What information would third-party models need to provide to employ the scoring capability of StarDrop?

The models would need to provide a measure of uncertainty and you would need an understanding of the output in order to choose appropriate scoring thresholds.