Project Status Report- Smoke Outlook Translation Model

Project Team Name: Team Fire Bot (Smoke Outlook Translation Model) [Aparajita Pathak, Samuel D. Ryan, Agampreet Singh, Frankie Roberts, Yiyu Chen]
Sponsor Name: Jonathan Callahan
Date: Feb 4, 2025

Project Purpose - The USFS AirFire group maintains systems that forecasters use to create Smoke Outlooks, i.e. wildfire smoke forecasts. These are initially written in English. Then they are auto-translated into Spanish with ChatGPT. Finally, the translations are manually edited by a native speaker. The goal of this project is to use our corpus of existing outlooks and translations to train a ChatGPT model for improved translations. A simple user interface will also be needed to allow professional translators to compare translations using the trained and untrained models.

Add value proposition

I. Summary of Progress/Milestones Reached (since last reporting period) (move deliverables up once completed. max 6. remove lower priority items)
We have successfully completed the following milestones:
- Weekly team and sponsor meetings
- Team members have added branches of their own information to the GitHub repository i.e notes, practice code
- Coding environment setup i.e Git, ExpressJS, Visual Studio Code, Node.JS
- Created initial test model on ChatGPT with rough translations
- Conducted initial testing of the development environment on OpenAI with API Keys from the sponsor
- Reviewed solution requirements with sponsor
- Established a workflow for team updates and task allocation to maintain consistent progress.
- Present to sponsor two options of AI models to create the new smoke outlook translations with (GPT-4 vs. GPT3.5)
- Decided to use both fine tuning and prompt engineering

II. Deliverable/Milestone Status (since last reporting period)
- Green – All on time
- Yellow – Some delayed
- Red – Ability to complete key deliverables/reach key milestones significantly at risk

Status color and explanation of above delays and/or significant deviation from project schedule and plan to get back on schedule.

Deliverable	Status	Explanation
Create initial document notes/step-by-step in GitHub	Completed ETA: Before Next Sponsor Meeting 2/7 Progress: 100%	Sponsor asked us to create a notes document describing all the steps we are taking while completing this project.
Acquiring new translation	Completed	Sponsor informed us that we

data	Completed ETA: Before Next Sponsor Meeting 2/7 Progress: 100%	would be receiving more translation data from past smoke outlooks in order to improve translations on the new model
Acquiring more API tokens from USAirFire	Completed ETA: Before Next Sponsor Meeting 2/7 Progress:100%	In order to run more test smoke outlooks we need more API keys and hence we need some more budget approval again (An extra $100).
Choosing which AI model to select	On Track ETA: Discuss with sponsor during meeting 2/7 Progress: 50%	Choosing between GPT-4, GPT-3.5, and now the sponsor has presented two more models in OpenAI with API access, o1 and o3 mini.

III. Resource Status (since last reporting period)

task owner
Status color and explanation of above resource issues and steps planned to remediate.

Resource	Status	Explanation
OpenAI API	Acquired	OpenAI credentials are provided by AirFire, and the API is readily accessible for translation tasks. More keys/tokens to be acquired from sponsor in future
JavaScript/Node.js	Implementing	Development environment is set up locally, and team members are familiar with the required programming language or acquiring skills through material.

Project Status Report- Smoke Outlook Translation Model

ExpressJS	Implementing	Framework is available, and team members have experience or acquiring skills through working with it.
HTML/CSS/JavaScript	Implementing	Tools and skills needed for UI design are accessible and aligned with the project timeline.
Corpus of translations	Acquired more smoke outlook data	Existing initial data from Smoke Outlooks and translations are readily available for preprocessing and training. Much more data has been acquired
Visual Studio Code	Ongoing coding	Recommended IDE is installed and functioning, enabling efficient development and debugging.
GitHub repository	Utilizing	Repository has been set up and is ready for version control and collaboration.
Postman (or equivalent)	Utilizing	Tools for API testing are available and can be integrated into the development process.
OpenAI training guide	Referencing	Documentation is available online to guide model training and prompt engineering.
Budget for API usage	Acquired	Sponsor has approval of adding budget for API Keys and Tokens
User feedback for testing	Might face delays	Dependent on translator availability and feedback,

		which may delay iterative improvements to the system.
Development milestones	On Track	Initial phases (setup and basic API integration) are on track with no delays expected at this stage.
Documentation (HOWTO, etc.)	On Going	Can be prepared as part of ongoing development, with no significant barriers anticipated.

IV. Sponsor Status (since last reporting period)

Sponsor Status: Green

Our Sponsor, Jonathan has remained actively engaged, attending all scheduled meetings and responding promptly to updates. His support has been valuable in keeping the project on track.

Currently, we are working together to evaluate AI model options, which requires additional API tokens for further testing. This ongoing process involves continued collaboration to ensure we select the most suitable model.

No major issues have been identified at this stage. To maintain this status, we will continue providing weekly updates and proactively addressing any potential challenges to keep progress aligned with project goals.

V. Other Status Points

- Any other issues encountered that your TA and/or the Capstone Instructor needs to be aware of?
- Anything that your TA and/or the Capstone Instructor can provide to help ensure progress and project success?

There haven't been any major issues that need attention at this time. The project is progressing as planned, and no significant obstacles have been encountered.One area where additional input would be valuable is in optimizing the GPT model, particularly given the budget constraints. Any advice on how to fine-tune the model effectively without exceeding the budget would be helpful.

We're also open to any feedback or suggestions that could help improve our approach as we continue with the project.

VI. Planned deliverables for next reporting period.

Include a brief summary of your planned deliverables/milestones for the next reporting period.

Project Status Report- Smoke Outlook Translation Model

For our next major milestone, we aim to make significant progress on key technical development aspects of the project. The focus will be on further developing the ChatGPT model or the OpenAI model, refining the user interface, and ensuring the backend is ready for integration.

For the next reporting period, we plan to:
- Begin training the ChatGPT model using the prepared dataset.
- Finalize the tailored prompts to enhance translation quality.
- Set up the backend structure for the ExpressJS web service.
- Start building the user interface for comparing translations from both models.

We will also continue testing the system and gather feedback from translators to refine the model and interface.