Assignment 2 – Artificial Intelligence Tools

Topic: AI Bias & Fairness Testing Bias in AI Career Suggestions

Creators: Shaima Nigem, Selen Mahajna

1) Task (What we did)

The goal of this project was to test whether AI tools can produce biased or stereotype-driven outputs when asked about professions and career suggestions. We compared answers from two AI tools using the same prompts, then checked whether adding bias-reducing constraints would improve fairness and diversity in the suggestions.

2) Tools (and why we chose them)

Tool 1: ChatGPT (LLM / Generative AI)

- **Ease of use:** Very easy to use with clear prompts and fast responses.
- **Technology behind it:** A Large Language Model (LLM) that generates text based on patterns learned from large datasets.
- **Reliability:** Helpful for drafting and summarizing, but outputs can vary between runs and may reflect dataset patterns; requires careful prompting and human judgment.

Tool 2: Google Gemini (LLM / Generative AI)

- **Ease of use:** Simple interface and quick responses; good for alternative phrasing and comparison.
- **Technology behind it:** Also an LLM that generates text responses from learned patterns in data.
- **Reliability:** Often produces well-structured answers, but may still show bias depending on how the prompt is phrased; results should be checked and compared.

3) Method (Process)

We ran **three experiments**, each with two prompt versions:

- **Prompt A (general):** asked normally without constraints.
- **Prompt B (bias-reducing):** added constraints such as “use gender-neutral language” , “avoid stereotypes” and “balanced across fields” .

Experiments: **Engineer**, **Doctor**, and **Jobs for a caring person**.
For each experiment, we collected outputs from **ChatGPT** and **Gemini**, then noted bias indicators (stereotypes, narrow role clusters, gendered framing) and compared Prompt A vs Prompt B.

4) Final Output (Website)

Our final output is a website that presents the experiments, screenshots as evidence, and the comparison results. The website matched our expectations: it clearly shows how prompt wording affects the diversity of career suggestions. If we had more time, we would test more professions and traits and measure bias more systematically (e.g., more runs per prompt and clearer scoring criteria).

5) Reflection

Did we enjoy it? Yes , because it demonstrated that small prompt changes can improve fairness and diversity.
Advantages: fast to reproduce, easy to compare tools, shows the impact of prompt design.
Disadvantages / Challenges: outputs vary between runs; bias evaluation is partly qualitative; careful prompt design is needed to reduce stereotype-driven narrowing.