We would like to thank the reviewers for their comments and helpful suggestions. We respond below to the reviewers' comments and criticism.

- *R1: The term "privacy model" is confusing:* We will definitely think of a better term. Perhaps, "privacy-data" model is less confusing.
- *R1: More discussion of the broader impact"* We will elaborate more on the potential impact of the proposed framework. We believe it can be a basis for a more flexible, more expressive framework for DP learning.
- *R3: "Intuitively, the sample complexity should be comparable to that of the non-private one":* We disagree with the reviewer's intuition about the problem. The question of the sample complexity under this model is quite different from its non-private counterpart. In particular, our construction is tailored to halfspaces, and – as R4 pointed out – extending this to general VC classes is non-trivial (and is an open question).
- *R3: "I'd expect some simple but a bit more advance model, like a Bernoulli model . . . ":* The general framework enables studying settings such as the one suggested by the reviewer (as well as many others). In this work, we focus on this special setting (label-determined) to demonstrate the capacity of the proposed framework to capture realistic settings, which were not explored by previous works that studied utilizing public data.
- *R3 & R4: label-determined privacy model is too restrictive:* The main goal of this work is to propose a new, formal, more flexible framework for DP learning that captures more realistic scenarios than prior works. The result concerning the label-determined model serves mainly as a proof of concept that demonstrates the capacity of this framework to capture new interesting settings. Another contribution of our result for learning half-spaces is the kind of technical tools used in the construction (which are less commonly used in learning theory). See also our responses below to R4's comments. We also want to point out that the label-determined model does capture some realistic scenarios. For example, imagine a scenario where the data of individuals who tested positive for COVID-19 does not require privacy protection (to enable contact tracing and symptom analysis), while the data of those who tested negative remains protected. This is actually the case in some countries, and is exactly captured by the label-determined model.
- *R3: A short discussion on private knowledge transfer, and private learning halfspaces:* We will add a relevant discussion. Note that the prior works on private knowledge transfer (e.g., [Papernot et al. 2018, BTT18]) assume that public and private data are identically distributed. We also note that the target distribution in our case is a mixture of the (possibly different) private and public distributions. Hence, the knowledge transfer (or, domain adaptation) paradigm does not seem very useful in our case.
- *R4: "The extra novelty in the new privacy framework is that distributions of private and public data are different":* Our motivation for this new framework is to capture more realistic scenarios compared to prior work. One aspect, as the reviewer notes, is that the private and public data can arise from different distributions. Another aspect is that in this framework the examples are sampled from a single source (the mixture distribution) rather than assuming access to a separate oracle for each of the public and private examples (as in the prior work). So, unlike the prior work that utilized public data, here there is only one sample complexity, and it is given in terms of the sum of private and public examples drawn from the mixture.
- *R4: "I wonder what kind of results are even possible without making such assumptions":* There are various directions one can explore in this framework. Perhaps, among the most realistic ones are those based on a distribution-dependent model, where one also restricts the data distribution (and not only the concept class), or privacy models where the privacy status can be correlated with the feature vector not just the target label. Also, in the distribution-independent setting – as the reviewer notes – one may ask whether every VC class can be learned in the label-determined setting. (In the other extreme, where the privacy status and the label are independent, previous works [BNS13, ABM19] showed that any VC class is learnable, with significant savings in sample complexity.) Moreover, one may also consider privacy models that interpolate between these two extremes (label-determined and label-independent), and explore the sample complexity in this spectrum.
- *R4: The algorithm is improper and also not poly-time:* We note that improperness of the algorithm is not necessarily a drawback (e.g., boosting algorithms are improper). However, we think the question of whether one can construct a proper algorithm for this problem is an interesting open question as we mention in the paper. Concerning computational efficiency, note that, even without privacy, agnostic PAC learning of halfspaces is known to be computationally hard. However, the question of computational efficiency in the realizable case is yet another interesting question. Since our work makes the first step in studying this new framework, we believe it's useful to first study the sample complexity of this problem without computational restrictions.
- *R4: "Is it possible to get a sample complexity close to $O(d)$?" :* This is indeed a very good question for future work, which we have been thinking about. We will definitely add a relevant discussion to the paper.
- *R4: Comparison for the sample complexity of halfspaces with that of prior work:* We will add this comparison to the paper.