

### Motivation:

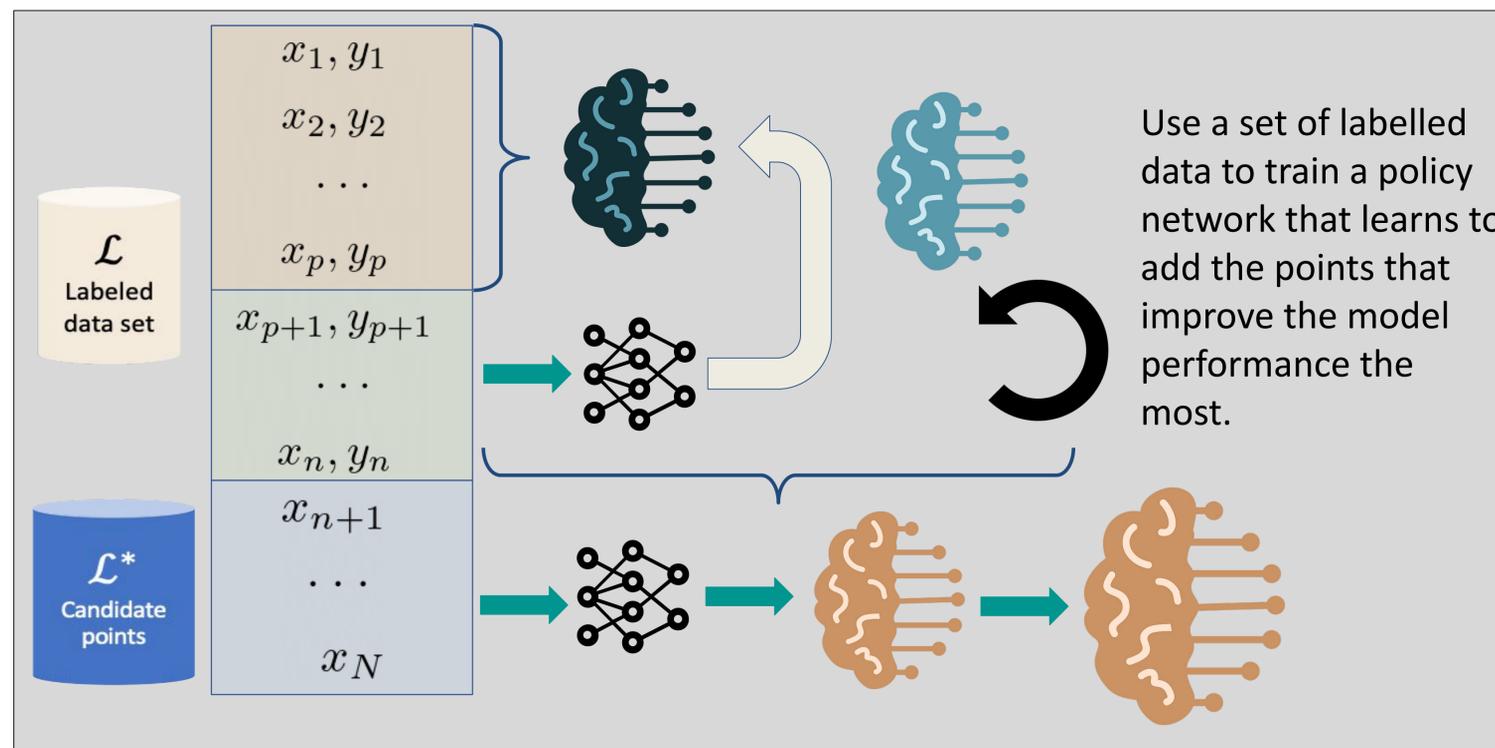
Efficient use of data is key when doing drug development given the high cost of production. Exploiting the structures of the sampled data could improve the way we use it

### Research Question:

1. Can a policy be designed that can improve the balance between exploration and exploitation?
2. Could this policy perform better than the previously heuristically determined policies thanks to the data structure?

### Method description:

We proposed a method based on deep Q learning that adds an ancilla model that aims to model the policy by finding a balance between exploration and exploitation.



### Description of results:

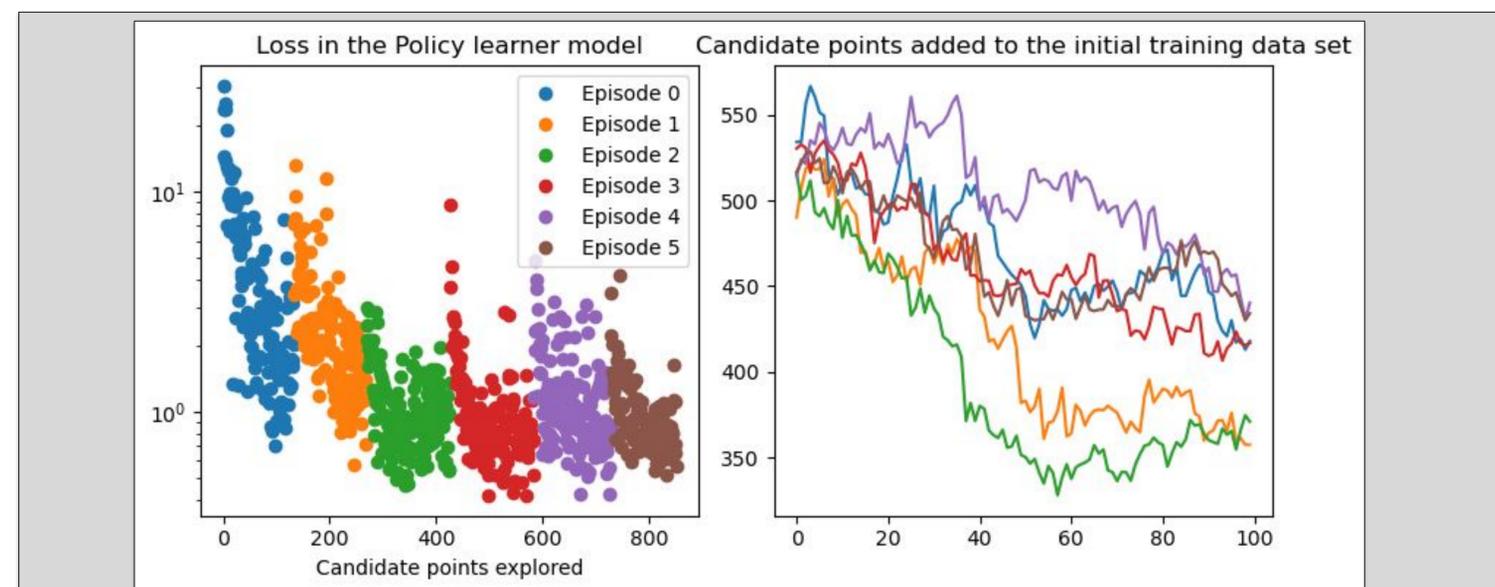
- We tried this technique on two simulated data sets: prediction of hydrogenic orbital probabilities and regression simulated data
- We found that training the learning network is a major challenge and that perhaps a single layer is not adequate
- The application of the method to model training seems discouraging

### Next Steps:

- Experimentation with the policy learner architecture
- Application to efficient optimization techniques like bayesian optimization.

### Acknowledgements

**Merck:** Xiang Yu, Terri Bui, Kenneth Wilson, Rajesh Desai  
**Purdue Data Mine:** Dr. Mark Ward, Maggie Betz, Nicholas Rosenom, Kali Lacy



For a simulated regression data set, from the sklearn package, we propose a single layer network as our policy learner and train in episodes feeding streams of labelled data and keep track of the loss of the principal model