

Project Write-up: News Chaos - Drawing with a Spreadsheet

Introduction

In this project, "News Chaos," I set out to address the prompt to "draw with a spreadsheet" and "turn noise into narrative" by transforming raw news data into an evolving, organic visual representation. I leveraged Generative AI to interpret complex information and translated it into an artistic display of sentiment and categorization. My core outline for this process was clear:

- I would not require manual training of datasets.
- News sentiment analysis would rely entirely on pre-trained models.

Process Details

1. How did I start? Where did the data come from? I began by identifying a suitable complex dataset: a comprehensive **news dataset about Indian media from 2001 to 2023, sourced from Kaggle**. This dataset contained key columns like `publish_date`, `headline_category`, and `headline_text`.

[Link for the dataset](#)

2. What was the involvement of Generative AI in data selection? Generative AI was **not involved in my initial data selection** (choosing the dataset from Kaggle). That was a manual decision on my part. GenAI's primary role was in the **interpretation and transformation of the raw textual content** within the selected data.

3. What were the challenges I faced?

- **Dataset Size:** The original dataset had over 1 million rows, posing significant memory and processing time challenges for my local execution.
- **GenAI Processing Time:** Running sophisticated AI models (like those from Hugging Face Transformers) on a large number of text entries proved computationally intensive on my CPU, potentially taking hours.
- **Iterative Visualization Design:** Translating abstract data concepts (sentiment, categories, temporal flow) into a precise and aesthetically pleasing generative art form required extensive iteration and refinement based on visual feedback I observed.
- **P5.js Drawing Performance:** Rendering thousands of complex, animated elements simultaneously in a browser environment demanded careful optimization to maintain fluidity.

4. How was the data categorized? Categorization was done entirely by **Generative AI, without any manual labeling or training of a custom dataset from my side.**

- I used the transformers library from Hugging Face.
- Specifically, I employed a pre-trained **Zero-Shot Text Classification pipeline** (facebook/bart-large-mnli model).
- This AI model read the headline_text of each news item and classified it into one of eight predefined categories: Entertainment, Education, Politics, Technology, Socio-Cultural, Economy, Sports, Crime. It achieved this by understanding the semantic content of the text relative to the provided labels, without needing any examples of labeled news data from me.

5. What are the decisions I took in the process?

- **Data Subsetting for Prototyping:** Due to the large raw dataset, I used a smaller, manageable sample (e.g., 10,000 rows, and later filtered to a single month for initial visualization steps) for development.
- **Python for Data Pipeline:** I chose Python (pandas, transformers, json) for its robust data manipulation capabilities and seamless integration with state-of-the-art Generative AI models.
- **Processing Environment:** All Python data processing was executed within a **local virtual environment (venv) on a virtual machine** (or my local machine configured as a VM for development purposes). This ensured an isolated and controlled environment for my library dependencies.
- **Processing Checkpointing:** I introduced an intermediate CSV file (intermediate_processed_news.csv). After the time-consuming GenAI sentiment and categorization steps were completed, the processed DataFrame was saved. On subsequent runs, my script checks for this file and loads it directly, skipping redundant GenAI inference.
- **P5.js for Visualization:** I selected P5.js as the visualization platform for its strength in creative coding, browser-native rendering, and ability to handle dynamic, animated graphics.
- **Visualization Style Evolution:** The design evolved through iterative feedback I observed:
 - Initial concepts of distinct "blobs" or "splines" per month were refined.
 - The core aesthetic shifted to a **Perlin noise-driven fluid flow field** for the entire canvas.
 - The "monthly" representation became **thicker, diffused, overlapping areas of color** generated by noise, rather than discrete shapes.
 - Individual headlines transformed into **finer, animating vector lines** ("sub-branches") emanating from within these monthly regions.
- **Sentiment-to-Visual Mapping:** I designed a precise mapping for sentiment:
 - **Neutral Sentiment (near 0.0):** I translated these to curves with **extremely low opacity** (very faded) and **very thin stroke weights** (0.1-0.5).

- **Negative Sentiment:** I translated these to curves with **darker opacity** (80-200 alpha) but **lesser stroke weights** (0.8-1.8).
- **Positive Sentiment:** I translated these to curves with the **darkest opacity** (150-255 alpha) and **higher stroke weights** (1.5-3.0).
- **Drawing Strategy:**
 - All curves are drawn **simultaneously** each frame.
 - The canvas background() is redrawn with a **low alpha value** (e.g., background(0, 10)) to create a subtle "fading trail" effect, allowing previous curve positions to slowly persist and blend into the overall dynamic image. This creates the ethereal, layered look.
 - Individual curves draw their full numSteps segments in every frame, ensuring complete lines.

6. How did I derive the visualization? My vision for this visualization was deeply inspired by nature; I initially wanted to draw **fractals**, as much of my artistic work comes from observing natural patterns and organic growth. I even explored giving it an evolving plant-like form. However, as I delved into the capabilities, I ended up working with **Perlin noise** because I truly liked the inherent fluidity and organic movement it creates. My goal shifted from focusing on individual headlines as discrete points to asking: "Can I get a glimpse of what the news is telling me? Can I visualize the overarching trends derived by AI to see what India was up to in 2001?" The entire visualization was designed with this intent in mind – to provide an intuitive, artistic overview of the data's flow and underlying sentiment, rather than just a granular display. The fluid animation and the layered curves are meant to convey a sense of continuous motion and changing moods in the news landscape.