1   We thank all Reviewers for their helpful comments; specific responses are given below.

2   **Reviewer 1**

3   1. *Supplement exact copy of manuscript :* No. Beyond the main body, the supplement contains 15 pages of appendices,
4   with pseudocode, proofs of theorems, and the results of our preliminary experiments (Sect. C.3) in a tabular form.

5   2. *Main flaw is lack of experiment ... speed should be addressed:* Our experiments are a preliminary validation of our
6   theoretical results (the latter being the main meat of the paper). Despite we used only one dataset, we intentionally
7   generated 3 very diverse input trees out of it (LL. 325-344, and Tables 1 and 2), so as to enlarge the scope of this
8   empirical validation. As for speed, please observe that our theoretical results come with running time analyses.

9   3. *line 46: even split of version space:* Even split of version space (whenever possible) is a standard baseline approach
10   to active learning [12,24,14,15,26,23].

11   **Reviewer 2**

12   1. *Notation to improve:* Thanks for pointing this out, we'll find a better notation so as to avoid notation overloading.

13   **Reviewer 3**

14   1. *On motivation of our problem:* It is often the case in big organizations that data processing pipelines are split
15   into *services*, making Machine Learning solution providers be constrained by the existing hardware/software
16   infrastructure. In the Active Learning (AL) applications that motivate this work, the hierarchy over the items to be
17   clustered *is not build by us*, it is rather provided by a third party, i.e., an exogenous data processing tool that relies
18   on side information on the items (e.g., word2vec mappings and associated distance functions) which are possibly
19   generated by yet another service, etc. In this modular environment, it is reasonable to assume that the tree T is *given*
20   *to us as part of the input* of our AL problem. The human feedback our algorithms rely upon may or may not be
21   consistent with the tree T at hand both because human feedback is generally noisy and because this feeback may
22   originate from yet another source of data, e.g., another production team in the organization that was not in charge of
23   building the original tree. In fact, the same tree over the data items may serve the clustering needs of different groups
24   within the organization, having different goals and views on the same data. This also motivates why we are led to
25   consider different noise scenarios (realizable, noisy realizable, and nonrealizable). In short, we are not artificially
26   constraining ourselves to clusterings realized by a tree, since the tree is itself part of the input. Our preliminary
27   experiments have perhaps been a bit misleading, in that the trees we managed to generate ourselves.
28   The above can be added to the paper to better motivate our investigation.

29   2. *On the objective of the learning process and clustering closeness:* We are not 100% sure we fully understood this
30   comment. If Σ (the human feedback) is consistent with T (realizable setting) then being close in Hamming distance
31   to the underlying clustering represented by Σ implies being close to that clustering in any "reasonable" sense. In
32   particular, zero Hamming distance implies the two clusterings coincide. On the other hand, if Σ is not realized by T,
33   then we are either in the noisy realizable or in the nonrealizable setting, where the goal is different (like bounding the
34   excess risk in Eq. (2)). On the contrary, if the reviewer is alluding to using a distance metric other than Hamming,
35   e.g., one that depends on the structure of T, this is a relatively easy adaptation of what is currently in the paper. The
36   very the reason why we restricted ourselves to Hamming distance (over matrices) stems from our need to treat in a
37   unified manner both the (noisy) realizable and the nonrelizable settings, since some such alternative distances would
38   only apply to the realizable setting (with or without noise), but not to the nonrealizable one with i.i.d. entries.

39   3. *It appears that any practical ... fall into the non-realizable case:* This statement looks too broad to be considered
40   undisputed ... As a striking counterexample, one of the findings of our experiments (see Appendix C.3) is that
41   AL algorithms assuming persistent noise can in practice be more effective than those making the more general
42   nonrealizable assumption. Notice that in our experiments Σ has been generated by the MNIST class labels, hence Σ
43   has virtually nothing to do with the trees we generated, which in turn do not rely on those labels at all. This offers a
44   hopefully clearer interpretation of our empirical findings, which we can further elaborate upon in the paper.

45   **Reviewer 5**

46   1. *On motivation:* Please see response 1 to Reviewer 3.

47   2. *Related work and tradeoffs:* Thanks for bringing the first three references to our attention, we shall duly compare to
48   them. In the time frame of this rebuttal, what we can say is that it seems like these papers are not readily comparable
49   to ours, since the noise assumptions are slightly different, e.g., the uneven noise gap analysis in Firmani et al. (Thm
50   1 and 2), and Prop. 3-5 therein. Anyhow, the reviewer is right in that, being restricted to a clustering realized by
51   T, we generally need less queries. For instance, combining our Thms. 2 and 3, our bound in the realizable case
52   is between $\mathbb{E}[K]$ and $\mathbb{E}[K]\log h$, while in general it takes $n\mathbb{E}[K]$ queries to fully reconstruct the clustering [13].
53   Also recall the lower bound $n - K + \binom{K}{2}$ in Wang et al., and Firmani et al. As for the other papers the reviewer is
54   mentioning, we'll definitely cite them, but they seem to be facing the problem of building the hierarchy, rather than
55   cutting a pre-existing one (by the way, another reference along this line is [11]). It is currently unclear to us which
56   comparison can be made.

57   3. *More details in the non-realizable case:* Thanks for your suggestions, we'll add more illustrations.