We would like to thank all reviewers for your valuable feedback to help improve our paper.

Reviewer #1: <u>3. Weaknesses</u>: • RGPE and TAF are the only points of comparison that we are aware of, and they can be easily adapted for the federated setting (lines 174-177). We'll revise the paper to make it clearer that RGPE and TAF are not developed for the setting considered in this paper. • We agree that other approximation techniques for GP (e.g., inducing point methods) could also be used to avoid the sharing of raw data. It would be interesting (and potentially challenging) to explore whether they can also tackle the key challenges of federated BO in a principled way. <u>4. Correctness</u>: • You are right that selecting x_t using ω_n is in fact a non-linear optimization problem due to cosine non-linearities. Fortunately, it can still be efficiently solved in our experiments using the DIRECT method (taking on average 0.76s/iteration for $M = 200$, landmine experiment). Thank you for pointing this out and we'll correct this. <u>8. Additional feedback</u>: • Regarding the computational cost of sampling ω_n, we'd like to clarify that the size of the matrix Σ_t^{-1} is in fact $M \times M$ (Eq. 3) and independent of t, where M is the number of random features and hence a constant. For the values of M used in our experiments (up to 200), we've found the computational cost to be reasonable, i.e., in the order of 0.1s. We'll discuss this in the paper. • The horizontal line in Fig. 2 represents the performance of TS.

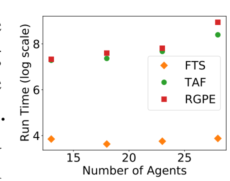
Thank you for your careful review and we'll also address all your other comments when revising our paper, e.g., discussing why FTS performs better than TS and revising inappropriate statements.

Reviewer #2: <u>3. Weaknesses</u>: • Your suggestion on perceiving our algorithm as BO under model mixtures is very interesting, and we'll explore whether it can be cast in this alternative interpretation. We'll explore these literature, and also add references to multi-fidelity/multi-source BO as you suggested. • We agree that considering the level of fidelity of different agents (i.e., similarity to the target agent) is an interesting extension and have discussed it in lines 344-346. <u>4. Correctness</u>: • The theoretical guarantee in lines 106-108 is in fact a high-probability guarantee, which we omitted for simplicity. We'll add this after revision. • For the claim in lines 131-132, you are right that we have assumed that the kernel is bounded (line 97). We'll revise and make our assumptions clearer.

Thank you for your constructive suggestions. We'll also take into account all your other comments to revise our paper.

Reviewer #3: <u>3. Weaknesses</u>: • We would like to clarify that we have shown that our FTS algorithm is efficient in terms of both computation (please refer to Fig. 3) and communication (please refer to Fig. 2).

Reviewer #4: <u>3. Weaknesses</u>: • *Experiments*: It's indeed of interest to us to optimize challenging functions, which we have done using the real-world experiments. Meanwhile, we use synthetic experiments to verify the practical relevance of theoretical results and to investigate the behavior of our method, which is made simpler by easier synthetic functions. Thompson sampling (TS) in the non-federated setting simply runs standard TS for a single agent, with no communication with any other agent. In Figs. 2 and 3, the number of sampling iterations for TS is 50, which is the same for all methods. The performance advantage of FTS is consistent across all 3 real-world experiments, showing that it's stable against variations in various factors such as optimization function, number of agents, etc. We'll add more details on vanilla TS and more discussions about the experiments.



• *Secondly*, as you suggested, we added an experiment to investigate scalability w.r.t. the number of agents (refer to the figure, landmine experiment). The results show that our FTS is more scalable w.r.t. the number of agents, which verifies our analysis in lines 194-202. Regarding our limitations, Fig. 1c in fact contains a failure case. It shows that when all other agents are heterogeneous, FTS can converge slightly slower than TS if p_t doesn't grow sufficiently fast. However, Fig. 1c also shows that FTS still converges to the same final performance as TS, and this limitation can be alleviated by making p_t grow faster (red curve). Another limitation on susceptibility to advanced privacy attacks and opportunity for future work are discussed in the Broader Impact section. • *Privacy Preservation (line 50)*: You are right that RFF only allows us to *retain the raw data* and avoid transmitting it, so that we can pass RFF parameters *in the same way as standard federated learning* (lines 60-63). We'll revise the paper to reflect this. We'll also explore effective means to preserve privacy in future work (Broader Impact section). • *Communication Efficiency (line 189)*: We agree that we can only claim our FTS is more efficient in communication for a fixed experimental setting and a fixed approximation quality (fixed M). We have shown this in our experiments (Fig. 2). We'll revise to make our claim more specific and accurate. <u>6. Relation to prior work</u>: • We agree that previous works on parallel/distributed BO/TS are also related. We'll cite these works such as the one you suggested, and discuss their relationship with our work. Please also note that there are fundamental differences: they usually optimize a single objective function while we need to consider different functions from (potentially highly) heterogeneous agents; they usually allow the sharing of raw data. <u>8. Additional feedback</u>: • We would like to clarify that sharing of random features $\phi(x)$ can be achieved by sharing a *finite* number of parameters whose dimensions depend on D and M (Appendix A), even if $\mathcal{X}$ is uncountable.

Thank you very much for you thorough and constructive review, and we'll also address your other comments to revise our paper. We hope that our clarifications and additional results can improve your opinion of our work.