Questionnaire 2

1. Provide an example of where the bear classification model might work poorly in production, due to structural or style differences in the training data.

   There are many cases that the bear classification model could fail, especially if these cases were not represented in the training data:

   - The bear is partially obstructed
   - Nighttime images are passed into the model
   - Low-res images are passed into the model
   - The bear is far away from the camera
   - The bear training dataset is highly biased toward one type of feature (i.e., color)

2. Where do text models currently have a major deficiency?

   Text models can generate appropriate text (like replies or imitating author style); however, text models still struggle with generating *correct* responses. Given factual information (such as a knowledge base), it is still hard to generate responses that utilizes this information to generate factually correct responses, though this text can seem very compelling. This can be very dangerous, because people unfamiliar with the subject may not be able to evaluate the factual accuracy of the generated text.

3. What are possible negative societal implications of text generation models?

   The ability for text generation models to generate context-aware, highly compelling responses can be used at a massive scale to spread disinformation and encourage conflict.

   Models also reinforce bias (like gender or racial bias) in training data and can create a vicious cycle of biased outputs.

4. In situations where a model might make mistakes, and those mistakes could be harmful, what is a good alternative to automating a process?

   The prediction could then be reviewed by a human professional, who can determine whether the prediction was accurate or not. For example, a machine learning model for identifying strokes in CT scans can alert high priority cases for expedited review, while other cases can be sent to radiologist for regular review.

   Other models can also augment a medical professional's abilities, reducing risk but improving efficiency of the workflow. For example, deep learning models can provide useful measurements for radiologists or pathologists.

5. What kind of tabular data is deep learning particularly good at?

   Deep learning is good at analyzing tabular data that includes natural language, or high cardinality categorical items (containing larger number of discrete choices like zip code).

6. What's a key downside of directly using a deep learning model for recommendation systems?

   Current (as of 2020) machine learning approaches for recommendation systems will only tell what products a user might like, and may not be recommendations that would be helpful to the user. For example, if a user if familiar with other books from the same author, it isn't helpful to recommend those products even though the user bought the author's book.

7. What are the steps of the Drivetrain Approach?
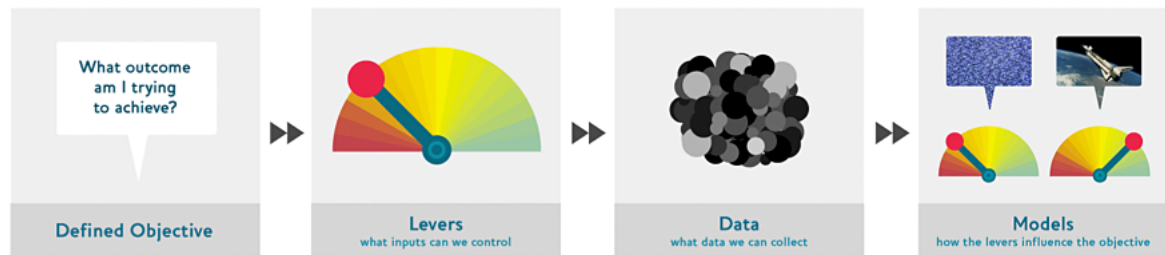
8. How do the steps of the Drivetrain Approach map to a recommendation system?

   The **objective** of a recommendation engine is to drive additional sales by surprising and delighting the customer with recommendations of items they would not have purchased without the recommendation. The **lever** is the ranking of the recommendations. New **data** must be collected to generate recommendations that will cause new sales . This will require conducting many randomized experiments in order to collect data about a wide range of recommendations for a wide range of customers. This is a step that few organizations take; but without it, you don't have the information you need to actually optimize recommendations based on your true objective (more sales!)

9. Create an image recognition model using data you curate, and deploy it on the web. Placeholder

10. What is `DataLoaders`?

    The `DataLoaders` class is the class that passes the data to the fastai model. It is essentially a class that stores the required `DataLoader` objects (usually for train and validation sets).

11. What four things do we need to tell fastai to create `DataLoaders`?

    - what kinds of data we are working with
    - how to get the list of items
    - how to label these items
    - how to create the validation set

12. What does the splitter parameter to `DataBlock` do?

    The splitter parameter is used as a way for fastai to split up the dataset into multiple subsets (usually train and validation sets). For example, to randomly split the data, you can use fastai's predefined `RandomSplitter` class, providing it with the proportion of the data used for validation.

13. How do we ensure a random split always gives the same validation set?

    It turns out that it is impossible to create truly random numbers. Instead, computers use a process known as a pseudo-random generator. However, this process can be controlled using a random seed. By setting a seed value, the pseudo-random generator will generate the same "random" numbers in a fixed manner which will be the same for every run as long as the seed is the same. By using a random seed, we can generate a random split that always gives the same validation set.

14. What letters are often used to signify the independent and dependent variables?

15. What's the difference between the crop, pad, and squish resize approaches? When might you choose one over the others?

16. What is data augmentation? Why is it needed?

17. What is the difference between item_tfms and batch_tfms?

18. What is a confusion matrix?

19. What does export save?

20. What is it called when we use a model for getting predictions, instead of training?

21. What are IPython widgets?

22. When might you want to use CPU for deployment? When might GPU be better?

23. What are the downsides of deploying your app to a server, instead of to a client (or edge) device such as a phone or PC?

24. What are three examples of problems that could occur when rolling out a bear warning system in practice?

25. What is "out-of-domain data"?

26. What is "domain shift"?

27. What are the three steps in the deployment process?

Further Research

1. Consider how the Drivetrain Approach maps to a project or problem you're interested in.

2. When might it be best to avoid certain types of data augmentation?

3. For a project you're interested in applying deep learning to, consider the thought experiment "What would happen if it went really, really well?"

4. Start a blog, and write your first blog post. For instance, write about what you think deep learning might be useful for in a domain you're interested in.