Individual Project Documentation

Start date: September 1st

Datasets:

**AI Global Index**

<https://www.kaggle.com/datasets/katerynameleshenko/ai-index>

CSV file contains 62 rows and 13 columns. Five columns are categorical, and the others are numerical in float format.

**SOURCES**

The data was collected in open sources: <https://www.tortoisemedia.com/intelligence/global-ai/>, <https://www.worldbank.org/en/home>, <https://ourworldindata.org/>

**COLLECTION METHODOLOGY**

The data was collected with web scrapping

**From Data Entry to CEO: The AI Job Threat Index**

<https://www.kaggle.com/datasets/manavgupta92/from-data-entry-to-ceo-the-ai-job-threat-index>

This CSV file contains 4706 rows and 6 columns. 3 columns are strings, 2 are integers and 1 is a float.

Contained within this file is a comprehensive dataset examining the interplay between job roles and the pervasive influence of artificial intelligence (AI). Each entry meticulously categorizes job titles, shedding light on AI's potential impact, the workload distribution between human tasks and AI models, and the domain to which each job belongs. This database contains 4706 unique values or jobs.

**SOURCES**

The data has been sourced primarily from leading AI tools websites that provide insights into job analytics, AI implementations in the corporate world, and the broader impact of AI on the job market. By combining data from multiple trusted platforms, the dataset offers a holistic view, capturing both the depth and breadth of the evolving job landscape in the AI era.

**COLLECTION METHODOLOGY**

The "Job Threat Index" dataset was meticulously assembled using a combination of web scraping techniques and manual curation. Advanced scraping tools were employed to extract relevant data from a selection of prominent AI tools websites known for their insights into job analytics and AI's workplace impact. These tools navigated the intricate web structures, extracting job titles, AI impacts, task counts, and other relevant metrics. Post-extraction, the data was subjected to rigorous manual reviews, ensuring any anomalies, outliers, or inconsistencies were addressed. This dual approach, blending automated data harvesting with human oversight, ensured the dataset's comprehensiveness, accuracy, and relevance to the overarching theme of AI's influence on the modern job landscape.