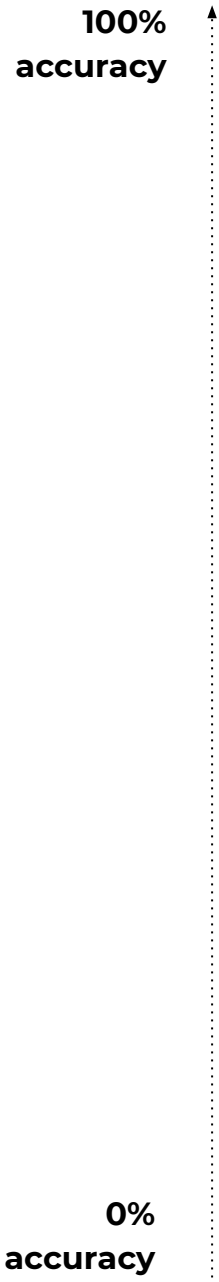


Worksheet: Benchmarking



Discuss and place on the spectrum between 1% and 99%:

Current human benchmark based on how people currently perform the task (optional: expert & layman)

Baseline model based on research and industry standard

Minimum confidence level based on cost of falses. How confident does the model need to be before presenting an answer? At what point is making the wrong prediction more harmful than not making one at all?

Minimum benchmark to provide value to user based on your use case and user research

Targeted benchmark based on the above and cost of falses.

At what point is the prediction correct enough to provide value? How confident does the model need to present an answer? At what point is making the wrong prediction more harmful than not making one at all?