I. REPLY TO THE FIRST REFEREE

II. REPLY TO THE SECOND REFEREE

Dear referee,

we appreciate the valuable insights on how to improve our article that you have provided us with. Below we address each of the emphasized points separately, as they were written in the reply that was sent back to us. We quote the issues raised in italic, and highlight in boldface the points that we thought were particularly relevant.

Major

I am not sure about the novelty and generality of their method. For example, ***their pipeline cannot work for complex experimental physical systems, where the Hamiltonian could be unknown [1]*** *. Moreover, the authors state that it is sufficient to use the input data in real space to predict the topological phase with high accuracy.* ***It is not clear to me what "sufficient" and "high accuracy" mean [2]*** *, and the advantages of using the input data in real space instead of the Hamiltonian in wavevector space.* ***I expect the evidence to indicate that their method is superior or comparable to any other method [3]*** *. It is worth mentioning that the author already discussed* ***the motivations for developing a data-driven approach based on real space [4]*** *in the "Learning topological phases from real space data" section. However, these arguments should be pointed carefully in the "Introduction" section.*

[1] Mention the possibility of using incomplete information with eigenvector ensembling.
[2] ...
[3] Mention trends in model explainability as well as performance scores.
[4] We agree that the motivation for using real space data to study topological phase transitions should be placed in the introduction. We have therefore moved the first paragraph of the section "Learning topological phases from real space data" to the introduction.

I am not sure which crucial problems their algorithm can solve that was not possible before with machine learning. Furthermore, along with the emerging research of unsupervised learning methods in realizing the topological phases, ***why should the supervised learning method be focused in this context? [5]*** *In the present stage of this manuscript, where I do not see the proposal's advantages, I would prefer the* ***unsupervised approaches that could grasp information from a given system without knowing their phases [6]*** *. In my opinion, a proper intrinsic route for understanding physical systems without much information on the dynamics of the system will lead to significant future developments and a better understanding of physic systems.*

[5] Mention semi-supervised learning, model explainability.
[6] Mention Hamiltonian compression, of unsupervised learning through dimensional reduction.

The details of ***the eigenvector ensembling algorithm should be addressed in the main text [7]*** *instead of supplemental material. What is the specific algorithm used in the paper to train on eigenvectors? I think it is important even if the readers are not familiar with some ML techniques. Some readers in physics may find it difficult to understand some ML technical terms such as bootstrapping, training and validation, test sets, etc. In addition,* ***I expect proof of what "physics" their method captures and why the eigenvector can play an important role in characterizing the topological phase [8]*** *. If this physical interpretability is not mentioned, it is very difficult to see the contribution of the method in physics.*

[7] We have moved the section "The eigenvector ensembling algorithm" to the main paper. The issue of specifying the particular learning algorithms used was also addressed.
[8] One of the strongest points of the paper is to introduce model explainability tools from machine learning in physics. The need to use eigenvectors is that they encode all the physical information from a system. The paper paves the way for new theoretical investigation on the relations between the shannon entropy of eigenvectors and topological phase transitions.

The authors made efforts to analyze how the algorithm was able to recover the Hamiltonians' global property in the "Information Entropy Signatures" section. The authors state that learning topological phases from local real-space data in bulk is still possible even for small subsets of lattice sites, then refer us to the section "Learning topological

phases from real space data" in the Supplementary Material. The authors mention on page 4 of the Supplementary Material that ***"key topological information can be said to be localized on a few lattice sites," which is a particularly interesting statement to me [9]*** *. However, I fail to understand the physical insights behind it. Without the proper explanation,* ***it is difficult to see the effectiveness of their method or verify the method with a more complex model [10]*** *instead of the simple SSH form.*

[9] Move this sentence to main paper.

[10] Add FFT of entropy signatures? Mention entropic uncertainty? Talk about information gain? Mention that interpretability in ML is often challenging and a fast evolving field. Furthermore, eigenvector ensembling can always be used as a preprocessing step to other ML models, such as in semi supervised learning tasks.

Minor

The authors state in the abstract that "model explainability in machine learning can advance the research of exotic quantum materials with properties that may power future technological applications such as qubit engineering for quantum computing." ***Could you explain a bit more about properties that may power future quantum computing? [11]***

[11] Insert comments about topological qubits?

On page 4, the authors state that "Figures 2 and 3 illustrate single iterations of experiments 1 and 2 as seen from parameter space". ***What are the experiments 1 and 2? [12]***

[12] Experiments 1 and 2 were mentioned as subtitles to the sections.

The authors should explain the evaluation metric in the main text [13] *, such as "the accuracy" and "the probability heatmap," "eigenvector accuracy," and "Hamiltonian accuracy."*

[13]

I prefer to ***put the definition of information entropy signatures in the main [14]*** *text to help the readers understand what the method tries to do.*

[14]