OBSERVE.AI - ML Engineer (NLP)

# Slot: 1

# Procedure

CV Screening > Interview Mode: Physical (Venue: OCCAP Building)

|  |  |  |
| --- | --- | --- |
| Interview Rounds | Duration | Interview Focus Areas |
| ML Understanding | 30 minutes interview + 5 mins feedback | * General understanding of Machine learning concepts * Walk through the Projects in the resume * Domain understanding - NLP and Speech Projects * Understand ML skills - level |
| ML Formulation & Hands on | 1 hour interview + feedback | * Domain-related problems will be shared with them. * Scope the problem Clarify the right questions and data set * Recommend the solution |
| HM Round | 30 Mins | * Understanding of the Role and the interest in the work |
| HR Round | 15 minutes | * Role fit and career aspiration |

Interview Process Overview:

* 9 students were selected for interviews, with an additional 3 on the waitlist.
* The interview process consisted of multiple rounds, with some candidates starting with ML understanding while others began with ML formulation and hands-on tasks.
* The HM and HR rounds were conducted only after clearing both technical rounds. The HM-to-HR order was followed, but the ML rounds could be in any order. All rounds were elimination rounds. Every interviewer was taking notes of the interview.
* **Final Selections:** 2 students were selected for NLP positions and 1 for Speech.

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

# Rushikesh Pawar

**Personal View**: Interviews focused on fundamental concepts as well as problem-solving approaches.

**Status**: Selected

## Round 1: ML formulation and Hands on

There were multiple problems, I got the problem related to RAG.

The interviewer shared a Google Colab notebook, and I was asked to share my screen while being recorded. I was allowed to use the internet for documentation but could not directly search for the solution. LLM-based solutions were not permitted.

I had to go through the problem and then discuss the possible approaches.

**Problem Statement:** Given a query string, retrieve the closest document from a set of documents.

I discussed both a brute force and the RAG approach. The interviewer was polite and asked questions to understand my thought process. This discussion lasted about 20 minutes.

**Implementation:** I was tasked with implementing the algorithm using NumPy, given an example query and a small toy dataset of documents. The colab notebook had a rough structure with function names and docstrings to guide implementation.

After the implementation, he asked me if I had any questions for him. I asked 1-2 questions.

**Skills Required:** Familiarity with basic string operations (split, replace, lower, isalnum) to preprocess the strings, Python loops, lists, if-else statements, and NumPy to implement different functions for embedding and calculating similarity.

## **Recommendation:** It's helpful to practice implementing the multi-headed attention mechanism's forward pass in 3 ways: first using PyTorch, then using NumPy (without using PyTorch), and finally, implementing it in Python without NumPy and PyTorch. Having done this before, I had no difficulty with the implementation.

## Round 2: ML Understanding

The interviewer asked me to choose an architecture or algorithm of my choice and explain it. I chose Transformers and explained the architecture.

Key Questions:

* What is positional encoding?
* What are the learnable parameters in a Transformer, and how are they learned?
* What is the loss function?
* What would happen if we initialized all the Wk, Wq, and Wv matrices with a constant value?
* What if we used, He initialization for Wk and then used the same matrix for Wq and Wv as well?
* What are the regularization methods used in Transformers?
* What are the differences between encoder-only, decoder-only, and encoder-decoder Transformer models?

Then he asked me two formulations/approach-based questions

**Scenario 1:** If a trained language model (decoder-only) encounters a random token during inference, such as "IISc is <randomword>," how would you handle it?

(I asked him do we want to correct this mistake/random token, or do we want to generate next tokens correctly even if the current token is wrong/random/low probability)

I was asked to provide solutions for both correcting the token and generating the next tokens correctly, even if the current token is incorrect. (I discussed my approach and thought process, in between he asked me questions to check my understanding and reasoning)

He asked me to explain different decoding strategies and questions related to decoding strategies and hyperparameters (temp, top\_p, top\_k) etc. He was hinting that beam search is also a possible solution to correct the current wrong token and I got it. (my solution was a bit different).

**Scenario 2:** I was asked to design a classification system to categorise a product into categories and subcategories. I discussed my thought process and reasons for my approach, while the interviewer asked me about the potential drawbacks of my approach and asked how to address them.

Asked me if I had any questions.

## Round 3: Hiring Manager

This was a rapid-fire round, with the interviewer asking questions continuously for about 30-40 minutes. The focus was on my thought process and attitude toward problem-solving and challenges.

(As there were so many questions asked in about 30-40 mins, I don’t recall all of them but below are some of the questions that I remember.)

Key Questions:

* How were the previous interviews?
* Why do you want to work at Observe?
* What is your MTech dissertation? What are the challenges and problems that you are trying to solve and what is the real-world impact of the project?
* How would you measure the success of your MTech dissertation?
* What are other applications (in other areas except materials/physics) of your MTech dissertation?
* If you were to create a product based on your MTech project, what would it be?
* Your project uses GNNs, but Observe works with LLMs, which are orthogonal areas. Why do you want to join?
* What challenges did you face transitioning from a mechanical engineering background to data science, which is more related to CS and math? How did you overcome them?
* What is your long-term goal or dream?

**Q&A:** I asked 2-3 questions, leading to a nice discussion. The interviewer was an alumnus of IISc (specifically from SERC).

## Round 4: HR

This was another rapid-fire round, lasting about 30-35 minutes. The HR was taking notes throughout, and the questions were intense. I don’t recall all the questions but below are the few I can recall.

Key Questions:

* How did you find the previous interview process?
* What did you learn from it?
* Why do you want to join Observe?
* What do you know about Observe and its products?
* Tell me about your BTech project. How could you turn it into a startup?
* What are your expectations from the company and your team?
* What kind of manager would you like to work with?
* How was your manager in your internship?
* What are your strengths, and how did you develop them?
* How do you handle uncertainty?
* Out of Observe's five values (Customer First, We, Own, Grit, Agile), which one do you relate to the most?
* Why do you relate with “own” value the most? Did you own a project or some event in the past?
* Which side are you more interested in: applied ML or research?

Do you have any questions for me? (I again asked 2-3 questions)

Note: This was one of the most interesting HR rounds I experienced.

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

# Vivek Dhamale

Personal View: Status:

Interview Description

## Round 1:

## Round 2:

## Round 3:

## Round 4:

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

# Sameeksha Bhatia

Personal View: I was first waitlisted but called later for the interview. The interview process they have is rigorous and they test all aspects of your profile.

Status: Rejected

Interview Description

## Round 1: Hands on ML

Duration: 50-60 minutes

Same problem statement as Rushikesh. To obtain the top similar words from the given corpus based on the input string.

## Round 2: ML understanding

1. Write cross entropy loss function
2. How is it related to KL divergence
3. Explain transformer architechture
4. Role of multihead attention in detail, why is it done
5. Difference between word2vec, glove and bert embeddings
6. Asked to write the dimensions of each layer in a transformer.
7. Why is attention scaled
8. Difference between the posititional encoding of transformer model and T5
9. What was the problem with RNNs, how do transformers solve it.

## Round 3: Hiring Manager

1. How were your previous interviews?
2. Any difficulty you faced
3. What were the learning outcomes of your internship
4. Gave some situational questions about stakeholders and designing a problem statement.

(I made the mistake of assuming that this was supposed to be more of a behavioral round and was answering with that in my mind, but the interviewer expected some technical point of view as well, which I did not give very clearly. There was a disconnect and i was not shortlisted for the next round)

## Round 4: HR – Not shortlisted further