**1- If the data quality is good, how is it possible to get a bad/wrong output from an ML algorithm?**

The ML algorithm could be a bad algorithm or there could be too many irrelevant features. Bad/wrong output can also occur if one over-fits the training data. Having good data quality is important, but it’s also important on utilizing it correctly.

**2- How is data integrity/quality maintained in an online ML system?**

Some ways to preserve data integrity is to validate both the input and the data. In addition, it is important to keep the data unique and remove duplicates. Having duplicate data can cause the data to be skewed. It is also important to keep the data up-to-date and to back it up occasionally. You can also check the consistency of data by comparing them to other data sets. Another way is to also have a continuous check of descriptive statistics to determine if any pre-processing needs to be added.

**3- Explain how Google's automated video captioning system is categorized according to different ML types?**

Google’s automated video captioning system is a deep neural network(DNN) model that was trained with a lot of labeled data. Google initially would use weakly labeled data to train the neural network to detect all ambient sounds such as laughter, music, applause, etc. This makes this system a semi-supervised, model-based learning system. The labeled data is weak, which means there is some labeled data and unlabeled data which makes the system semi-supervised. Separating all of the data into different kinds of models for different sounds such as laughter, music, and applause makes the system model-based.

**4- What are the key elements of a successful ML-based application?**

The key elements are representation, evaluation, and optimization. Representation is how to represent knowledge, which includes many models and rules. Evaluation is a way to evaluate hypotheses, which includes accuracy, cost, error margins, and so on. Optimization is the search process through represented models to obtain better evaluations/programs. Data integrity is also very important, to ensure data quality, a suitable algorithm, and use of cross-validation to avoid over-fitting.