

# QnA data guidelines

---

Starter for some guidelines around building good QnA training data for a bot service.

## Questions

- To help the models the questions in the training data need to be succinct:
  - Not this: "How safe will staking be? Will the "average Joe" be able to stake without having to worry about his/hers OMGs? Will there be certain precautions that will need to be taken in order to prevent being hacked?"
  - ...but this: "How safe will staking be?"
  - ...or this: "Will staking be safe?"
- We probably need to be careful about having overlapping Q&As in different places
  - **TODO:** Can we prioritise one answer over another with metadata?
- If the questions are short we can probably automate generating phrasing variations to help the models
  - You can plug [Active learning](#) into the Azure bot service now too

## Answers

- Should be triaged to make sure they are acceptable (clear, correct, free from the usual troll stuff if they are from the community)
  - **TODO:** how to sign these off? Pull request?
- The QnA bot should provide the **pointer** to the correct response, not the answer. **IMO:** [PickleN](#)