In the City of Peacetopia, you are a celebrated researcher tasked with a critical mission. The entire population shares a phobia of birds, and your job is to create an algorithm that identifies any birds flying over the city, alerting residents promptly.

You are given 10,000,000 images of the sky from the city's security cameras, labeled as:

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

**Your mission is to develop an algorithm to classify new images from security cameras accurately**. This involves crucial decisions on the evaluation metric and how to organize your data into train/dev/test sets.

The City Council desires an algorithm that:

1. Achieves high accuracy.
2. Quickly classifies new images.
3. Uses minimal memory to be compatible with small processors in various security cameras.

**True or False:** You acknowledge that having multiple evaluation metrics may complicate the decision-making process and slow down iteration speed.

- (●) True
- ( ) False

**Correct**
While it's important to consider various performance aspects, focusing on a single evaluation metric simplifies decisions and accelerates the development cycle, enabling faster iterations and optimizations.

**2.** The city revises its criteria to:

- "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 seconds to classify a new image."
- "We want the model to fit in 10MB of memory."

Given models with different accuracies, runtimes, and memory sizes, how would you choose one?

- ( ) Create one metric by combining the three metrics and choose the best performing model.
- ( ) Accuracy is an optimizing metric, therefore the most accurate model is the best choice.
- ( ) Take the model with the smallest runtime because that will provide the most overhead to increase accuracy.
- (●) Find the subset of models that meet the runtime and memory criteria. Then, choose the highest accuracy.

**Correct**
Once you meet the runtime and memory thresholds, accuracy should be maximized.

3. Based on the context of a city's data analysis project, **which of the following statements is true regarding the metrics used?**

- ○ Accuracy is a satisficing metric; running time and memory size are an optimizing metric.
- ○ Accuracy, running time, and memory size are all satisficing metrics because you have to do sufficiently well on all three for your system to be acceptable.
- ◉ Accuracy is an optimizing metric; running time and memory size are satisficing metrics.
- ○ Accuracy, running time, and memory size are all optimizing metrics because you want to do well on all three.

**Correct**
Accuracy is a metric we aim to maximize, while running time and memory size have acceptable thresholds.

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**, given that the total data set contains 10,000,000 data points?

- ◉ With a dataset comprising 10,000,000 individual samples, 2.5% represents 250,000 samples, which should be more than enough for dev and testing to evaluate bias and variance.
- ○ 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.
- ○ The emphasis on the training set will allow us to iterate faster.
- ○ The emphasis on the training set provides the most accurate model, supporting the memory and processing efficiency.

**Correct**
The purpose of dev and test sets is fulfilled even with smaller percentages of the data in large datasets, allowing for effective evaluation of model performance.

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 have a different distribution from the images the City Council originally provided, but you think they could help your algorithm.

**Which of the following is the best use of that additional data?**

- ○ Split it among train/dev/test equally.
- ○ Do not use the data. It will change the distribution of any set it is added to.
- ◉ Add it to the training set.
- ○ Add it to the dev set to evaluate how well the model generalizes across a broader set.

**Correct**
It is not a problem to have different training and dev distributions. Different dev and test distributions would be an issue.

6. One member of the City Council wants to add 1,000,000 citizen data images evenly to the training, development (dev), and test sets. Your original data is from security cameras, and you object because:

- ◯ The additional data would significantly slow down training time.
- ◯ The 1,000,000 citizen data images do not have a consistent input-output relationship as the security camera data.
- ◯ The training set will not be as accurate because of the different distributions.
- ◉ If we add the images to the test set, then it won't reflect the distribution of data (security cameras) expected in production.

**Correct**
The test set must accurately represent the real-world data distribution to properly evaluate model performance. Adding citizen data images to the test set would skew this distribution.

7. Human performance for identifying birds is < 1%, training set error is 5.2%, and dev set error is 7.3%.

**Which of the options below is the best next step?**

- ◉ Train a bigger network to reduce the 5.2% training error.
- ◯ Get more data or apply regularization to reduce variance.
- ◯ Try an ensemble model to reduce bias and variance.
- ◯ Validate the human data set with a sample of your data to ensure the images are of sufficient quality.

**Correct**
Avoidable bias is 4.2%, which is larger than the 2.1% variance, so reducing bias is the priority.

8. You want to define "human-level performance" for a bird species identification project to present to the city council. Which of the following is the best way to define it?

- ◯ The average performance of regular citizens of Peacetopia (1.2%).
- ◯ The average of all recorded error rates (0.66%, ornithologists and citizens).
- ◯ The average performance of all the city's ornithologists (0.5%).
- ◉ The performance of the city's best ornithologist (0.3% error rate).

**Correct**
The best human performance, represented by the lowest error rate, is the closest practical estimate of Bayes' error.

9. Which of the below shows the optimal order of accuracy from worst to best?

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

**Incorrect**
In an optimal scenario, the algorithm's performance can be better than human-level performance, but it cannot be better than Bayes error.

10. Which of the following best describes the **most effective next step in your project**, given the following performance metrics?

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

(●) Prioritize actions to decrease bias by increasing model complexity, as the training error significantly exceeds human-level performance.

( ) Evaluate the test set to determine the variance.

( ) Deploy the model to target devices to evaluate against satisficing metrics.

( ) Continue tuning until the training set error matches human-level performance, focusing solely on the optimizing metric.

**Correct**
Yes, addressing the largest performance gap (between human-level and training error) is the most efficient strategy.

11. You've now also run your model on the test set and find that the error rate is 7.0% compared to a 2.1% error rate for the dev set. What should you do? (Choose all that apply)

[ ] Try decreasing regularization for better generalization with the dev set.

[ ] Try increasing regularization to reduce overfitting to the dev set.

[ ] Get a bigger test set to increase its accuracy.

[x] Increase the size of the dev set.

**Correct**
A larger dev set can help provide a more accurate estimate of model performance and reduce overfitting.

12. After working on this project for a year, you finally achieve: Human-level performance, 0.10%, Training set error, 0.05%, Dev set error, 0.05%. Which of the following are likely? (Check all that apply.)

[ ] There is still avoidable bias.

[ ] This result is not possible since it should not be possible to surpass human-level performance.

[x] The model has recognized complex, emergent features that humans may not readily perceive. (Chess and Go, for example).

**Correct**
In domains like Chess and Go, AI models have demonstrated the ability to discover and utilize strategies beyond typical human comprehension.

[x] Pushing to even higher accuracy will be slow because you will not be able to easily identify sources of bias.

**Correct**
Exceeding human performance means you are close to Bayes error, making bias identification difficult.

13. It turns out Peacetopia has hired one of your competitors to build a system as well. You and your competitor both deliver systems with about the same running time and memory size. However, your system has higher accuracy!

    Still, when Peacetopia tries out both systems, they conclude they like your competitor's system better because, even though you have higher overall accuracy, you have more false negatives (failing to raise an alarm when a bird is in the air).

    What should you do?

    - ○ Ask your team to take into account both accuracy and false negative rate during development.
    - ◉ Brainstorm with your team to refine the optimizing metric to include false negatives as they further develop the model.
    - ○ Pick false negative rate as the new metric, and use this new metric to drive all further development.
    - ○ Apply regularization to minimize the false negative rate.

    **Correct**
    The target has shifted so an updated metric is required.

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 model is being tested on a new type of data.

    There are only 1,000 images of the new species. The city expects a better system from you within the next 3 months.

    **Which of these should you do first?**

    - ○ Put the 1,000 images into the dev set to evaluate the bias and re-tune.
    - ○ Add hidden layers to further refine feature development.
    - ○ Add the new images and split them among train/dev/test.
    - ◉ Augment your data to increase the number of images of the new bird species.

    **Correct**
    Generating a sufficient number of images of the new species is crucial for your model to learn its features effectively.

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.

    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.)

    - ☐ 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.
    - ☑ 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**
    A smaller dataset can expedite training and allow for more iterations, potentially leading to a satisfactory model faster.

    - ☑ Needing two weeks to train will limit the speed at which you can iterate.

    **Correct**
    The long training time constrains how quickly you can test and refine models.

    - ☑ Buying faster computers could speed up your team's iteration speed and thus your team's productivity.

    **Correct**
    Enhanced computational resources can reduce training time and improve productivity.