

Team 3

1. Why is a complete and meaningful sentence specifically selected for the content of a speech dataset? Does this provide any benefit for training the model?
2. How to preserve the needy part for data in the preprocessing step in detail?
3. How to identify the sensitive part?
4. Have you considered using real-world data as training data to make your model handle real inputs better?
5. Why are the sample frequencies of speech data and noise data different?
6. I found that there are many ways to perform feature extraction, such as using wavelet transforms or MFCCs. Why did you choose STFT specifically for your analysis?

Team 7

1. I really like your videos, and your motivation is admirable. However, it seems that the videos don't mention how you plan to implement things, such as how to collect voice data. That seems like it would be a very tedious part. Also, how do you plan to measure the performance of your model at the end? Additionally, I would like to know how to extract the features you mentioned from the audio.
2. I would like to know more details about the implementation, including what training data you will use and how you plan to evaluate your model's performance.
3. This is a really impressive way to represent your goals and motivation. However, it seems that you didn't mention how to collect the training dataset and how you evaluate the accuracy of the predictions. Are you using publicly available datasets, or do you plan to gather your own data?
4. I could see that you're trying to solve real problems through AI. Though you've described that you want to implement the model through CNN, I wonder about further details and where your data comes from? And how do you want to evaluate your model?
5. I wonder why you chose CNN as the training model. Have you compared the benefits of CNN with other models?
6. The video is incredibly creative and has sparked my curiosity about the implementation of the project. I'm particularly interested in understanding the specific approach you took in developing it. For instance, what model did you use to train the data, and how did you choose it?

Team 32

1. How will you measure the performance of your model? Because this part seems to be ignored in the video.
2. How to extract the feature of the audio?
3. Why did you choose the sound of birds as the learning target? Is it hard to record the realistic sound?
4. I'm curious about what kind of goal you're trying to reach, or what kind of result you expect. If you just want to generate bird sounds but without any limitation or approach, what is the purpose of generating new bird sounds using ML? Or do you expect to have better performance than any traditional method?
5. Will the sample species of birds be constrained to a certain region or type?
6. Considering the goal of generating realistic bird sounds with machine learning, how do you plan to address challenges such as regional variations in calls of the same species?

Team 35

1. Will the speech output also convey emotions?
2. Are the training data coming from the internet, or do you plan to record them yourself? How do you ensure that those are sufficient to reflect real-life emotions?
3. I think you can not only use the user's speech to detect their emotion but also consider the user's text.
4. I think the idea you came up with is pretty impressive, but I want to know how you will evaluate your model performance, whether by objective evaluation or statistics. Besides, since the scope is quite big, maybe you can consider whether your time and hardware support such a model.
5. I think there might be some situations that contain multiple emotions; how will you handle this when detecting emotion?
6. What strategies will you employ to handle potential ambiguities in emotions during both the training and evaluation phases?

Team 42

1. I'm curious about how you will compare the performance of your models. Will it be a pixel-by-pixel comparison, or is there another method?
2. I would like to see some details on how to balance between the generator and discriminator of the GAN.
3. Do you consider any preprocessing to make the features of your images represented in a better way?
4. I would like to know more about how you plan to implement the "privacy protection" part.
5. How do you plan to handle cases where PSNR and SSIM metrics do not align with human perception of image quality, especially in terms of color accuracy and detail?