Preregistration Statement # 1

Hypothesis: Remote work reduces the salary premium associated with higher experience levels in the US AI job market. Specifically, the percentage increase in salary from entry-level to senior-level positions will be smaller for fully remote workers compared to on-site workers.

Analysis: Run a linear regression model with salary_in_usd as the dependent variable and experience_level (categorical: Entry, Mid, Senior, Executive), remote_work (binary: 0 = on-site, 1 = fully remote), and their interaction term (experience_level × remote_work) as independent variables. The interaction coefficient will test whether the relationship between experience level and salary differs based on remote work status. If the interaction coefficient is negative and meaningful, this supports the hypothesis that remote work flattens the salary hierarchy. If the coefficient is near zero or positive, remote work does not reduce experience-based salary differences.

Preregistration Statement # 2

Hypothesis: Specialized technical roles (Machine Learning Engineers, Research Scientists) show a steeper salary trajectory across experience levels compared to generalist roles (Data Scientists, Data Analysts), and this difference is more pronounced for on-site workers than remote workers.

Analysis: Calculate median salaries for each combination of experience_level (Entry, Mid, Senior, Executive), role_type (Specialized vs Generalist), and remote_work (on-site vs fully remote). Compute the experience premium as: (Senior Median - Entry Median) / Entry Median × 100% for each of the four groups (Specialized On-site, Specialized Remote, Generalist On-site, Generalist Remote). Compare these premiums to determine whether specialized roles have larger premiums than generalist roles overall, and whether the difference between specialized and generalist premiums is larger for on-site workers than remote workers. Visualize results with grouped bar charts showing premiums across the four groups.

Questions for reviewers:

1. Does the interaction term mentioned in the preregistration statement #1 approach adequately test whether remote work reduces salary hierarchies, or is there a more direct way to test this hypothesis?

2. Is comparing entry-to-senior premiums the most meaningful way to measure salary trajectory differences, or would examining the full progression (entry → mid → senior → executive) be more informative?

3. Given that my dataset has binary remote work (79% on-site, 21% fully remote with minimal hybrid), does this limitation fundamentally undermine my ability to draw conclusions about remote work's effect on salary hierarchies? If so, what can I do to possibly make up for this lack of complexity?