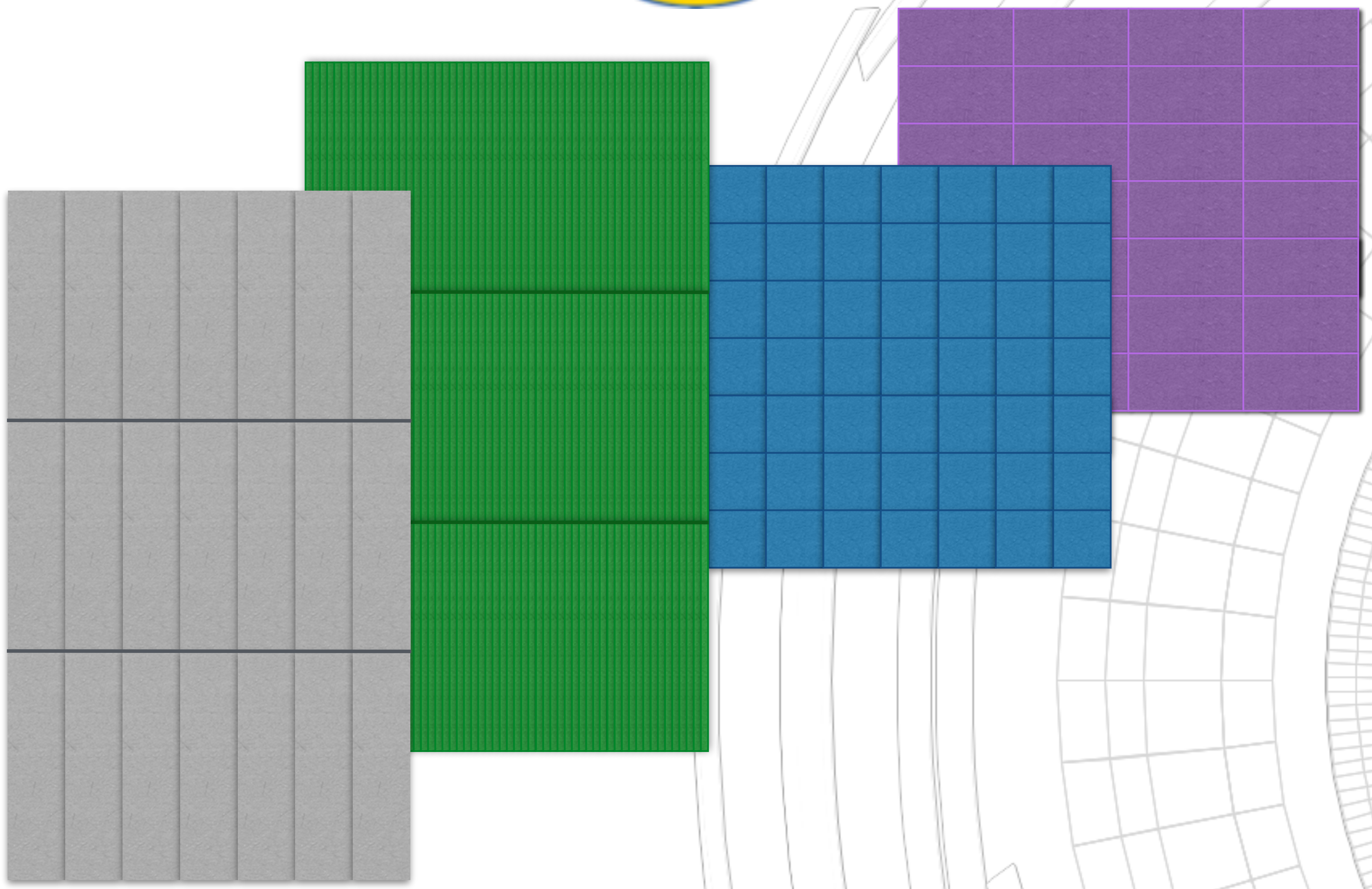
Summary: Uncertainty, Validation, Interpretation of ML

Aishik Ghosh (for all the contributors)

Seattle Snowmass Summer Meeting

19 Jul 2022



Science demands more from ML

When studying the universe, we don’t just want the facts, we want to understand why things work the way to do, and not some other way

Similarly demand explanations, not just accurate predictions from ML

Important to validate that patterns learnt from simulations will translate to data

Techniques emerging to use ML to propagate and mitigate uncertainties with ML and also quantify uncertainties of ML models themselves

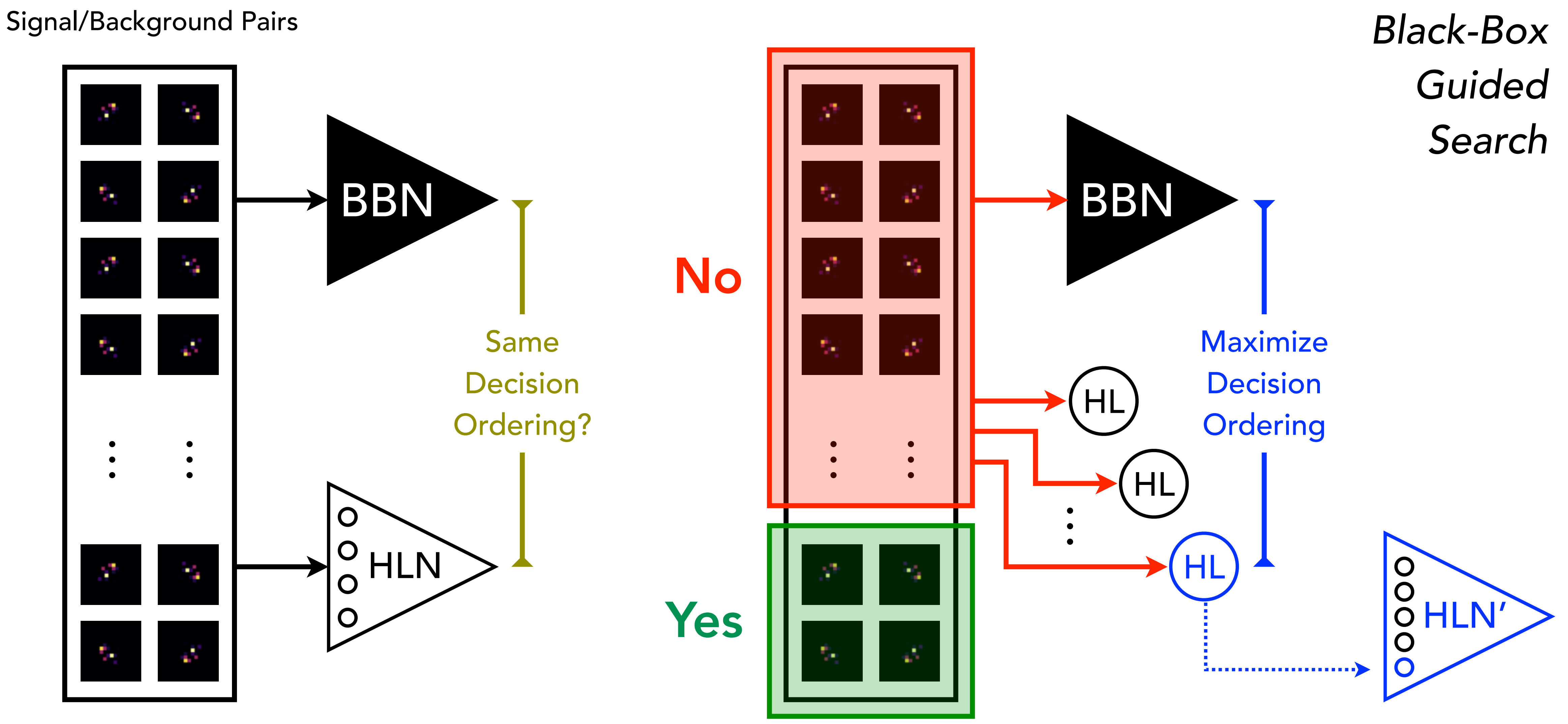
Interpretable ML

What information is the model using?

In simple cases (using high level features), drop an input feature or decorrelate the model from feature to find feature importance

For unstructured data like images, new methods are being explored.

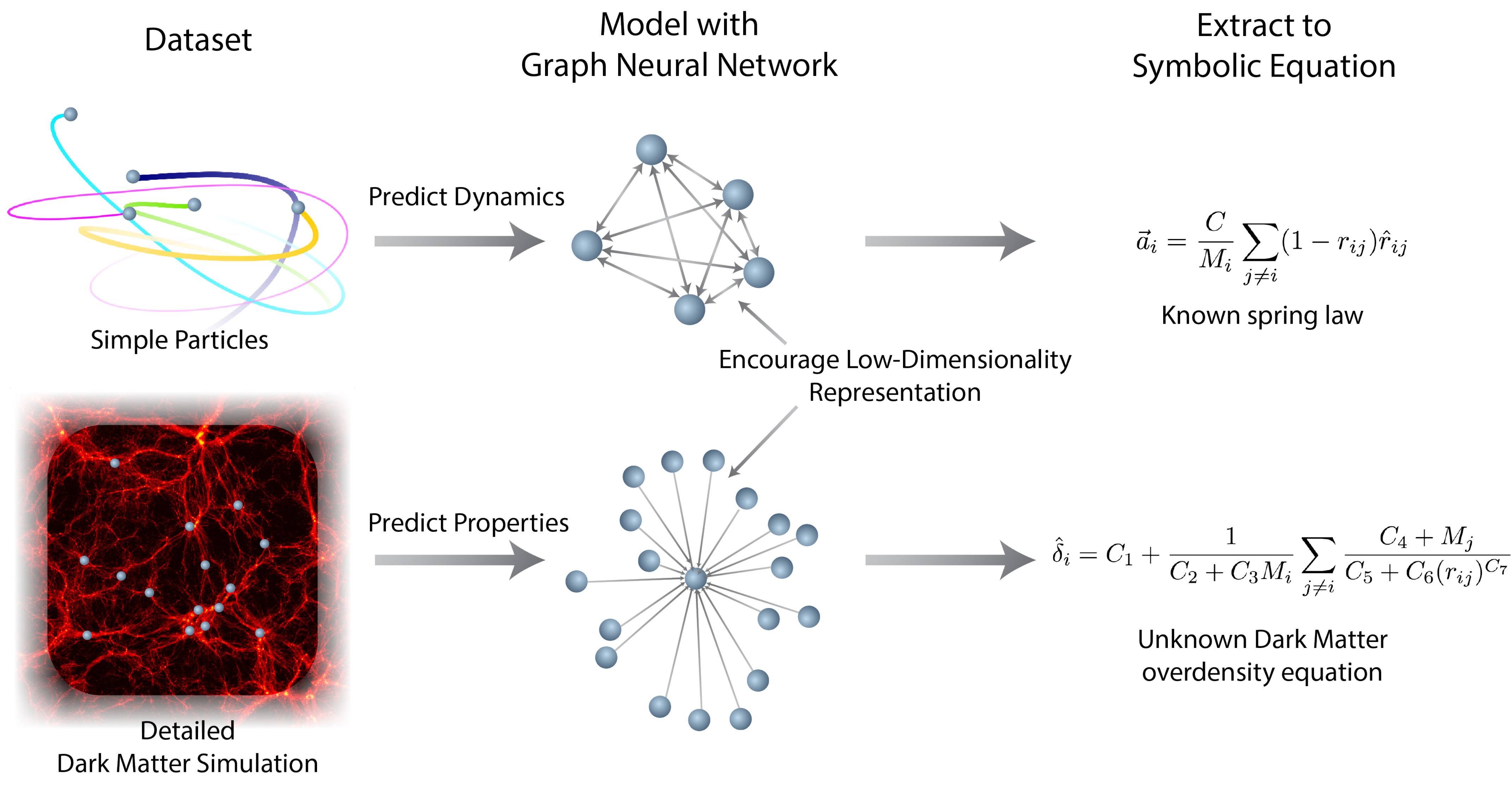
Example [2010.11998]: Automated way to find a set of high level features [EFNs] that include all the information a CNN is implicitly learning from raw calorimeter images



Constrain structure of model

Embed physics symmetries / interactions in the network structure

Symbolic regression [2006.11287]



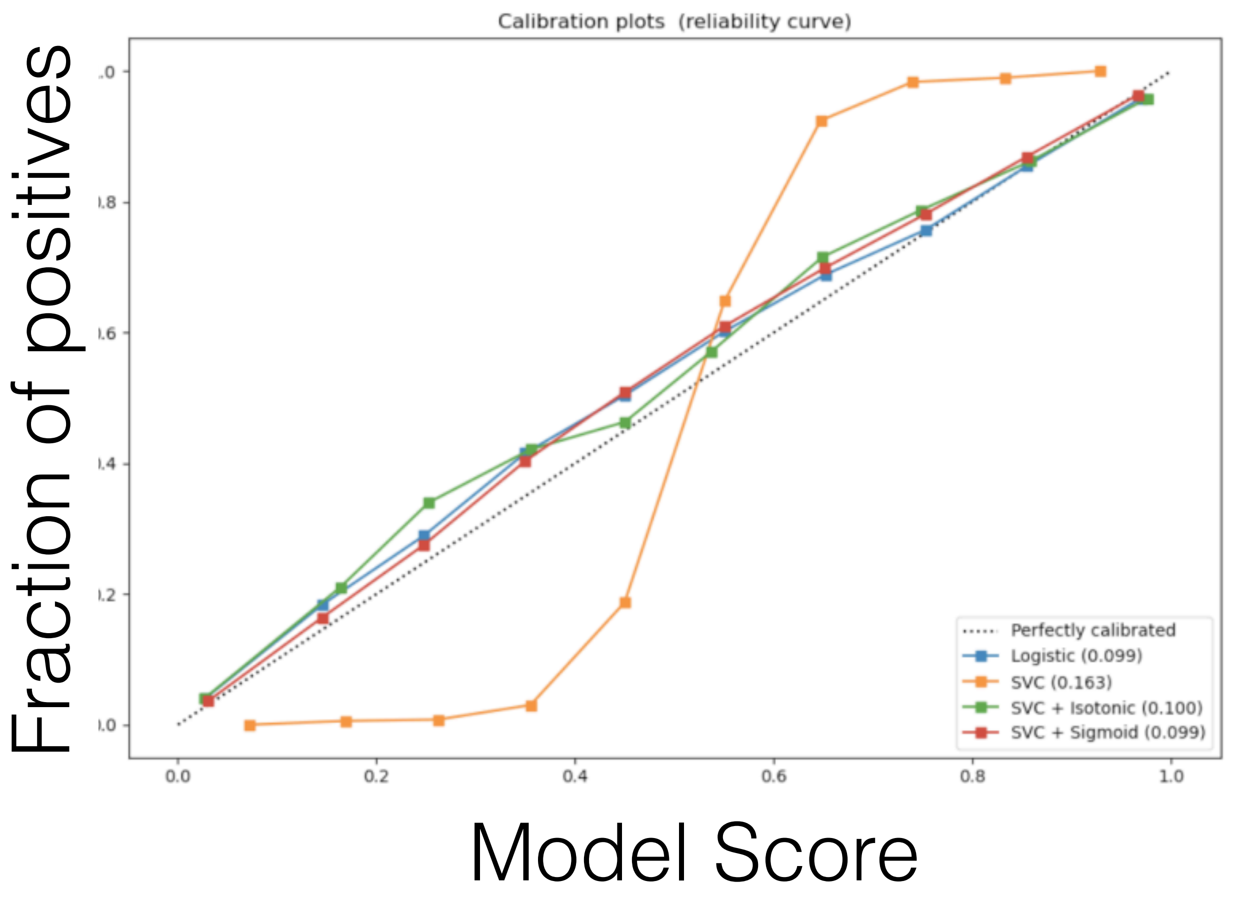
Construct networks that do not violate theoretical constraints (infrared / collinear safety for jet images) [EFPs], or are provably monotonic w.r.t. some feature [2112.00038]

Validation

Power of ML comes from finding subtle non-linear patterns in training data, this also makes it more susceptible to discrepancies between simulation – data.

Sometimes we want to quantify uncertainty of a model itself (eg. Likelihood ratio estimations) - <u>sometimes you don't</u> (eg. when histogramming the output

Calibration of a model: Is the model over/under-confident of its predictions?



Source: https://neptune.ai/blog/brier-score-and-model-calibration

Do generative models covert statistical uncertainties in training sample to systematic uncertainties in generated data? How do you deal with that

Uncertainty Propagation / Mitigation

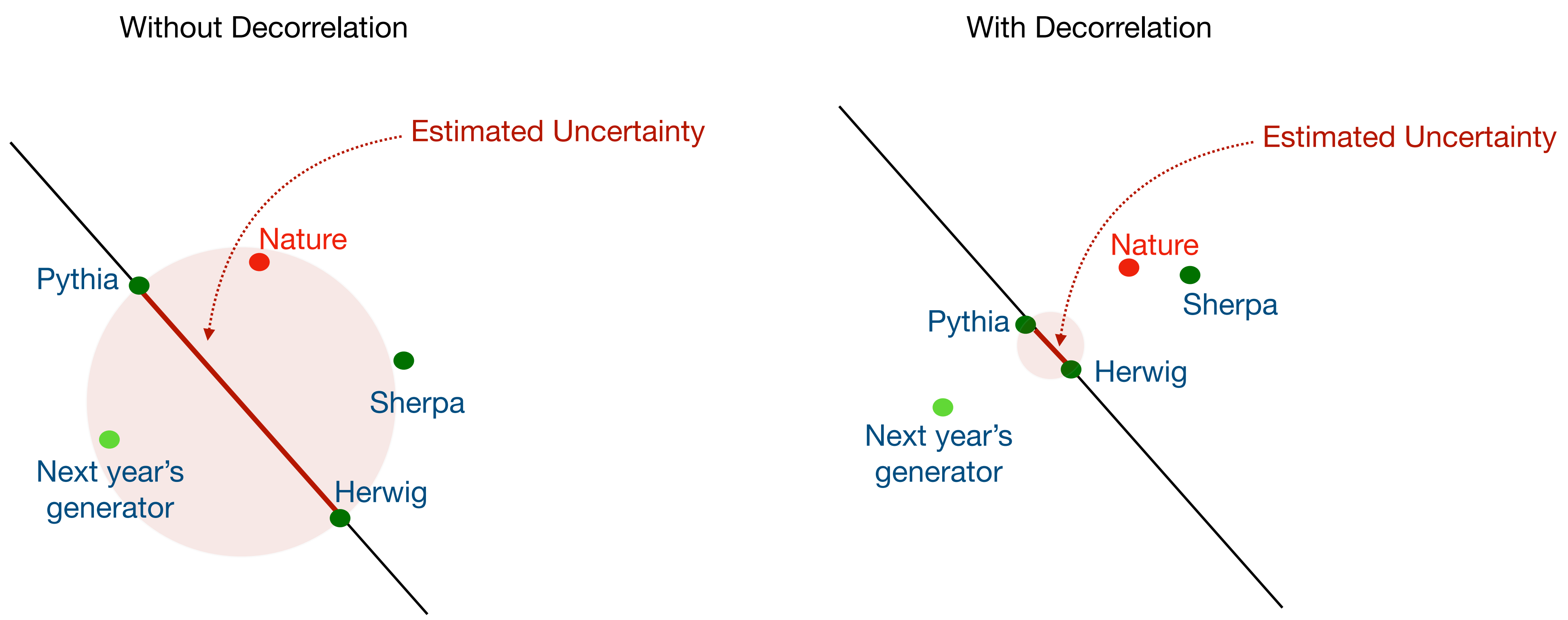
Decorrelation of a model to the source of uncertainty (eg. uncertain detector response like an energy scale) [eg. 1611.01046, 2001.05310 for DNNs, 1810.08387 for BDTs]

Data augmentation and other domain-adaptation techniques

Alternatively: Uncertainty Aware Learning to propagate uncertainties to the final result [2105.08742]

Differentiable program to optimise the final physics goal, including uncertainties [eg. 1806.04743, 2203.05570]

Beware the danger: Sometimes bias mitigation techniques only serve to hide the true uncertainties, rather than reducing them



Reducing uncertainties in other ways

Generative models learning from data could help reduce uncertainties currently coming from simulation mismodelling like for Hadronization [2203.12660]

ML empowers large experiments to report unbinned measurements which allows better re-use of data [2109.13243]

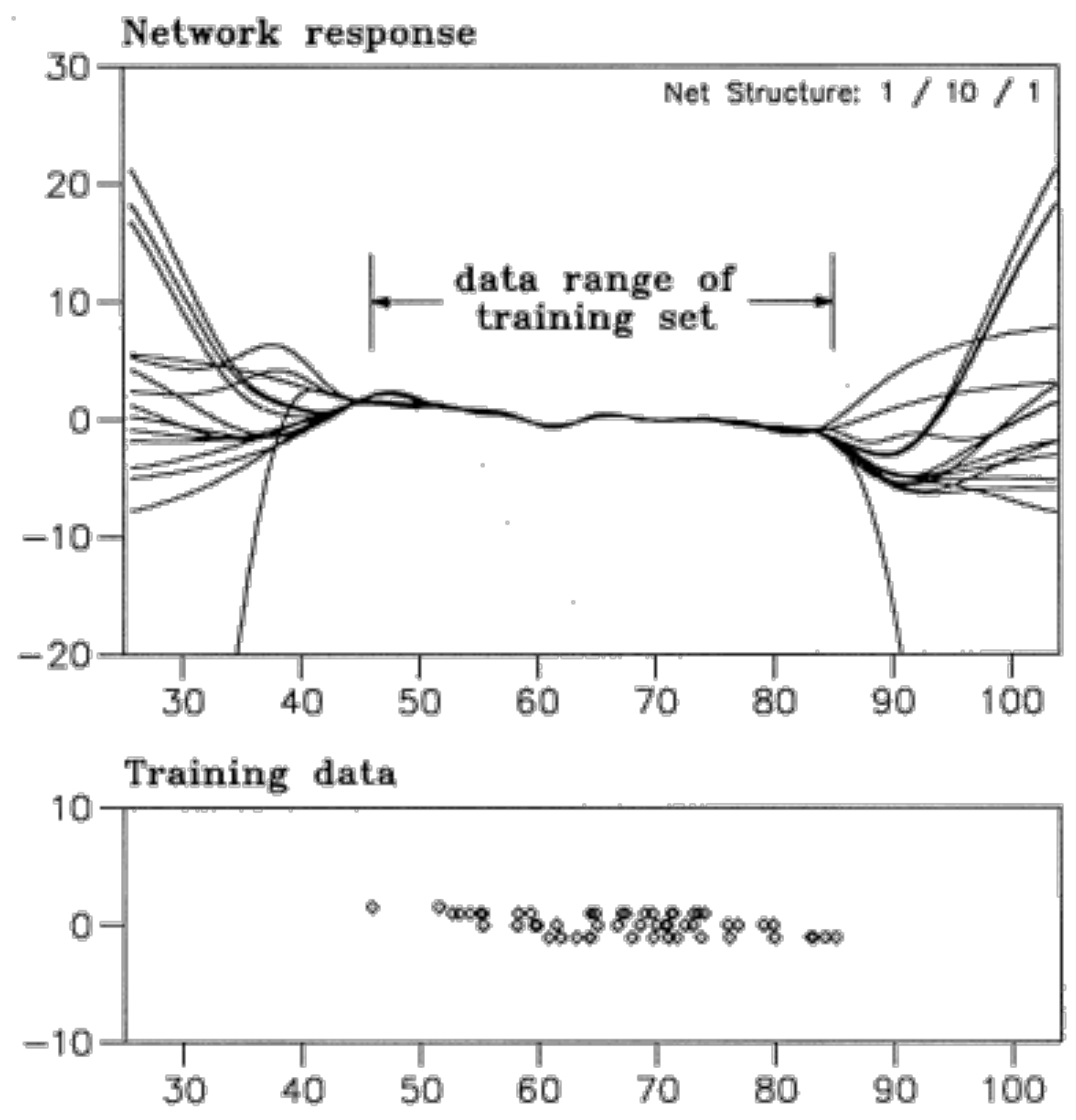
Quantification of model uncertainty

AI solutions growing to quantify the uncertainty of a model prediction

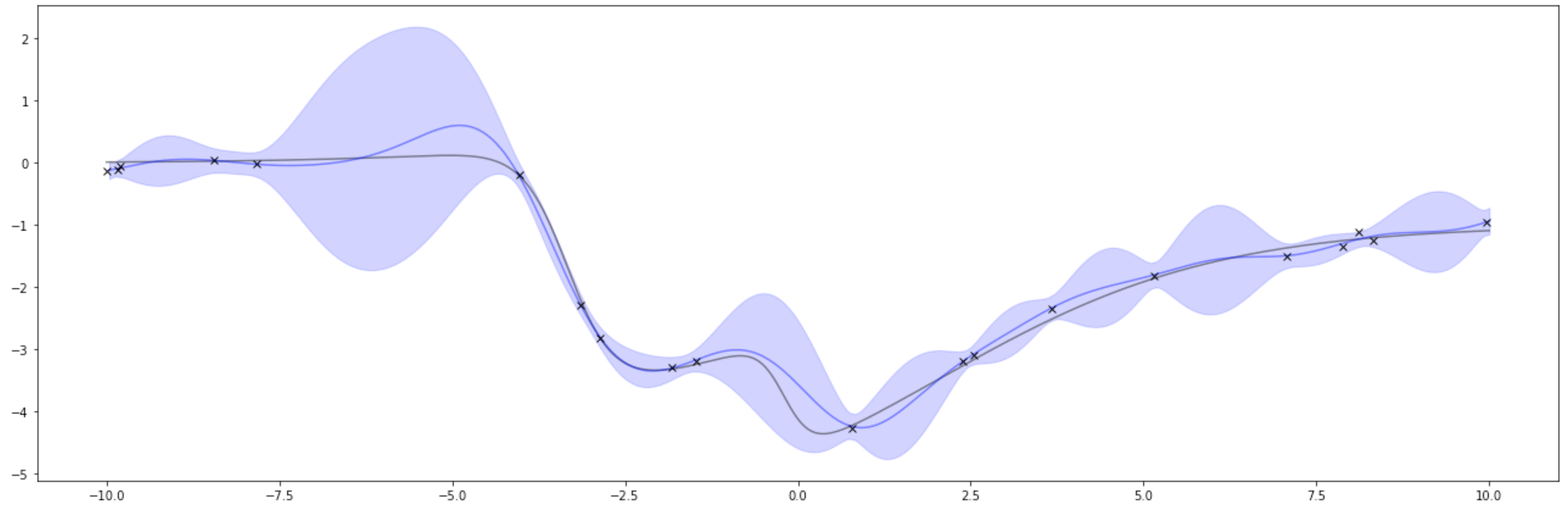
Ensemble of NNs, Bayesian Networks [eg. Generative models with uncertainties: 2110.13632], Gaussian Processes

But what is your uncertainty on that uncertainty? If model is poorly trained, it could give a wrong prediction with incorrect uncertainties.

The physics and AI communities are starting to engage on these topics



NN Ensemble



Gaussian Process

Lack of common definitions

Physicists	AI researchers
Systematic and Statistical Uncertainties	Aleatoric and epistemic uncertainties
Coverage	Calibration
…	…

Recommendations

Physics-AI community can put out benchmark datasets for uncertainty quantification and mitigation

Encourage community to make code public to test reproducibility of ML heavy studies

Funding agencies should endorse challenges to create and compare UQ, UM solutions

Common UQ methods should be incorporated into standard packages like SKLearn for ease of adoption

Outlook

Sometimes a balancing act between performance and interpretability

Problem not unique to physics but we could be among leaders this line of research in AI

Putting out public benchmark datasets will spur creation and comparison of such techniques and increase engagement between physics and AI communities

Harmonising terminology would help interaction between communities on this topic