1. This example is adapted from a real production application, but with details disguised to protect confidentiality. **1 / 1 point**

You are a famous researcher in the City of Peacetopia. The people of Peacetopia have a common characteristic: they are afraid of birds. To save them, you have **to build an algorithm that will detect any bird flying over Peacetopia** and alert the population.

The City Council gives you a dataset of 10,000,000 images of the sky above Peacetopia, taken from the city's security cameras. They are labeled:

- y = 0: There is no bird on the image
- y = 1: There is a bird on the image

Your goal is to build an algorithm able to classify new images taken by security cameras from Peacetopia.

There are a lot of decisions to make:

- What is the evaluation metric?
- How do you structure your data into train/dev/test sets?

Metric of success

The City Council tells you the following that they want an algorithm that

1. Has high accuracy.
2. Runs quickly and takes only a short time to classify a new image.
3. Can fit in a small amount of memory, so that it can run in a small processor that the city will attach to many different security cameras.

You meet with them and ask for just one evaluation metric. True/False?

- False
- True:

Expand

Correct
Yes. The goal is to have one metric that focuses the development effort and increases iteration velocity.

2. After further discussions, the city narrows down its criteria to: **1 / 1 point**

- "We **need** an algorithm that can let us know a bird is flying over Peacetopia as accurately as possible."
- "We *want* the trained model to take no more than 10 sec to classify a new image."
- "We *want* the model to fit in 10MB of memory."

If you had the three following models, which one would you choose?

Test Accuracy	Runtime	Memory size
97%	3 sec	2MB

Test Accuracy	Runtime	Memory size
98%	9 sec	9MB

Test Accuracy	Runtime	Memory size
97%	1 sec	3MB

Test Accuracy	Runtime	Memory size
99%	13 sec	9MB

Expand

Correct
Correct! This model has the highest test accuracy, the prominent criteria you are looking for, compared with other models, and also has a runtime <10 seconds and memory size < 10MB.

3. Which of the following best answers why it is important to identify optimizing and satisficing metrics? **1 / 1 point**

- It isn't. All metrics must be met for the model to be acceptable.
- Knowing the metrics provides input for efficient project planning.
- Identifying the metric types sets thresholds for satisficing metrics. This provides explicit evaluation criteria.
- Identifying the optimizing metric informs the team which models they should try first.

Expand

Correct
Yes. Thresholds are essential for evaluation of key use case constraints.

4. You propose a 95/2.5%/2.5% for train/dev/test splits to the City Council. They ask for your reasoning. Which of the following best justifies your proposal? **1 / 1 point**

- The emphasis on the training set provides the most accurate model, supporting the memory and processing satisficing metrics.
- The most important goal is achieving the highest accuracy, and that can be done by allocating the maximum amount of data to the training set.
- With a dataset comprising 10M individual samples, 2.5% represents 250k samples, which should be more than enough for dev and testing to evaluate bias and variance.
- The emphasis on the training set will allow us to iterate faster.

Expand

Correct
Yes. The purpose of dev and test sets is fulfilled even with smaller percentages of the data.

5. Now that you've set up your train/dev/test sets, the City Council comes across another 1,000,000 images from social media and offers them to you. These images are different from the distribution of images the City Council had originally given you, but you think it could help your algorithm. You should add the citizens' data to the training set. True/False? **1 / 1 point**

- False
- True

Expand

Correct
Yes. This will cause the training and dev/test set distributions to become different, however as long as dev/test distributions are the same you are aiming at the same target.

6. One member of the City Council knows a little about machine learning, and thinks you should add the 1,000,000 citizens' data images to the test set. You object because: **1 / 1 point**

- [x] The test set no longer reflects the distribution of data (security cameras) you most care about.

Correct

- [ ] The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- [ ] A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.
- [x] This would cause the dev and test set distributions to become different. This is a bad idea because you're not aiming where you want to hit.

Correct

Expand

Correct
Great, you got all the right answers.

7. You train a system, and its errors are as follows (error = 100%-Accuracy): **1 / 1 point**

Training set error	4.0%
Dev set error	4.5%

This suggests that one good avenue for improving performance is to train a bigger network so as to drive down the 4.0% training error. Do you agree?

- Yes, because this shows your bias is higher than your variance.
- No, because this shows your variance is higher than your bias.
- Yes, because having a 4.0% training error shows you have a high bias.
- No, because there is insufficient information to tell.

Expand

Correct

8. You want to define what human-level performance is to the city council. Which of the following is the best answer? **1 / 1 point**

- The average of all the numbers above (0.66%).
- The average of regular citizens of Peacetopia (1.2%).
- The average performance of all their ornithologists (0.5%).
- The performance of their best ornithologist (0.3%).

Expand

Correct
Yes. The best human performance is closest to Bayes' error.

9. Which of the below shows the optimal order of accuracy from worst to best? **0 / 1 point**

- The learning algorithm's performance -> human-level performance -> Bayes error.

- Human-level performance -> the learning algorithm's performance -> Bayes error.
- Human-level performance -> Bayes error -> the learning algorithm's performance.
- The learning algorithm's performance -> Bayes error -> human-level performance.

Expand

Incorrect
No. in an optimal scenario, your algorithm's performance would be better than HLP but it can never be better than BE.

10. Which of the following best expresses how to evaluate the next steps in your project when your results for human-level performance, train, and dev set error are 0.1%, 2.0%, and 2.1% respectively? **1 / 1 point**

- Keep tuning until the train set accuracy is equal to human-level performance because it is the optimizing metric.
- Port the code to the target devices to evaluate if your model meets or exceeds the satisficing metrics.
- Based on differences between the three levels of performance, prioritize actions to decrease bias and iterate.
- Evaluate the test set to determine the magnitude of the variance.

Expand

Correct
Yes. Always choose the area with the biggest opportunity for improvement.

11. You also evaluate your model on the test set, and find the following: **1 / 1 point**

Human-level performance	0.1%
Training set error	2.0%
Dev set error	2.1%
Test set error	7.0%

What does this mean? (Check the two best options.)

- [ ] You have underfitted to the dev set.
- [x] You have overfit to the dev set.

Correct

- [ ] You should get a bigger test set.
- [x] You should try to get a bigger dev set.

Correct

Expand

Correct
Great, you got all the right answers.

12. After working on this project for a year, you finally achieve: **1 / 1 point**

Human-level performance	0.10%
Training set error	0.05%
Dev set error	0.05%

What can you conclude? (Check all that apply.)

- [x] If the test set is big enough for the 0.05% error estimate to be accurate, this implies Bayes error is ≤ 0.05

Correct

- [x] It is now harder to measure avoidable bias, thus progress will be slower going forward.

Correct

- [ ] This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.
- [ ] With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%

Expand

Correct
Great, you got all the right answers.

13. Your system is now very accurate but has a higher false negative rate than the City Council of Peacetopia would like. What is your best next step? **0 / 1 point**

- ( ) Pick false negative rate as the new metric, and use this new metric to drive all further development.
- (•) Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.
- ( ) Expand your model size to account for more corner cases.
- ( ) Reset your "target" (metric) for the team and tune to it.

Expand

Incorrect
No. You must maintain accuracy and include false negatives.

14. You've handily beaten your competitor, and your system is now deployed in Peacetopia and is protecting the citizens from birds! But over the last few months, a new species of bird has been slowly migrating into the area, so the performance of your system slowly degrades because your data is being tested on a new type of data. **1 / 1 point**

You have only 1,000 images of the new species of bird. The city expects a better system from you within the next 3 months. Which of these should you do first?

- (●) Use the data you have to define a new evaluation metric (using a new dev/test set) taking into account the new species, and use that to drive further progress for your team.
- ( ) Put the 1,000 images into the training set so as to try to do better on these birds.
- ( ) Try data augmentation/data synthesis to get more images of the new type of bird.
- ( ) Add the 1,000 images into your dataset and reshuffle into a new train/dev/test split.

Expand

✓ Correct

15. The City Council thinks that having more Cats in the city would help scare off birds. They are so happy with your work on the Bird detector that they also hire you to build a Cat detector. (Wow Cat detectors are just incredibly useful, aren't they?) Because of years of working on Cat detectors, you have such a huge dataset of 100,000,000 cat images that training on this data takes about two weeks. Which of the statements do you agree with? (Check all that agree.)

1 / 1 point

- [x] If 100,000,000 examples is enough to build a good enough Cat detector, you might be better off training with just 10,000,000 examples to gain a ≈10x improvement in how quickly you can run experiments, even if each model performs a bit worse because it's trained on less data.

  ✓ Correct

- [x] Buying faster computers could speed up your teams' iteration speed and thus your team's productivity.

  ✓ Correct

- [ ] Having built a good Bird detector, you should be able to take the same model and hyperparameters and just apply it to the Cat dataset, so there is no need to iterate.
- [x] Needing two weeks to train will limit the speed at which you can iterate.

  ✓ Correct

Expand

✓ Correct

Correct

Great, you got all the right answers.