities in CS widen tremendously and I can use the skills I gain from this research to work on other ML applications.

What will the deliverable be from your research (article, play, website, story map, etc.)?

This project will result in a Neural Network model that can write simple stories that are (hopefully) countering biases in some way, such as by picking the most diverse person when presented with multiple character types. I will also be keeping an updated blog with my research findings from reading the various journals and ML lessons.