

# CASEFlow - AI-Powered Email Response System Using Fine-Tuned LLMs for Customer Service

**Team members:** Cameron Clyde, Angela Harris, Sohil Marreddi, Emma Smith | **Faculty Advisor:** Preetam Ghosh, Ph.D. | **Sponsor:** CoStar Group | **Mentor:** Keroles Hakem

## Background

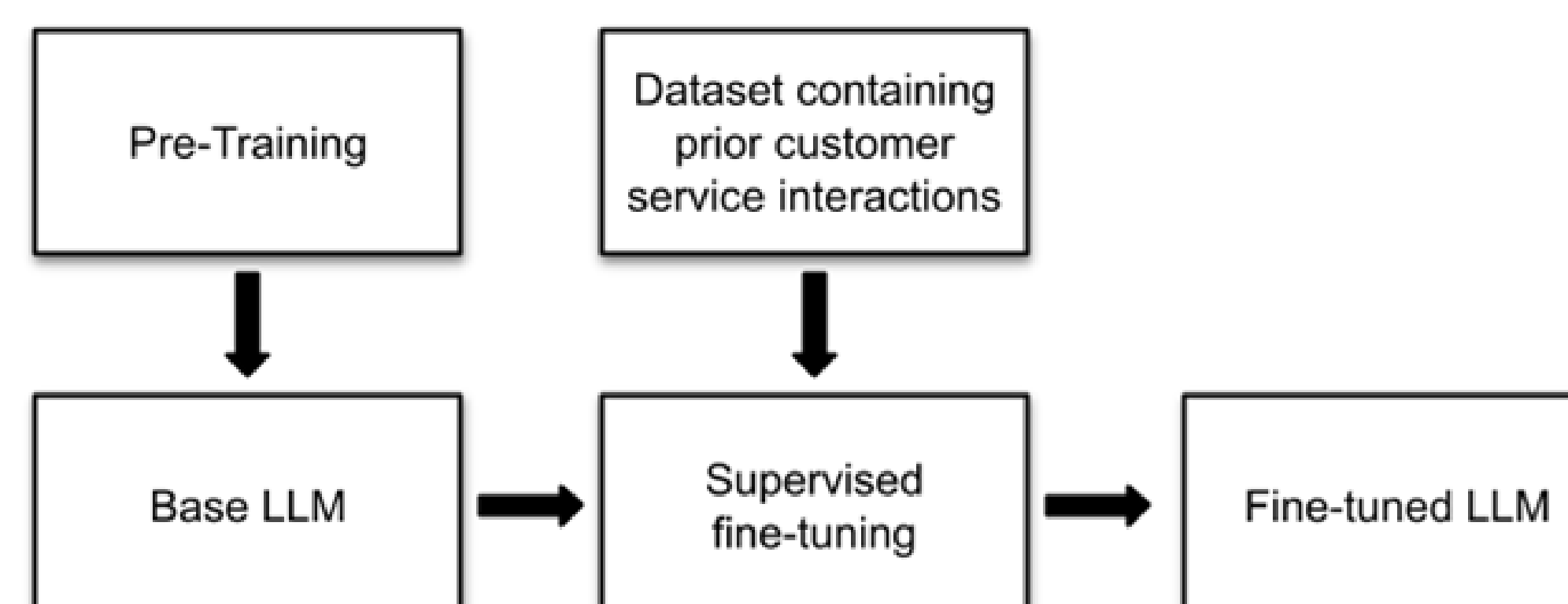
- **Current Method:** In CoStar's current customer service process, representatives respond to issues by selecting a prefabricated response template, then customize it with personalized details the customer's name and issue-specific information.
- **Increasing Workload:** As CoStar Group continues to grow, the demands on its customer service teams increase significantly.
- **Balancing Tasks:** Customer service employees are responsible for handling complex cases and responding to customer emails, which is challenging to balance.
- **Need for Efficiency:** These challenges highlight the need for a solution that streamlines the email process, allowing employees to respond quickly and efficiently without compromising the quality of service.

## Objectives

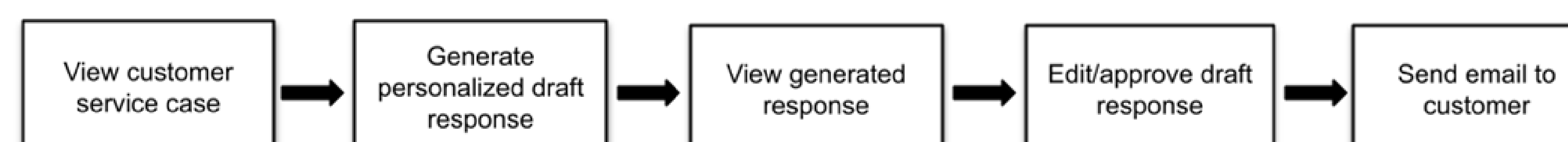
- **Improved Email Response Times:** AI generated draft replies aimed to reduce employee response time
- **Enhanced LLM Performance:** Fine-tuning of a LLM using customer service data, with the goal of improving speed and accuracy
- **Drive Continuous Improvement:** Use employee feedback to refine the AI model, enhancing speed, accuracy, and response relevance.

## Approach

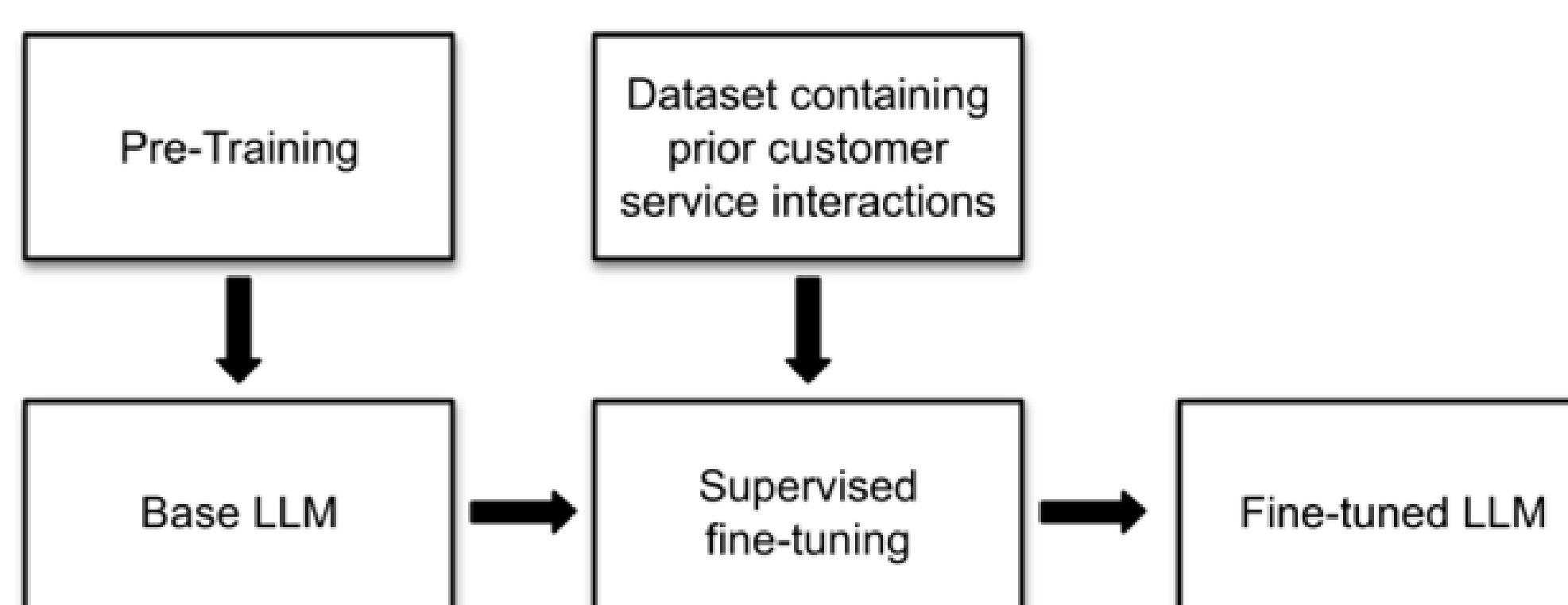
- **Fine-tune LLM:** Fine-tune LLaMa 3.1 using Amazon SageMaker with real customer service message dataset.



- **Web Application:** Create a web application using React and FastAPI to allow representatives to submit/view customer service cases. Representatives will be able to use the AI response tool through this web application.



- **Continuous Improvement:** Implement a feedback loop, allowing customer service representatives to give ratings to the generated responses. This feedback will further fine-tune the model and improve the performance of the generated responses.



## Features

- **Web App with Authentication:** Develop a secure web application with login using JWT tokens.
- **UI for Review & Edits of Generated Responses:** UI provided to allow employees to review and edit responses generated by our fine-tuned LLM, with success measured by usage and edit rates.
- **Feedback Loop for Model Improvement:** Employees can rate model-generated email responses, and improvement will be measured by the model's ability to learn from feedback
- **Rich Text Editor:** We are implementing Lexical rich text editor to provide options to format text in case descriptions.

## Tech Stack

