

## Coding Report for AD-AutoGPT Recreation

Author: Adit Potta

The goal of the project is to replicate the function of AD-AutoGPT through a small-scale language model that generates spatiotemporal data as found in online articles. This experiment takes a step beyond most summaries of articles into visualizing the locations and times when the events occurred. The experiment would also explore the capabilities of a simpler model-such as FalconsAI-in performing tasks for which larger LLMs would typically be deployed.

Given that more and more summaries of articles are available, this method brings an added value of an added dimension of identifying geographical and temporal trends. The visual presentation of information in this fashion can serve as a useful tool, helping scholars and researchers in their work.

### Implementation Details

#### 1. Data Gathering through APIs

- This notebook extracts data from web links with the use of various APIs.
- For collecting relevant articles manually, `BeautifulSoup` along with a `requests` library is used to maintain far better control over what data gets fed to the model for processing and training.

#### 2. NLP Pipeline and Model Choice

- The LLM has chosen FalconsAI for text summarization. This model has been tested to see if a lightweight LLM could perform good enough on tasks that a larger model could handle.
- The construction is initiated by drawing context from Adit's inspirations from the GeoText GitHub repository as well as LangChain documentation for implementation on leveraging NLP and geographic frameworks.

#### 3. Performance Evaluation of FalconsAI

- The experiment could report that notwithstanding the fact that FalconsAI is easier than AD-AutoGPT, it still does a fair job generating fairly specific summaries. Future work could involve a run on medical LLMs, which might yield a better output since they are tuned for specific tasks. The outcome could offer some insights into how specialized training could impact the outputs. Thus:

### Spatiotemporal Data Visualization and Results

#### -LDA and Geographic Mapping:

- The processed data has been visualized using Latent Dirichlet Allocation (LDA) and geographic maps built using GeoPandas and Matplotlib. The base map for showing locations mentioned in the scraped articles is derived from the Natural Earth dataset.

