Experimentation Plan for TrialGPT

Goals:

1. **Evaluate Accuracy**: Assess TrialGPT's ability to match patients with appropriate clinical trials with a focus on criterion-level and trial-level accuracy.
2. **Measure Efficiency**: Quantify time savings for clinicians using TrialGPT versus manual trial matching processes.
3. **Evaluate Usability**: Gather feedback from clinicians on the interpretability and ease of use of TrialGPT's explanations and rankings.

Plan Overview:

1. Setup and Data Preparation

- **Patient Data**: Initially it would be a good choice to use synthetic patient data from Synthea for monitoring the model performance and then conduct physical trials.
- **Trial Data**: Use clinical trials from ClinicalTrials.gov that are actively recruiting.
- **Test Groups**:
    - **Control Group**: Clinicians perform manual patient-to-trial matching.
    - **Experimental Group**: Clinicians use TrialGPT for patient-to-trial matching.

2. Evaluation Metrics

- **Accuracy**:
    - Criterion-level prediction accuracy compared to expert annotations (target ≥ 85%).
    - Trial-level ranking precision (NDCG@10, P@10).
- **Efficiency**:
    - Measure reduction in time spent on matching tasks (target ≥ 40% time reduction).
- **Usability**:
    - Collect clinician feedback through surveys, focusing on the clarity of explanations and the ease of integrating TrialGPT into their workflow (target ≥ 80% satisfaction).

3. Pilot Testing

- **Sample Size**: Conduct a pilot with 20 clinicians, using 100 patient-trial pairs.
- **Tasks**: Each clinician will screen 50 patient-trial pairs, half with and half without TrialGPT.

- **Outcomes**: Compare accuracy, time taken, and satisfaction between the control and experimental groups.

4. Launch Decision Criteria

- **Accuracy**: TrialGPT achieves accuracy within 5% of expert-level predictions.
- **Efficiency**: At least 40% time savings for clinicians in the experimental group.
- **Usability**: ≥ 80% of clinicians report that TrialGPT explanations are clear and useful.

Launch Decision:

- If TrialGPT meets or exceeds the goals for accuracy, efficiency, and usability, proceed with a broader clinical launch. If any metric falls short, refine the model and repeat testing with a larger sample size before launch.