**Weekly Report**

**Week 1 — Planning (Days 1–6)**

**Day 1**

* Defined project scope: create a fine-tuned code model for a CrewAI multi-agent coding system.
* Selected Hugging Face + LoRA fine-tuning approach for Colab compatibility.
* Decided to use a multi-agent setup: Creator, Debugger, and Reviewer.

**Day 2**

* Evaluated pretrained coding models for Colab: CodeGen-350M, Starcoder, CodeLlama.
* Chose **Salesforce/codegen-350M-multi** for manageable VRAM footprint.
* Drafted high-level plan with planning and execution phases.

**Day 3**

* Researched datasets for supervised fine-tuning.
* Initially picked OmkarPathak/python\_code\_instructions\_18k (later found unavailable).
* Outlined training parameters and LoRA config.

**Day 4**

* Identified ipywidgets + IPython display for interactive multi-agent UI.
* Planned integration of fine-tuned model into CrewAI tool.
* Designed sequential task execution for agents.

**Day 5**

* Drafted dataset pre-processing pipeline with prompt/completion formatting.
* Created plan to test model output after training using Hugging Face pipeline.
* Noted importance of matching prompt style during training & inference.

**Day 6**

* Reviewed potential risks: dataset availability, GPU limits, output quality.
* Finalized decision to use **CodeAlpaca** dataset as backup due to stability.
* Scheduled error-tracking and debugging during execution phase.

**Day 7 — Rest Day** 💤

**Week 2 — Execution (Days 8–13)**

**Day 8**

* Began dataset loading: encountered DatasetNotFoundError for original dataset.
* Switched to codeparrot/self-instruct-code, then finally to **CodeAlpaca**.
* Adjusted pre-processing to match CodeAlpaca’s "instruction"/"output" fields.

**Day 9**

* Implemented LoRA fine-tuning with 8-bit quantization in Colab.
* Hit ImportError for outdated bitsandbytes → fixed by upgrading package.
* Verified model loading with tokenizer padding fix (tokenizer.pad\_token = tokenizer.eos\_token).

**Day 10**

* Ran short fine-tuning (~200 steps on 1% dataset) → output was [No new tokens generated].
* Diagnosed issue as undertraining + prompt mismatch.
* Increased dataset slice and max\_steps for more meaningful updates.

**Day 11**

* Retested base model to confirm it generates code as expected.
* Adjusted inference prompts to include explicit "Code:" cue.
* Successfully got valid code output from fine-tuned model.

**Day 12**

* Integrated fine-tuned model into CrewAI multi-agent setup.
* Configured Creator → Debugger → Reviewer sequence with hf\_tool pipeline.
* Added ipywidgets UI to show agent discussion and final code in Colab.

**Day 13**

* Validated full pipeline end-to-end: fine-tuning → model testing → CrewAI multi-agent execution.
* Captured all warnings (unused weights, padding id) for documentation.
* Confirmed system produces reviewed, functional Python code