Interview guide

Notes:

- When a practitioner mentions a quality issue, always ask how they currently handle it.

Guide

Introduction

▼ Short introduction of the interviewers.

▼ Description of the objectives of the interview

- **What**: Develop a catalog of quality issues in Machine Learning Software Systems.
  - **What is a quality issue**: Any issue that does not affect the functionality of a system, but only its serving quality. For example, a recommender system whose predictions are accurate but not explainable has quality issues, but not functional issues.
  - **Quality aspects of ML**: robustness, scalability, explainability, model complexity, resource demand, etc.
    - Note: Correctness might be a functional issue, if a failure in the correctness of the ML component's predictions leads to a failure of the whole system (exp: a self-driving car must be correct, or else it will fail to its purpose: driving a car)
  - Furthermore, we are interested in any data or model quality issue you encountered when building a ML system
- **Why**: Guide future work on improving the quality of ML systems

▼ Setting up the interview

1. Ask for permission to record the interview. Explain to the interviewee that it is our intention to release an anonymized version of the interview transcript publicly.
2. Some background information
   a. Current position
   b. Experience (general/specific to ML)
3. For what purpose do you use AI at your company?
   a. Description of the data
      i. type (image, text, tabular, time series, etc.)
      ii. how is data collected

iii. for what problem

b. Which models, algorithms, and frameworks do you usually use?

c. Where are your models deployed (locally hosted, cloud-based, Azure, AWS, GCP, etc.)?

d. Who implements your solutions? Do you have a team of data scientists to develop your models or do you resort to consultants?

▼ If they build and deploy their models

- What are your development practices (e.g. do you use Agile, DevOps)?

Body of the interview

▼ A general and open-ended question to start the interview:

- What are the main quality issues that you have encountered with your data/model or system so far?

▼ Data collection questions

- How do you collect data to train your models?

▼ If they have data collectors (SME):

1. how do you verify the quality of the data collected?
2. what are the problems you encountered the most in the data collection process?
3. How do you prevent them from happening?

▼ If they use data generated internally

1. Where does the data usually come from (is it manually entered or generated by a computer)?
   a. Which data source had the most problems? Why?
   b. How do you fix these issues?
2. How do you merge data coming from different sources
3. What are the problems you encountered the most after merging data from different sources?
4. Does understanding the dataset require expertise?
5. Is adding features to a dataset a long process? Why?

▼ If they use public datasets

1. If yes:
   a. which one

b. did you use the whole dataset to train the model
    i. if no:
        1. why
        2. how did you decide which part of training data to use
c. What are the main problems you have encountered using these datasets?
    i. How did you handle these problems?

▼ If they use external services to obtain data (e.g. weather API)

1. Which external service did you use
    a. Which issues have you faced?
    b. How have you handled the issues?

▼ If they use the predictions of a model in their training data (i.e., cascading models: the predictions of one model are used by another model to do predictions)

1. Have you ever encountered problems with this data source?
2. How have you mitigated this risk?

- Have you encountered any other data quality issues caused by data collection?

▼ Data preparation questions (data cleaning + data transformation)

1. Which tools/frameworks do you use when preparing your data?
2. How long do you spend collecting and preparing your data vs developing the model?

    ▼ If it is close to 80/20

    1. Why does it take so long?

    ▼ If it is more balanced

    1. Some studies show that, in the industry, preparing data takes 80% of the time while developing models takes 20%. Why is it different for you?

3. What are the pain points you repetitively encounter when preparing data for ML?
    a. Why and how do these problems happen?
4. Is there any other data quality issue we missed that you consider relevant?

▼ Model evaluation questions

1. How do you evaluate the quality of models?
    a. Have you used existing qualified models to evaluate your model?
2. How long do you spend on tuning and debugging your ML model vs designing it?

    ▼ If it is close to 80/20

1. Why does it take so long?

▼ If it is more balanced

1. Some studies show that 80% of the time spent on the model is on tuning and debugging and only 20% on building the architecture. Why is it different for you?

3. Did you use any user acceptance measurement for trained models?

▼ If yes

What kind of measurements?

▼ If no

Why? Do you consider user acceptance a negligible issue?

4. Are your models used in scenarios involving different groups of people?
   a. Was there a situation where it performed poorly on some group of people?

▼ If yes

1. Why?
2. How did you fix it?

5. Have you encountered any other quality issues during the evaluation of your models?

▼ Model deployment questions

1. Where are your models deployed (on which platform/hardware)?
2. How are your models deployed (manually vs. automatically)?
3. Did you ever deploy a model that performed well locally but poorly once deployed?
   a. What was the problem?
   b. How did you handle the problem and have you taken any measures to prevent it from happening in the future?
4. Have you encountered any other quality issues during the deployment of your models?

▼ Model maintenance questions

1. Do you monitor your ML models once deployed?

▼ If yes

- How do you monitor them?
- Which metric do you monitor?
- What happens when a model is stale (do you have a platform to automatically re-train and re-deploy the models)?

2. Have some of your models gone stale after some time?

- ▼ If yes
    1. What was the root cause of your models' staleness?
    2. How do you handle model staleness?
3. Have you had other issues regarding the maintenance of your ML models?

- ▼ Quality measures of ML models questions
    1. Have you ever faced problems related to the scalability of trained models?
        a. Which kind of scalability issues?
        b. How does your team currently handle these issues?
    2. Is robustness a significant quality issue when building ML models?
        a. How do/did you evaluate robustness?
        b. How does your team currently handle robustness issues?
    3. Have you ever investigated the explainability of your trained models? (trying to explain the final decision of the model)
        a. How? Which measurement did you use?
    4. Is there any other quality issue in ML systems that you have experienced and that we did not inquire about in this interview?
- ▼ Other questions
    1. Do you have AI projects that took longer than expected?
        - ▼ If yes

          What were the main reasons?
    2. Did you have any AI projects that have been aborted because your team could not get good results with the given dataset?
        - ▼ If yes
            1. what was the problem?

Conclusion

- ▼ One closing question

  In your opinion, what is the most pressing quality issue researchers should try to solve?

  Do you have any other comments about the quality of ML systems?