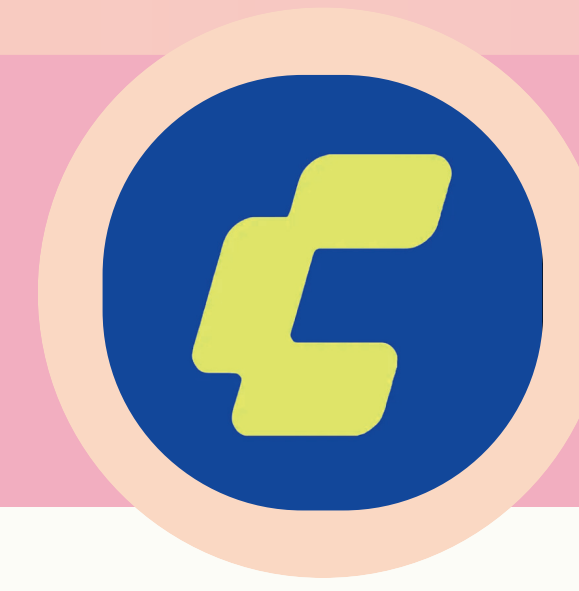


# DOODLE-TO-IMAGE

## Enhancing Text-Guided Image Generation with Sketch Prompts in ComfyUI

**Osen Mac-Iriase**

Department of Mathematical and Computational Sciences



**Advised By Dr. Daniel Palmer**

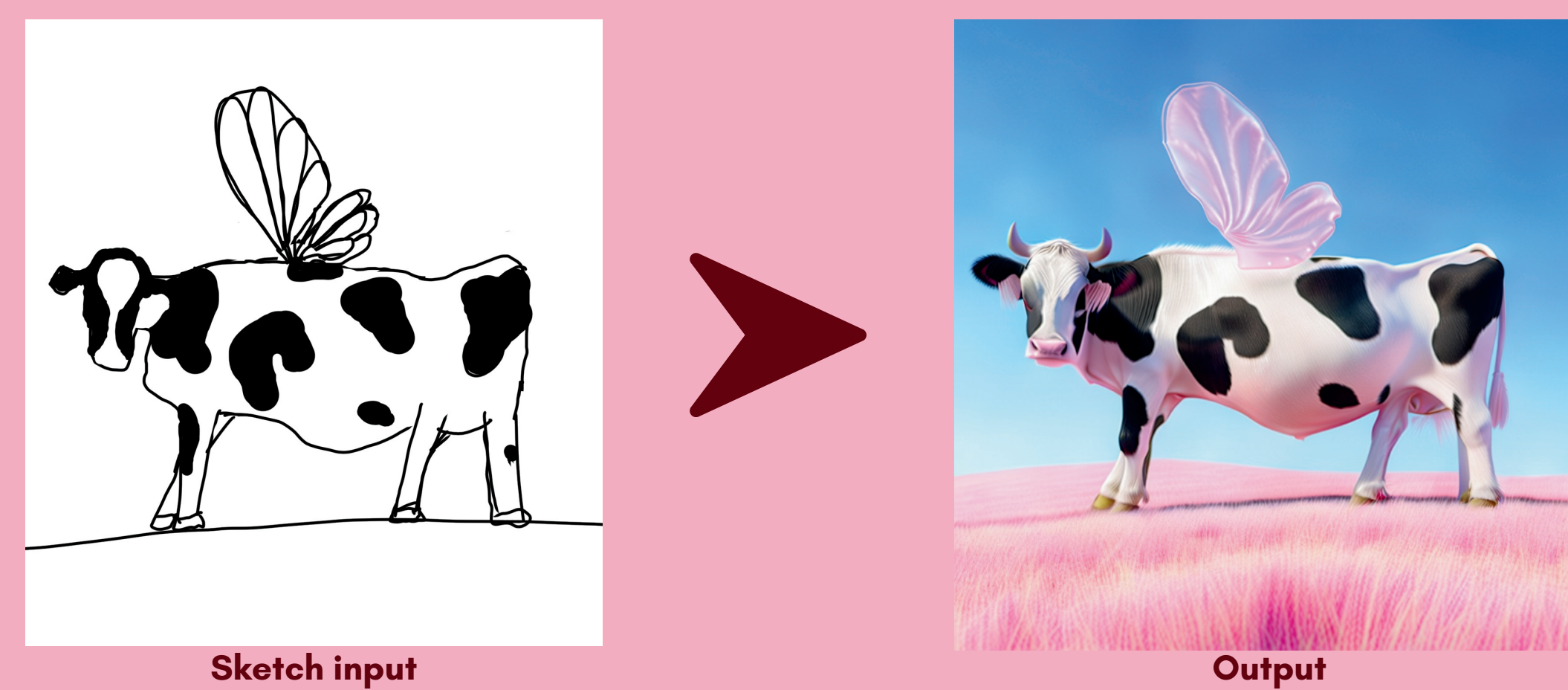
The College of Wooster

### Introduction and Background

**The Problem:** Text-only prompts often fail to dictate precise spatial layouts, leading to semantic drift where AI interpretations ignore user intent for object placement.

**The Solution:** This research utilises a Sketch-to-Image workflow in ComfyUI. By using Canny Edge Detection, hand-drawn sketches become spatial anchors for the Stable Cascade model.

**Research Objective:** To demonstrate that multimodal inputs (Text + Sketch) provide superior compositional stability and user control over standard text-only workflows.

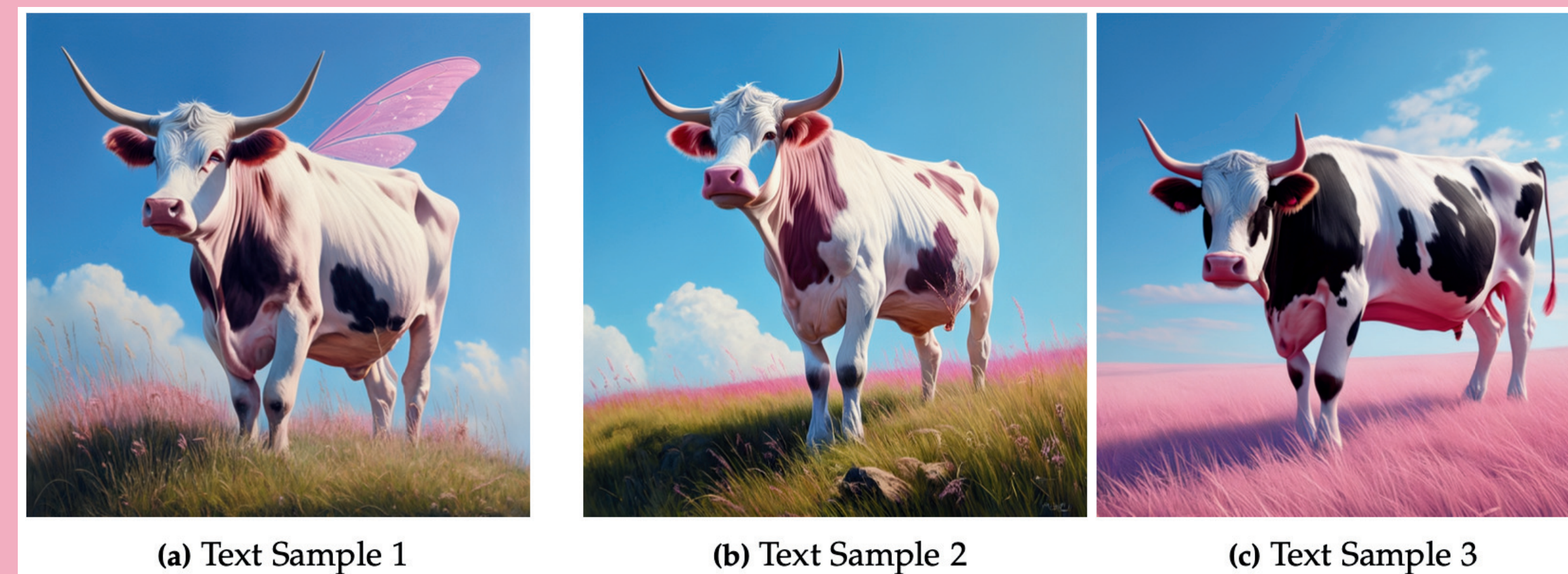


### Experimental Result & Analysis

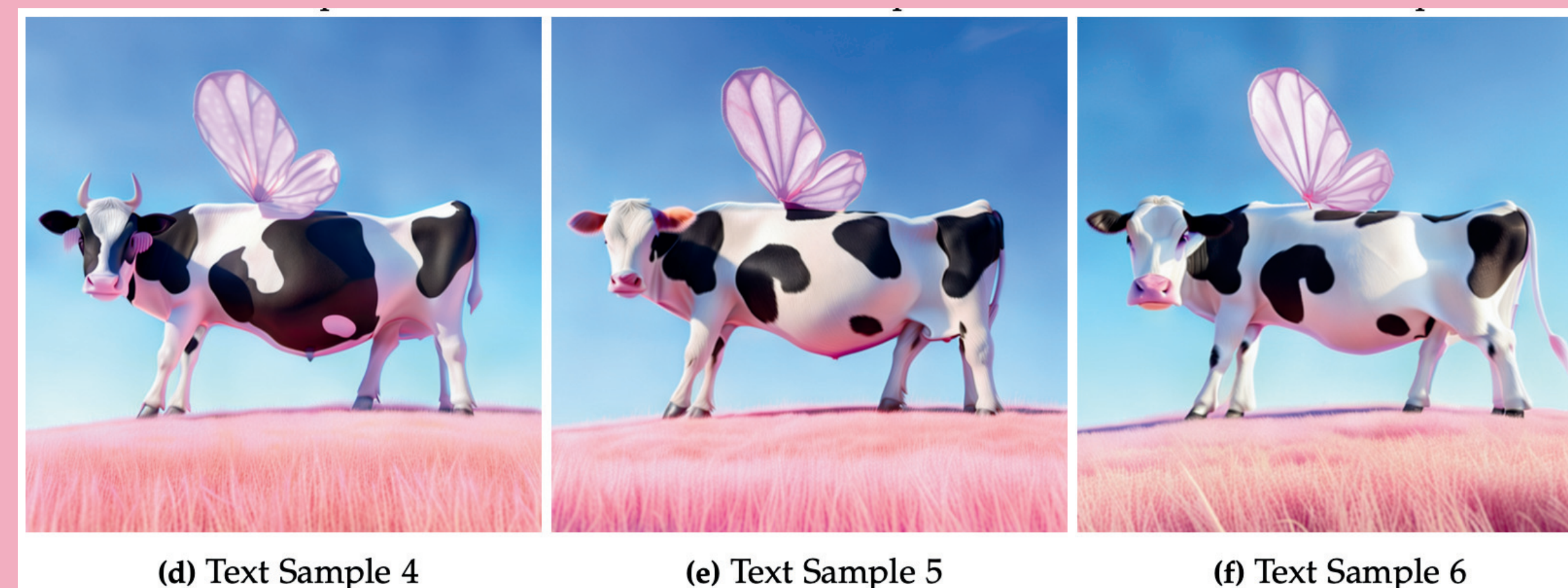


Target Image: Made in Canva

Text-Only Generation



Sketch+Text Generation



### Discussion and Impact

**Key Findings:** Better Control with Sketches. Adding sketches helps the model stay closer to what the user intends. In text-only generation, images can drift or place things incorrectly, especially at higher CFG values. Using Canny edges as a guide keeps the structure consistent while still allowing the model to add details and style.

**Impact:** This approach shows that tools like ComfyUI can give users more control over AI, instead of feeling like a "black box." It can be useful not just for creative work, but also in fields where accuracy matters, like data visualisation or scientific imaging.

### Conclusion

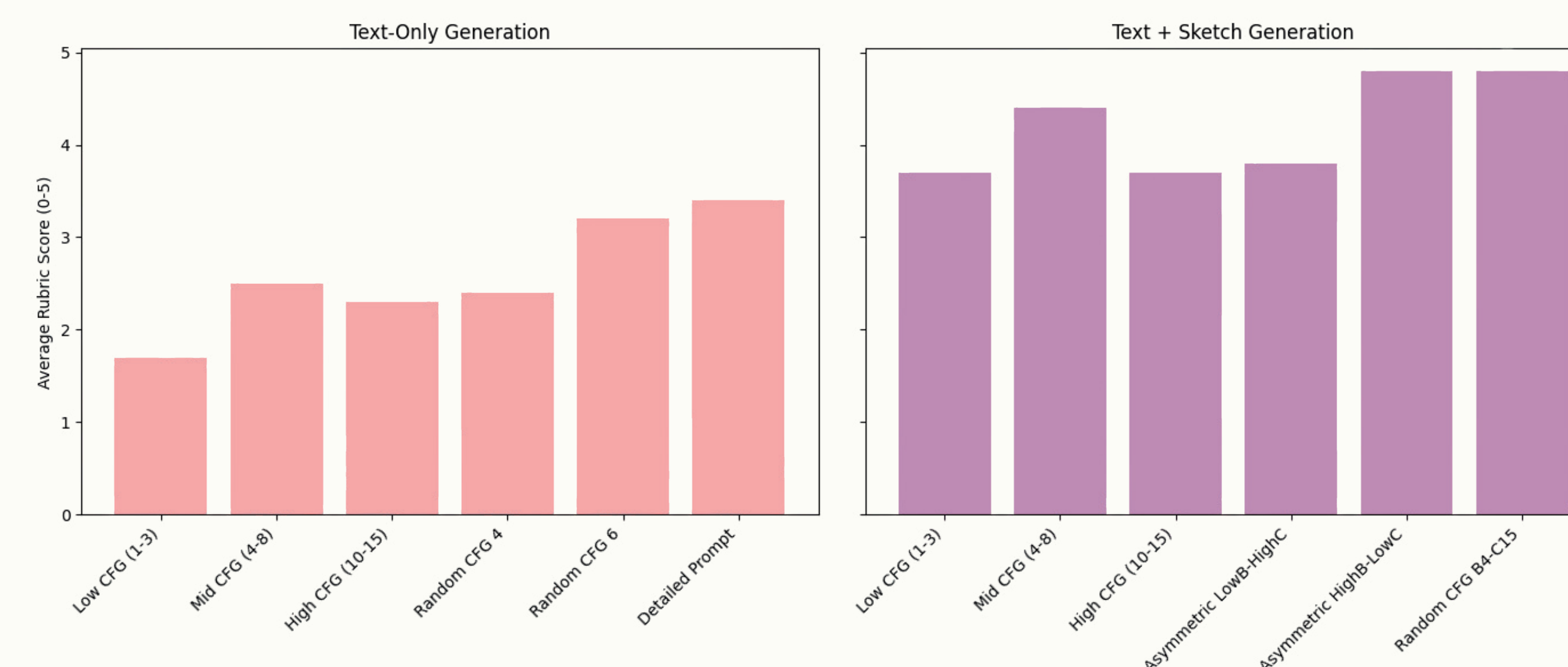
This project shows that combining sketches with text prompts improves control in image generation. By guiding the model with simple structural outlines, it becomes easier to produce images that match both the intended layout and meaning, reducing unpredictability in the results.

For future work, this approach can be strengthened by adding evaluation metrics to better measure accuracy and consistency. It could also be expanded to more advanced applications, such as 3D modeling or video generation, where maintaining structure over time is even more important.

### Methodology: Structural Conditioning

This project combines simple sketches with text prompts to guide image generation.

The sketch is first turned into a clean outline using edge detection, which acts as a guide for the image's structure. This outline and the text prompt are then fed into a diffusion model using ComfyUI. The model first creates a rough version of the image and then refines it into a detailed final result while following the sketch. Different CFG values were tested to see how much the model follows the sketch versus adding its own creative details.



### Acknowledgements

I would like to thank my advisor, Daniel Palmer, for his guidance and support throughout this project. I am also grateful to Jérémie Wenger for introducing the core idea and providing valuable direction. I acknowledge the support of The College of Wooster and the Copeland funding program for making this research possible.

