OLA Interview Experiences

# **Slot: 1**

# **Procedure**

1. Test: Yes. (Online)
2. Interview Mode: Virtual (Zoom)

-----------------------------------------------------------------------------------------------------------------

# Subhasis Biswas

**Personal View**: Less theoretical, more of DevOPs + AI/ML mix.

**Status**: Shortlisted through online test. Didn’t proceed to the second technical round.

**Interview Description**

## **Round 1:**

Started with a brief generic mutual introduction. Interviewer worked on OLA Krutrim LLM related stuff. Asked me regarding the overall architecture of my internship product.

As evident from the topic of the internship, it was on safeguarding LLMs from misusages and hallucinations etc; I was also asked about the prospects of its scalability. Then I was asked about my MTech project.

The topic then moved to ViT, the overall patching of the images (An Image is Worth 16x16 Words) and the high-level overview of the model architecture. Questions on class imbalance scenarios and their mitigations.

Conversation then shifted to general very in-depth transformer architecture. We were asked to implement a single self-attention unit (used torch). It was being coded live.

Then some questions as follows:

• A model with a quite high parameter count is being trained on a small dataset, what is expected to happen?

• How to detect overfitting?

• How to prevent?

• Is accuracy a good metric in all scenarios?

• Explain AUC-ROC.

Finally, the coding question:

If the letters in a string ***s*** are shifted by a certain fixed amount, say ***k***, then it can act as an encrypter. Write out the functions to produce the encrypted string and from the encrypted string, decode the original given key.

**Possible Cause of Rejection**: Unknown. I was very positive about the interview.

-----------------------------------------------------------------------------------------------------------------

# Sameeksha Bhatia

Personal View:

Status: Rejected

Interview Description

## Round 1: Technical

1. what is the difference between generative and non-generative models? (in terms of learning probability distribution)

2. What is MSE loss, why don’t we use it without the squared term (not MAE)

3. what is regularization, types of regularization, how does it introduce sparsity?

4. Why does l2 regularization does not introduce sparsity? Why do the weights not go to zero exactly?

5. What will you do for classification of an imbalanced data if you are not allowed to use sampling methods.

6. Asked to explain about projects from resume, details of vision transformer.

7. How does cross attention help?

8. Asked me to share the screen and open “attention is all you need” article. From the transformer diagram, also asked about all the layers, their functions, why masked attention in decoder, how is this autoregressive, batch v/s layer normalization, why residual connection, why is it called scaled dot product attention.

9. Problem with RNNs

10. How does attention solve it?

11. Asked to open VS code and code for the fair dice roll

12. Asked about variational autoencoders, how can we do fraud detection using VAEs

Told me I will be shortlisted for the next round and the interviewer said he was impressed and mentioned that I should be prepared for the next round, but they did not shortlist me further:”)

Round 2: HR round - not shortlisted