Bonus

Suggest a better implementation plan.

Here is what we can do to improve the implementation:

1. **Dynamic Epsilon Decay in Epsilon-Greedy**
   - Instead of using a fixed decay rate eps=1/t, we can implement a more flexible decay mechanism, such as exponential decay: eps=eps0*e^(-kt), where eps0 is the initial value, and k is a decay constant.
   - Additionally, we can set a lower bound for epsilon (e.g., eps ≥0.01) to ensure continued exploration throughout the experiment.
2. **Bayesian Updates for Precision in Thompson Sampling**
   - We can use Bayesian updating to adaptively calculate the precision of each bandit's distribution based on observed rewards, rather than assuming a known precision.
   - For instance, we could model rewards using a Beta distribution for binary outcomes or a Gaussian distribution for continuous outcomes and update the parameters dynamically as new data arrives.
3. **Optimize for Bandit-Specific Parameters**
   - Before starting the experiment, we can introduce a tuning phase to determine the optimal initial parameters for each bandit.
   - This could include setting the starting epsilon for Epsilon-Greedy or selecting the initial priors for Thompson Sampling.
4. **Multi-Armed Bandit with Contextual Data**
   - To enhance decision-making, we could extend the implementation to a **Contextual Bandit** problem if user-specific data is available.
   - This would involve incorporating features (e.g., user demographics, behavior) into the algorithm to optimize ad selection for different user segments.
5. **Hybrid Algorithm**
   - A potential improvement is to combine Epsilon-Greedy and Thompson Sampling into a hybrid approach.
   - For example, we could use Epsilon-Greedy in the initial phase of the experiment to ensure broad exploration and then switch to Thompson Sampling for more effective exploitation based on learned parameters.