CSE 3000 - Quantitative Bias Analysis

Objective

Quantitatively analyze a chosen platform for bias (e.g., racial, gender, or political) and compare it against at least two competitors in the same field.

Assignment Details

Part 1: Experiment Design and Execution

- **Choose a Bias Type and Platforms:**
  - Pick a type of bias (e.g., racial, gender, political).
  - Select one platform (e.g., DALL-E, Zillow, LinkedIn) and at least two competitors (e.g., Stable Diffusion, Redfin, Indeed).
- **Define Test Cases:**
  - Design a set of quantitative tests related to your chosen bias. For example:
    - For DALL-E and competitors: Create prompts you suspect will demonstrate bias (e.g., "professor," "nurse," "prisoner").
- **Collect Data:**
  - For image generation platforms:
    - Generate a fixed number of images per prompt (e.g., 50 images for "professor").
    - Use image analysis tools (e.g., facial recognition APIs) to predict demographics like race or gender.
  - For text or recommendation platforms:

    - Analyze textual outputs or recommended options (e.g., job titles, property listings).

- **Analyze Results:**
  - Quantify outputs using metrics like:
    - **Distribution Accuracy:** Compare demographic outputs against real-world data (e.g., employment or population statistics).
    - **Bias Indicators:** Use statistical methods (e.g., T-Tests, ANOVA, Kullback-Leibler divergence) to quantify disparities.
  - Visualize results with examples like:
    - Stacked bar charts showing demographic representation.
    - Heatmaps comparing platforms.

- **Define Accuracy Metrics:**
  - Students must justify a quantitative definition of "accuracy," such as:
    - **Real-world Accuracy:** Closeness to real-world employment statistics.
    - **Ideal Representation:** Closeness to equal representation across demographics.

Part 2: Ethical Analysis

- **Identify Disparities:**
  - Discuss any observed demographic disparities in outputs.
  - Highlight areas where platforms reinforce stereotypes or misrepresent groups.

- **Propose Solutions:**
  - Suggest actionable ways to reduce bias, such as:
    - Balancing training datasets.
    - Adjusting output algorithms to favor diversity.
  - Reflect on trade-offs (e.g., accuracy vs. fairness).

Deliverables

- **Code Submission:**
  - Provide a Python script or Jupyter Notebook that implements the tests and analysis.
  - Include well-documented code and instructions.
- **Presentation (5-10 minutes):**
  - Present findings using visualizations (e.g., charts, graphs).
  - Discuss results, accuracy metrics, and ethical implications.
- **Technical Report (3-4 pages):**
  - Summarize methodology, results, and ethical analysis.
  - Include quantitative results, justifications for accuracy definitions, and proposed solutions.

Example Workflow

- **Generate Images (for DALL-E, MidJourney, Stable Diffusion):** Produce outputs for analysis.
- **Analyze Demographics:** Use a facial recognition API to predict gender or race.
- **Quantify Bias:** Compare distributions to real-world or ideal data.
- **Visualize Results:** Use `matplotlib` or `seaborn` for comparisons.

Ethical Reflection Prompts

- Should platforms strive for real-world accuracy (mirroring employment statistics) or ideal representation (equal demographics)?
- How might biased training data affect outputs, and what is the platform's responsibility to address this?
- What are the risks of over-correcting for bias? How could this impact user trust?

Grading Criteria

- **Presentation and Report (40%):** Clarity, coherence, and depth of the report and presentation.
- **Experimental Design and Implementation (20%):** Quality of tests, thoroughness of data collection, and correctness of code.
- **Quantitative Analysis (30%):** Depth of insights, quality of statistical analysis, and visualizations.
- **Ethical Reflection (10%):** Originality and rigor in addressing trade-offs and proposing solutions.