1. Group Members
    a. Luke Hartzell
        i. Skills: Python, Java, C, Google DDss\Vertex AI
        ii. 1 credit hour
        iii. Task: Design and Implementation of line by line coding and insights
    b. Dongzhao Song
        i. Skills: Python, java, C, C++, Pytorch
        ii. 2 credit hour
        iii. Task: Helping doing Research on the idea of Llama
            1. Research and run the Llama model locally. Try to understand how to use it as code helper
    c. Mario Goranov
        i. Skills: Python, Java, Pytorch, Node.js, experience with other LLMs such as finetuning GPT-2 on Google Colab
        ii. 1 credit hour
        iii. Task: Backend, Maintenance/bug fixing, Marketing
        iv. I basically don't want this to be hidden away at the end, I want people to actually use it
    d. Jean-Jacques Muteteke(JJ)
        i. Skills: java, javascript, node, azure cloud development, pace cluster and cuda development and python
        ii. 2 credit hours
        iii. Task: Researching VS code extensions and their development
            1. Also Research GA Tech's current Hardware capabilities to see if it will suit our needs.

2. Project Goals

The goals of the project are to create a code helper for students or people that want to learn how to code as well as to explore the Llama model. We hope to explore the capabilities and implementation of the Code Llama model and potentially test how it compares to ChatGPT (with benchmark metrics – time permitting). The goal of the code helper is to provide line by line coding help – explaining code, providing tips, etc – to help coders learn how to code better functionally and more efficiently. We want to see and attempt to place our solution into a VS Code extension, for others to download and use.

We will measure progress based on phases of development. Each member is assigned a certain feature they will implement. Luke will do line by line comments and insights. Mario will create the entire release plan as well as work to squash all bugs. JJ will research the creation of VS Code extensions and the process that goes into that. JJ will also research current GA Tech hardware capacities and see if it will suffice in our goal to host the model ourselves. Finally Dongzhao will start out by researching the LLama models and its variants. He will test them out and see which one will best suit our needs. The first stage is initial research, all information and

designing of the respective feature (1 week). The second stage is implementation specifics research (1-2) weeks. And the final stage is actual implementation, which will be measured by completion and functionality (3-4) weeks. Time permitting, the final stage will be benchmark comparisons with ChatGPT.

-

Milestone	Date
Team Accustomization	(02/02)
Individual piece research	(02/09)
Implementation of research	(02/16)
Implementation of research	(02/23)
First Line of Code	(02/23)
prototype	(03/02)
prototype	(03/09)
prototype/potential ChatGPT Comparison	(03/16)- (03/30)
Refactoring/integration/bug fixes	(04/06) - (04/20)
Presentation and main product	~(04/27)

-

3. problem to be solved, interesting ideas that you would like to study
    a. Most Points have been answered elsewhere in the Proposal.
4. proposed solution / approaches / foreseeable problems and potential pitfalls
    a. Potential Pitfalls
        i. The foreseeable problems and potential pitfalls that we see are the use of lots of new technologies
        ii. We foresee potential problems in the training/finetuning of these open source models, as obviously they aren't as robust and documented with videos as ChatGPT. Since the Open-source models tend to be smaller in parameters there is a potential pitfall in simply the lack of embedded knowledge in these models.
        iii. We might have deal with unforeseen issues related to VSCode

iv. We want to attempt to Host the model on Georgia Tech servers, This might cause issues as we aren't fully aware of the new things Georgia Tech is doing on their Pace Clusters.
v. Overall We have limited knowledge of IDE extensions and Llama, so this project will be very exploratory for us (causing time delays, new document exploration, etc.)

5. implementation, tools needed, resources, file management
    a. **Implementations**
        i. Tools
            1. FIle Management
                a. Github
            2. Programming Language
                a. Python and javascript/typescript
            3. Cloud servers
                a. Pace clusters
    b. **Resources**
        i. Weekly meetings on discord
            1. We already have audio chat working.
        ii. Google Drive
        iii. Teams and Slack for any extra communications

6. Project Description

The problem to be solved is overcoming the barrier to entry for new coders. Often, navigating documentation and understanding fundamental coding concepts is difficult for new coders. This will streamline and personalize the process, cutting out the middleman by providing direct feedback and explanation of code lines in the IDE. Some interesting ideas we would like to study are IDE extensions, Llama performance and implementation, and retrieval augmentation generation. We will implement using Llama for the AI, VS code for the text editor, python for the test language, and GitHub for file management.