

## Live Data Dangers

TOTAL POINTS 2

1. Which of the following statements are true?

1 / 1 point

- ☐ The distribution of your live data is always consistent.
- ☒ If you train your model on one data pipeline, that does not guarantee the model will work for all data pipelines.
- ☐ Once you've trained a model, it can be used indefinitely.
- ☐ If you properly train your model, you can entirely eliminate the issue of overfitting.

✓ Correct

Correct! It can be useful to try to reuse as much code as possible in both the training data pipeline and the operational (live) data pipeline, but you will need to test to make sure your live data pipeline integrates well with your model.

2. Which of the following are potential issues that can arise with live data? (Select all that apply.)

1 / 1 point

- ☒ The sheer volume of live data might be much larger than first expected.

✓ Correct

Correct! Live data increases over time, and then keeps increasing. Live data may come in at a much higher speed compared to what you traditionally have with training data.

- ☒ There might be different distributions in live data than there was in your training data.

✓ Correct

Correct! Your training data is just a sample of the data out there so it is possible the live data has a different distribution (which can change over time).

- ☒ The owner of a dataset might unexpectedly make a change to the data.

✓ Correct

Correct! People have a tendency to change things without thinking about the larger impact of their changes.

- ☐ There are no issues with live data if you've properly trained your model.