Semi-Structured Questionnaire for Interviews

i) Best Practices and Protocols

- What do you consider as a best practice?
- Which process, guidelines, or pipelines you follow when employing ML in SE?
- What do you consider are the "must" practices or protocols you use or implement in your research/papers?
- Which are the most recurrent challenges when using ML for SE in any of the ML workflow stages?
- Which practices do you use from each ML pipeline stage?
- How did you employ the components of learning to guide your ML model design? [Abu-Mustafa, 2012]
- Do you take into consideration any of the learning principles (*e.g.,* data snooping, Occam's razor, or sampling bias)? [Abu-Mustafa, 2012]
- Which quality attributes are important for SE systems that use ML? ⇒ **ANSW1**
- Do you have any particular difficulties/challenges when building an ML-enabled system to ensure **ANSW1** attributes?

ii) Education

- What educational resources have you employed to perform ML4SE? (*e.g.,* tools, languages, books, etc..)
- How do you learn ML4SE?
- How do you educate your students to enable ML4SE?
- How do you promote **ANSW1** when teaching about ML4SE?

iii) Reviewer's Perspective

- What issues have you observed when reviewing papers?
- What ML4SE areas are not being covered in conference reviews?
- As a reviewer, have you seen that the **ANSW1** attributes are being addressed?

iv) ML Workflow Stages (*Amershi et al.*)

- Model Requirements
  - What was the rationale of selecting or proposing an ML model you have utilized in your papers?

- Data Collection
  - Do you collect your own data, use previous datasets, or both? How do you select data/datasets?
  - Which data collection strategies or protocols you employ in your studies?
  - How do you promote **ANSW1** when collecting data for ML-enabled systems?
- Data Cleaning
  - Which pipelines do you employ to address data exploration?
  - How do you perform data cleaning?
  - How do you handle exploratory analysis in your studies?
  - How important are exploratory analyses for data cleaning?
  - How do you promote **ANSW1** when cleaning data in ML-enabled systems?
- Data Labeling (if applicable)
  - Which data labeling protocols do you use for your supervised tasks?
  - How do you promote **ANSW1** when labeling data for ML-enabled systems?
- Feature Engineering
  - Which feature engineering methods you have employed?
  - How do you promote **ANSW1** when selecting and building features in ML-enabled systems?
- Model Training
  - What frameworks have you employed for model training?
  - How do you parameterize your models?
  - How do you promote **ANSW1** when training models in ML-enabled systems?
- Model Evaluation
  - How do you evaluate or validate your ML models?
  - Do you employ any measurements to control for bias?
  - Do you use any interpretability technique to guide your evaluation?
  - How are measurements aligned to study RQs, goals, or business objectives?
  - How do you promote **ANSW1** when evaluating ML-enabled systems?
- Model Deployment

  - How do you do promote **ANSW1** when deploying ML-enabled systems?
- Model Monitoring
  - Do you employ any measurements to handle concept drift?
  - How do you monitor models during operation (*e.g.,* user studies, industry projects, production)?
  - How do you promote **ANSW1** when monitoring ML-enabled systems?

v) Potential Follow-up Questions