A Checklist of Quality Concerns for Architecting ML-Intensive Systems

by Alessio Bucaioni, Rick Kazman, and Patrizio Pelliccione

System Concerns

- [ ] **SU1** Do you have data visualization techniques in place? (ML Spec: H)
- [ ] **SU2** Have you considered visualization techniques to highlight relationships between data and computing tasks? (ML Spec: H)
- [ ] **SDQ1** Do you have strategies for data preparation and for making statistics on data? (ML Spec: H)
- [ ] **SDQ2** Is your dataset clean, of good quality, and free from potential bias? (ML Spec: H)
- [ ] **SDQ3** Are you concerned about data cleaning in ML processing? (ML Spec: H)
- [ ] **SDQ4** Do you have a well-sized dataset for training the ML component? (ML Spec: H)
- [ ] **SDQ5** Do you engineer your ML-based system to adapt to input data changes, also known as concept drift? (ML Spec: H)
- [ ] **SC1** Do you have proper techniques for ensuring system correctness? (ML Spec: H)
- [ ] **SMV1** Are you performing the validation of the model to predict behavior on new data? (ML Spec: H)
- [ ] **SMV2** Are you combining model validation with data validation to better detect corrupted training? (ML Spec: H)
- [ ] **SMD1** Are you building a component-based distributed system where parts may need to be upgraded? (ML Spec: M)
- [ ] **SMD2** Are high cohesion and low coupling important? (ML Spec: L)
- [ ] **SMD3** If you are interested in maintainability and modifiability, did you consider using a microservice architecture? (ML Spec: L)
- [ ] **SMD4** Can you decompose your system into discrete services? (ML Spec: L)
- [ ] **SMN1** Can you explicitly model the intrinsic uncertainty of ML components and assess its impacts? (ML Spec: H)

A Checklist of Quality Concerns for Architecting ML-Intensive Systems

by Alessio Bucaioni, Rick Kazman, and Patrizio Pelliccione

- [ ] **SMN2** Do you have mechanisms like monitoring and a-posteriori analysis for time predictability? (ML Spec: H)
- [ ] **SMN3** Do you have tests that monitor changes in input distributions? (ML Spec: H)
- [ ] **SDE1** Can you use continuous integration techniques for system development? (ML Spec: M)
- [ ] **SDE2** Do you manage the IT infrastructure needed to build and deploy your ML system? (ML Spec: O)
- [ ] **SDE3** Are you including Blue/Green or Canary testing in your standard MLOps pipelines? (ML Spec: O)
- [ ] **SA1** Did you consider failure recovery strategies to avoid propagation of failures? (ML Spec: L)
- [ ] **SA2** Do you have the required domain knowledge to take availability decisions? (ML Spec: M)
- [ ] **SA3** Can you cleanly split business logic from ML components using layered/tiered architecture? (ML Spec: M)
- [ ] **SR1** Do you have complete information on the uncertainty of the ML components at design time? (ML Spec: H)
- [ ] **SS1** Do you have techniques for reaching safe states quickly when needed? (ML Spec: H)
- [ ] **SS2** Have you included an evaluation process for architectural safety design choices? (ML Spec: L)
- [ ] **SS3** Do you use strict and certified coding standards when developing safety-critical ML components? (ML Spec: L)
- [ ] **SS4** Are you having your system safety-certified by an external body? (ML Spec: L)
- [ ] **SS5** Are you explicitly designing and developing your ML system to defend against cyber-attacks? (ML Spec: H)

- [ ] **SS6** Do you have a way of systematically ensuring safety and fairness in your system? (ML Spec: H)
- [ ] **SP1** If you need to reduce data loss as well as improving privacy, one way is to use federated learning. (ML Spec: H)

A Checklist of Quality Concerns for Architecting ML-Intensive Systems

by Alessio Bucaioni, Rick Kazman, and Patrizio Pelliccione

Process Concerns

- [ ] **PD1** Do you have proper documentation or a plan to document your ML system? (ML Spec: M)
- [ ] **PT1** Do you have heterogeneous teams mixing ML developers, data engineers, and architects? (ML Spec: D)
- [ ] **PT2** Do you have a test-driven development strategy for your QA and testing process? (ML Spec: M)
- [ ] **PSP1** Do you separate the branches for training pipelines from model training? (ML Spec: H)
- [ ] **SML1** Do you have expertise to customize and reuse models? (ML Spec: H)
- [ ] **SML2** Do you manage and version ML models? (ML Spec: H)
- [ ] **SMI3** Have you defined ML infrastructure and deployment processes? (ML Spec: H)
- [ ] **SML4** Are you taking care of testing the quality and performance of the model? (ML Spec: H)