CSCE 567 Final Visualization Project Proposal

Name: Khushal Pandala

Project Purpose

The project's goal is to analyze the increasing energy demands of AI models and encourage more sustainable computing methods. It aims to use data visualization to explore trends in AI energy consumption, compare the efficiency of various models, and explore potential solutions for reducing environmental impact.

Introduction / Background

Artificial intelligence (AI) has changed many industries, but as the complexity and size of AI models increased, there has been an increase in energy consumption and carbon footprint. Large models such as GPT-3, GPT-4, and BERT require enormous computational power, which reflects high electricity usage and environmental impact.

The aim of this project is:

- To visualize critical data points related to AI energy consumption
- Compare the energy efficiency of different AI systems and computer systems
- Detect patterns of AI power usage over time and by geography
- Detect potential solutions for improving AI sustainability

The project will deliver interactive visualizations to clearly communicate these results.

Audience

The primary audience for this project includes:

- **AI researchers and developers** interested in optimizing model efficiency
- **Environmental advocates** concerned about AI's carbon footprint
- **Policymakers** considering regulations on sustainable computing
- **Technology enthusiasts** curious about AI's energy consumption trends

Dataset(s)

To effectively analyze and visualize AI energy consumption, the project will utilize the following datasets:

1. ML CO_2 Impact Calculator Data

Description: Estimates the carbon emissions produced by training machine learning models based on hardware type, usage duration, and cloud provider details.
Access: While the calculator itself does not provide a downloadable dataset, it allows estimates for specific models. The mlco2/impact GitHub repository may offer further insights.
Attributes:

- **Inputs:** Hardware type, hours used, cloud provider, region
- **Outputs:** Estimated carbon emissions (kg CO_2)

2. DeepEn2023 Dataset

Description: Focuses on energy measurements for edge AI applications, covering kernels, deep neural networks, and AI workloads.
Access: Available at DeepEn2023 project page and GitHub repository.
Attributes:

- Kernel-level, model-level, and application-level energy consumption (Joules or kWh)

- AI workload types (e.g., CNNs, RNNs, transformers)
- Edge device specifications

3. Our World in Data: CO_2 and Greenhouse Gas Emissions

Description: Provides global CO_2 and greenhouse gas emissions data, offering context for AI's impact.
Access: Available for download in CSV format: CO_2 and Greenhouse Gas Emissions - Our World in Data
Attributes:

- Annual CO_2 emissions per country (metric tons)
- CO_2 emissions per capita
- Greenhouse gas emissions by sector

Proposed Visualizations

The project will present interactive visualizations to analyze AI energy consumption and carbon emissions.

1. Line Chart – AI Energy Consumption Trends Over Time

Data Source: DeepEn2023 (Model-level energy consumption trends)

- **X-axis:** Years (2010–2025)
- **Y-axis:** Energy consumption (kWh)
- **Interactivity:** Hover for details on major AI breakthroughs (e.g., GPT-3 release)

2. Bar Graph – Energy Usage Across AI Models

Data Source: DeepEn2023 & ML CO_2 Impact Calculator

- **X-axis:** AI model (GPT-3, GPT-4, BERT, ResNet, etc.)
- **Y-axis:** Energy consumed per training session (kWh)
- **Interactivity:** Filter models by category (NLP, vision, reinforcement learning)

3. Heat Map – Carbon Footprint of AI Training by Region

Data Source: Our World in Data (Global CO_2 emissions)

- **Color-coded regions:** AI-related energy consumption
- **Interactivity:** Click on a region to view AI-related emissions data

4. Pie Chart – AI Energy Consumption by Computing Infrastructure

Data Source: ML CO_2 Impact Calculator

- **Categories:**
  - Cloud providers: Google, AWS, Microsoft
  - On-premise clusters
  - Supercomputers
- **Interactivity:** Click to view trends over time

Additional Visualizations

- **Scatter Plot:** AI model performance (accuracy) vs. energy consumption (kWh)
- **Network Diagram:** AI model complexity vs. power usage

Plan

Tools & Technologies

- **R (ggplot2, plotly, Shiny)** – For data visualization and interactivity
- **Python (Pandas, Matplotlib, Seaborn)** – For data cleaning and preprocessing
- **D3.js / Tableau** – For enhanced visualization options

Delivery Method

Final Product: A publicly accessible website presenting interactive visualizations.

- **Plan A:** Shiny Web App (Users explore AI energy trends interactively)
- **Plan B:** Web-hosted report (Embedded visualizations for static analysis)

Conclusion

This project will quantify the environmental impact of AI models and provide insights into sustainable computing strategies. Through interactive data visualizations, it will raise awareness and help AI researchers, policymakers, and environmental advocates make informed decisions.