**LLM EXPERIMENT 5**

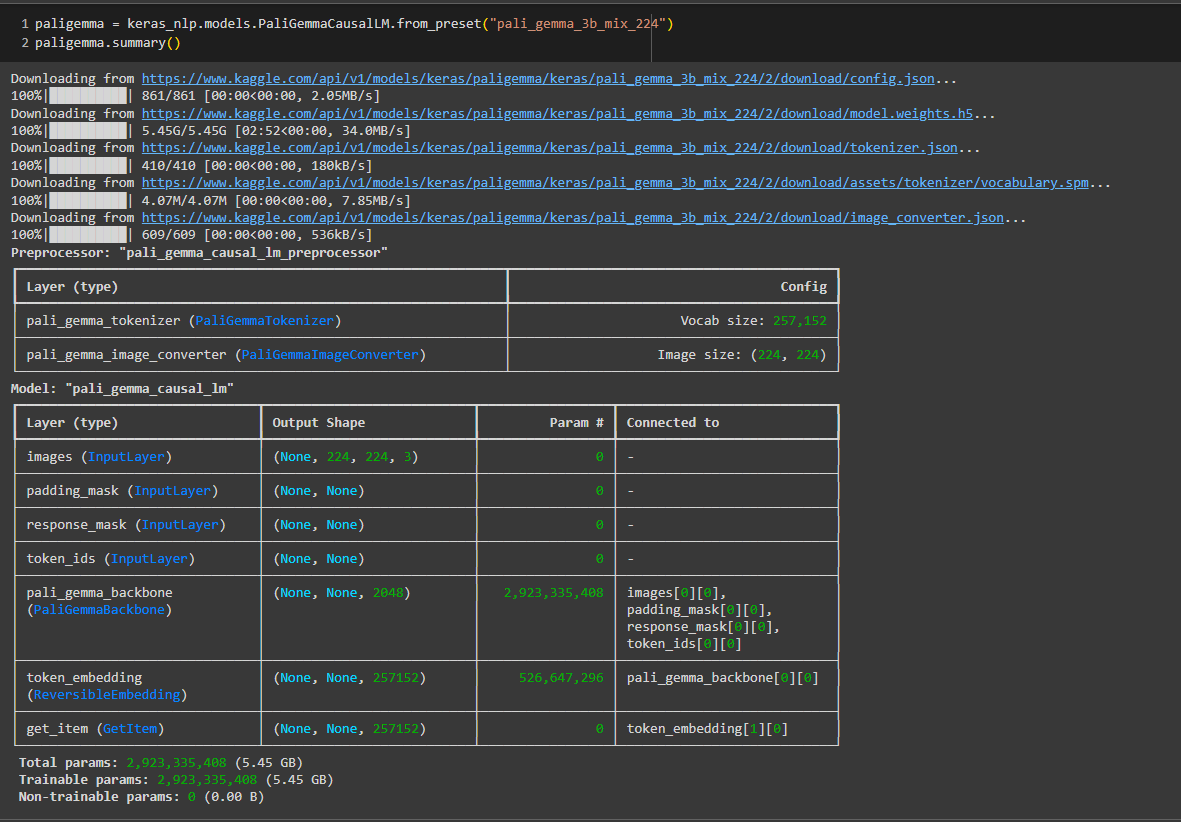
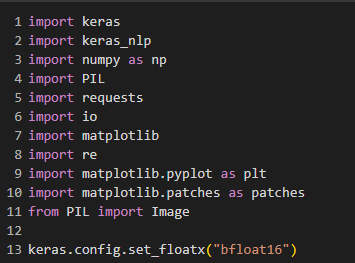
Vedant Kambli

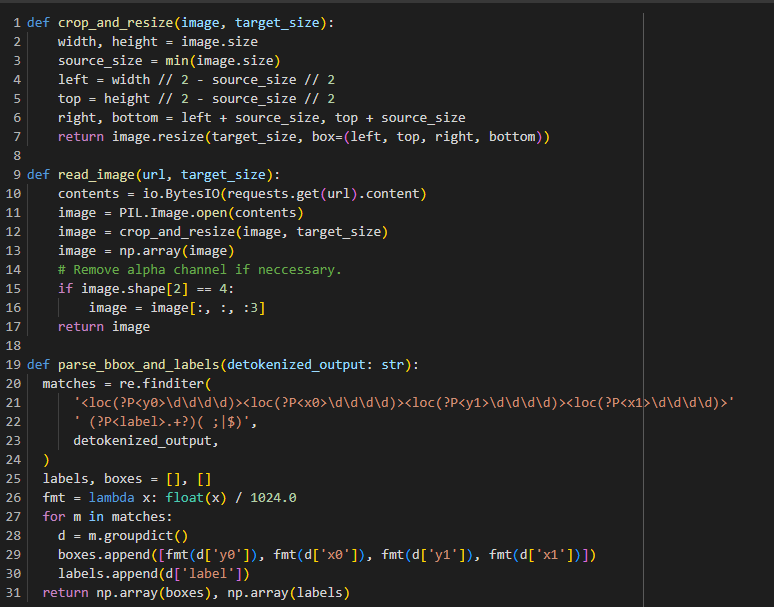
60017220096

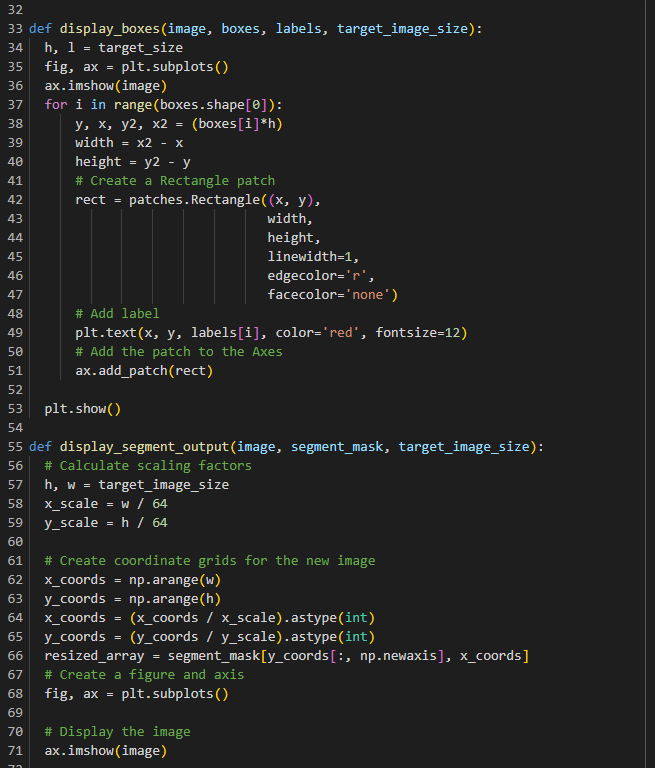
**Aim**: Build a simple multimodal generative model that combines text and image inputs to generate captions.

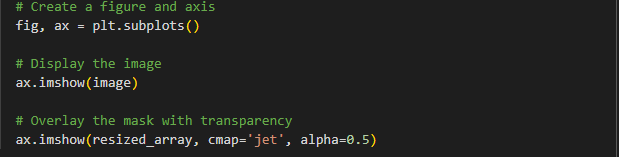
**Theory:**

**Code & Screenshot:**

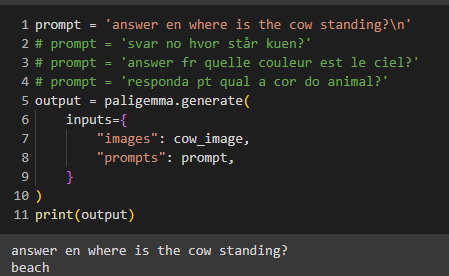


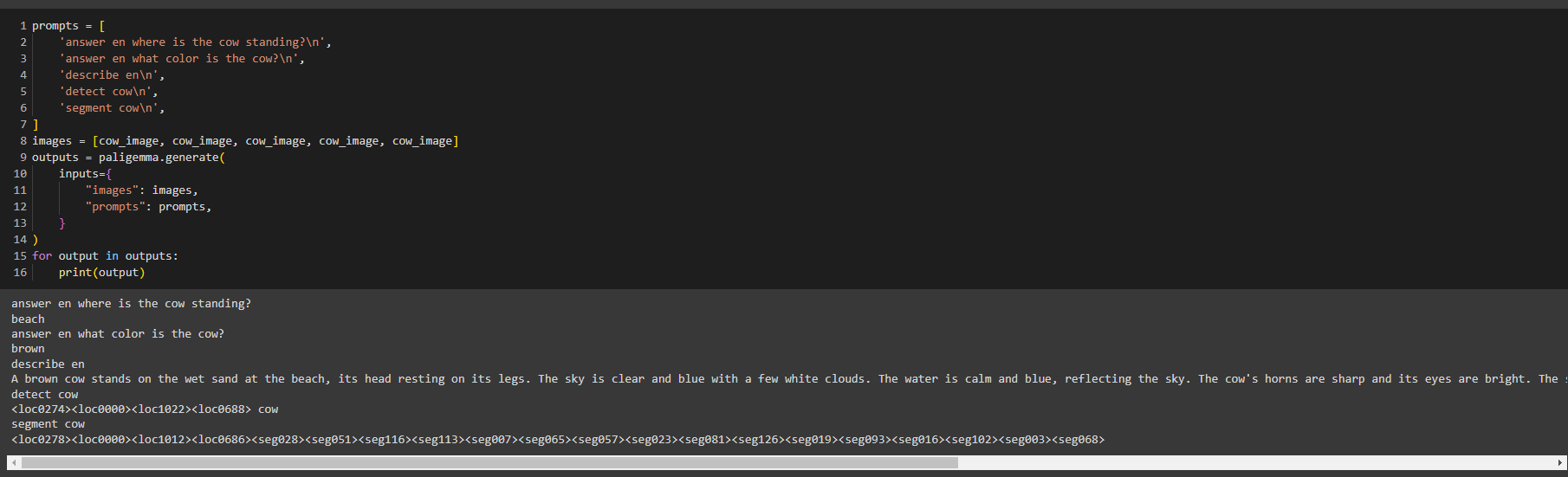


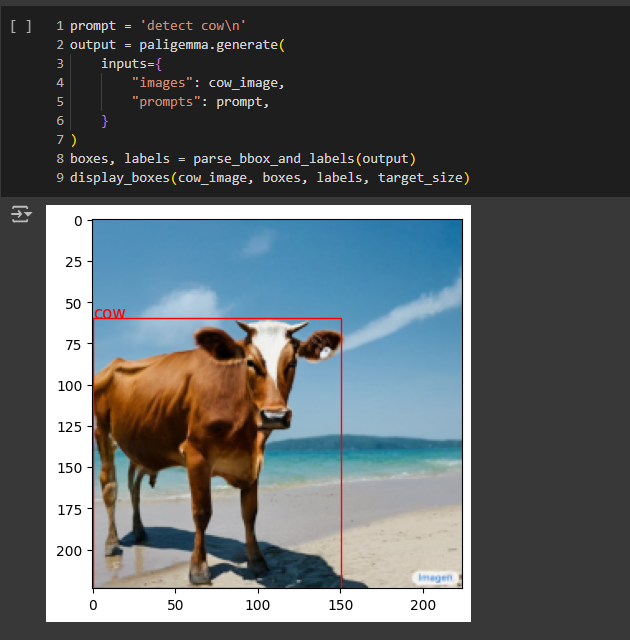


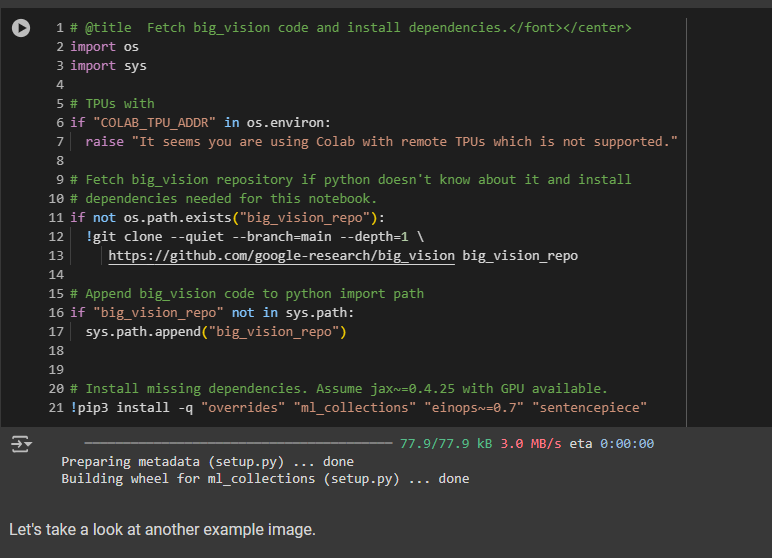




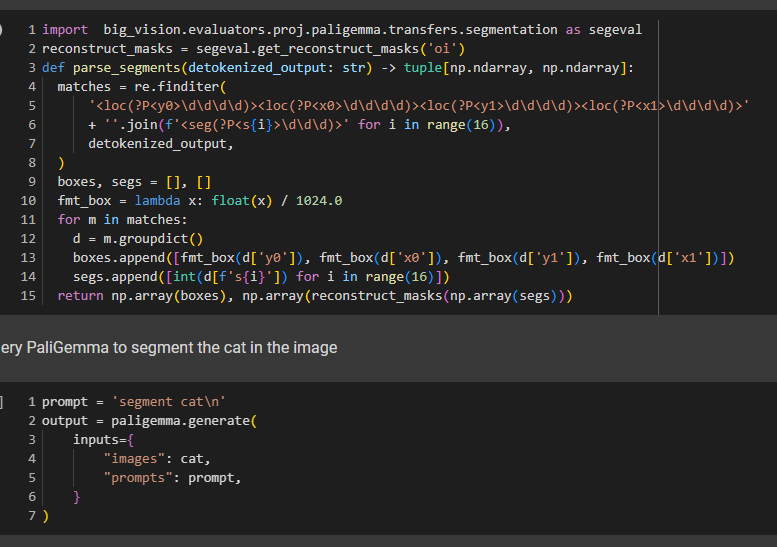


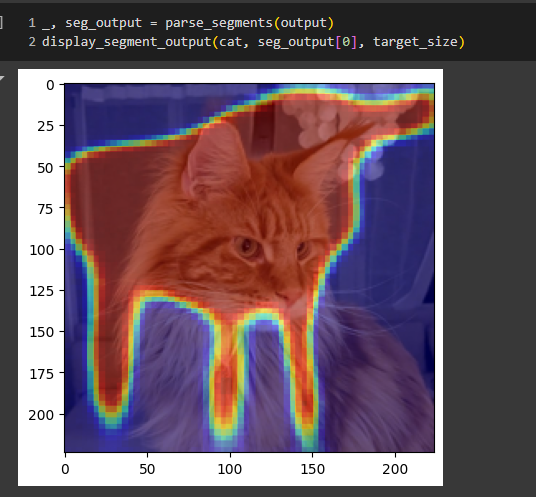












**Conclusion:** Therefore, we built a simple multimodal generative model that combines text and image inputs to generate captions.