Workshop Exercise Synthetic Data for Research: Epistemic and Practical Directions

Tsehaye Haidemariam, PhD

Prepared for:

Generative AI for Social Science Research

Norwegian Business School Workshop | 2025

30-Minute Practical Exercise (Group-Based or Individual)

 Title: *Stakeholder Perceptions of CSR – A Synthetic Survey*

Goal: Generate and explore a synthetic dataset on public trust in corporate social responsibility using GPT-4.

 Instructions for Participants

Step 1: Generate Data (5–8 min)

Prompt GPT-4 with:

"Create 15 survey responses. Each should include: name, age, industry, CSR_awareness (1–5), Trust_in_corporation (1–5), and a short opinion on the company's social responsibility."

You may use ChatGPT (Pro/GPT-4) or paste this prompt into a shared notebook if using code.

Step 2: Analyse (10 min)

Ask GPT-4 (or a provided script) to:

- Convert it to a DataFrame
- Group by industry and show average trust

Optional: Correlation between age and trust

Step 3: Visualize (5–7 min)

Ask GPT-4:

"Create Python code to show a bar chart of average CSR awareness by industry."

Run it in Jupyter/Colab or just interpret the code if no runtime is available.

Step 4: Reflect & Share (5 min)

Quick group discussion or post-it notes on:

- Did the data look realistic?
- What would you trust this data for—or not?
- How could this help business/social science researchers?