

# Top 10 Scenario-based NLP Interview Questions

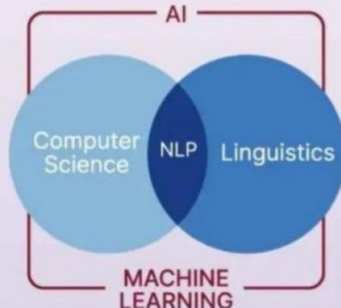


Shiksha Online 

Updated on Apr 19, 2023 18:25 IST

Natural Language Processing (NLP) has revolutionized the way computers interact with humans by enabling them to understand and analyze natural language. With the ability to hear, analyze, quantify, and identify significant parts of speech, NLP has opened up a world of possibilities, from chatbots to sentiment analysis to market intelligence.

## Scenario Based NLP Interview Questions



In this article, we have provided you with a list of top scenario-based **NLP interview questions**. These questions are designed to assess a candidate's understanding of NLP, from the basics of NLP and its limitations, to more advanced concepts like evaluating the performance of an NLP model.

The applications of NLP are diverse and expanding rapidly, with devices like Amazon's Alexa becoming ubiquitous in households worldwide. For businesses, NLP offers powerful tools for business intelligence and consumer monitoring, helping to



**Disclaimer:** This PDF is auto-generated based on the information available on Shiksha as on 01-Nov-2023.

drive innovation and gain a competitive edge.

## Table of Content

- [Top NLP interview questions](#)

## Top NLP Interview Questions

### Q1. What is natural language processing (NLP), and how does it relate to artificial intelligence?

This is one of the fundamental NLP Interview Questions that you must be aware of before proceeding to advanced level questions. [Natural Language Processing](#) is a subfield of [artificial intelligence](#) that focuses on enabling computers to understand and analyze human language. It involves the use of computational techniques to process and analyze large amounts of natural language data, such as text or speech. The ultimate goal of NLP is to enable computers to interact with humans in a natural and intuitive way, as if they were communicating with another person.

NLP is closely related to artificial intelligence because it relies on many of the same techniques and approaches used in other AI fields, such as [machine learning](#) and [deep learning](#). By leveraging these techniques, NLP researchers and practitioners are able to develop sophisticated [algorithms](#) and models that can process and analyze natural language data with high accuracy and efficiency.

*Explore:*

<a href="#">online deep learning courses</a>	<a href="#">online machine learning courses</a>
<a href="#">online data science courses</a>	<a href="#">free artificial intelligence courses</a>

### Q2. Can you explain the difference between syntax and semantics in NLP?

In NLP, syntax refers to the structure of language, such as the rules and patterns that govern how words are combined to form sentences. This might include things like sentence structure, word order, and grammatical rules. Syntax is important



because it provides a framework for understanding the meaning of sentences and the relationships between different words and phrases.

Semantics, on the other hand, refers to the meaning of language, such as the concepts and ideas that are conveyed by words and sentences. This might include things like word definitions, contextual meaning, and inferred meaning. Semantics is important because it allows us to understand the deeper meaning behind language and interpret the nuances and subtleties of human communication.

### **Q3. How would you approach developing a machine learning model for a named entity recognition (NER) task?**

Named entity recognition (NER) is a common NLP task that involves identifying and categorizing specific types of entities, such as people, organizations, and locations, in a piece of text. To develop a machine learning model for a NER task, we would first gather a large dataset of labeled examples, where each example contains a piece of text and the associated named entities.

Next, we would preprocess the text data by performing [tokenization](#), [normalization](#), and other techniques to create a standardized format for the text. Then, we would choose a suitable [machine learning algorithm](#), such as a conditional random field (CRF) or a recurrent neural network (RNN), and train the model on the labeled dataset.

During training, we would tune the hyperparameters of the model to optimize its performance on the validation set. Finally, we would evaluate the model on a test set to ensure that it is accurate and effective at identifying named entities in text.

### **Q4. What are some common challenges or limitations of NLP, and how do you overcome them?**

One common challenge in NLP is dealing with language ambiguity, which occurs when a word or phrase can have multiple possible meanings depending on the context. To overcome this challenge, we would leverage techniques like part-of-speech tagging, syntactic parsing, and named entity recognition to better understand the structure and meaning of the text.



Another challenge is dealing with out-of-vocabulary (OOV) words, which are words that the NLP model has not seen before and does not have a representation for. To overcome this challenge, we might use techniques like subword tokenization or character-level representations to enable the model to handle previously unseen words.

A final challenge is dealing with bias in the data or the model itself. To overcome this challenge, we would carefully analyze the data and the model to identify any biases or limitations, and take steps to mitigate or correct them. This might involve using techniques like adversarial training, data augmentation, or algorithmic fairness to ensure that the NLP model is accurate, reliable, and fair.

#### **Q5. How would you approach designing a chatbot for a customer service use case?**

When designing a chatbot for customer service, the first step is to understand the customer's needs and the types of queries they are likely to have. From there, we would develop a list of potential user intents and map them to appropriate responses. This would involve creating a set of rules or a decision tree to guide the chatbot's behavior.

Next, we would choose a natural language processing (NLP) framework that can handle the specific requirements of the customer service use case. This might involve training the chatbot on a large corpus of customer interactions to help it recognize common phrases and language patterns.

Finally, we would test the chatbot extensively to ensure that it can handle a range of scenarios and provide accurate and helpful responses to users.

#### **Q6. How would you use machine learning to extract meaningful insights from unstructured text data?**

To extract meaningful insights from unstructured text data using machine learning, we would first preprocess the data to remove any noise and standardize the format of the text. Then, we would use techniques like tokenization and lemmatization to



break the text down into smaller, more manageable units.

From there, we would choose an appropriate machine learning algorithm such as a [clustering](#) or classification algorithm to group similar pieces of text together or categorize them into different topics. We would train the algorithm on a labeled dataset and optimize its hyperparameters to achieve the best performance.

Finally, we would analyze the output of the algorithm to identify patterns and trends in the data and use these insights to make data-driven decisions.

### **Q7. Can you explain the difference between supervised and unsupervised learning in NLP?**

In [supervised learning](#), the machine learning model is trained on a labeled dataset, where each data point is associated with a predefined label or output. The goal of the model is to learn a mapping between the input features and the corresponding output labels, so that it can accurately predict the output for new, unseen data points.

In contrast, [unsupervised learning](#) does not rely on labeled data. Instead, the model is trained on a dataset of unstructured or unlabeled data and is tasked with finding patterns or structure within the data itself. This might involve techniques like clustering or dimensionality reduction to group similar data points together or represent them in a lower-dimensional space.

In NLP, supervised learning is commonly used for tasks like [sentiment analysis](#) or named entity recognition, where the output label is already defined. Unsupervised learning is often used for tasks like topic modeling or word embeddings, where the goal is to identify underlying patterns in the data without any predefined labels.

### **Q8. How would you handle a situation where a text classification model is misclassifying a significant number of documents?**

This is one of the important scenario based NLP Interview Questions. If a text classification model is misclassifying a significant number of documents, there are several steps we would take to diagnose and address the issue.



First, we would evaluate the performance of the model on a validation set and calculate metrics like [precision and recall](#), and F1 score to determine which classes are being misclassified the most.

Next, we would examine the misclassified documents themselves to see if there are any patterns or commonalities that might explain why the model is struggling to classify them correctly. This might involve manually annotating the documents with the correct labels and retraining the model on the updated dataset.

I might also try experimenting with different feature representations or tweaking the hyperparameters of the model to see if that improves performance.

### **Q9. How would you go about developing a system for sentiment analysis of social media posts?**

To develop a system for sentiment analysis of social media posts, we would first gather a large dataset of social media posts and their associated sentiment labels (e.g., positive, negative, neutral).

Next, we would preprocess the text by removing any noise, such as URLs or emojis, and performing tokenization and normalization to create a standardized format for the text.

Then, we would choose a suitable NLP framework or library, such as spaCy or NLTK, to perform sentiment analysis on the text data. This might involve training a machine learning model using a supervised learning approach, where the model is trained on the labeled dataset to predict the sentiment of new, unseen social media posts.

Alternatively, we might use a pre-trained language model, such as BERT or [GPT](#), which has already been fine-tuned for sentiment analysis tasks. These models are trained on large amounts of data and can often provide state-of-the-art performance on sentiment analysis tasks.

Finally, we would test and evaluate the system on a holdout dataset to ensure that it is accurate and can generalize well to new, unseen data. we would also monitor the performance of the system over time and retrain or update it as needed to



ensure that it remains effective in detecting sentiment in social media posts.

### Q10. How would you go about evaluating the performance of an NLP model?

To evaluate the performance of an NLP model, we would first define the evaluation metric or metrics that are most relevant to the task at hand. For example, if we are developing a sentiment analysis model, we might use metrics like accuracy, precision, recall, and F1 score to measure its performance.

Next, we would split my dataset into a training set, a validation set, and a test set. we would use the training set to train the model, the validation set to tune its hyperparameters and optimize its performance, and the test set to evaluate its final performance.

During evaluation, we would measure the model's performance on the test set using the chosen evaluation metrics. we would also analyze the model's outputs and errors to gain insights into its strengths and weaknesses, and to identify areas for improvement.

Finally, we would compare the performance of my NLP model to the performance of other state-of-the-art models or benchmarks in the field, to assess its relative effectiveness and identify opportunities for future improvement.



### Top 9 NumPy Interview Questions

In this article, we will discuss basic and intermediate NumPy Interview Questions in most simplified ways with example.



### Data Structure Interview Questions

Data structures and algorithms form a crucial component of any programming interview. **Data structure interview questions** are also asked at data science and Java interviews.





## Deep Learning Interview Questions and Answers

*Deep learning is the technique that comes closest to the way we humans learn. It has become increasingly important in the field of artificial intelligence (AI) and jobs...[read more](#)*

### Endnotes

NLP has already become an indispensable tool in today's digital world, and its significance will only grow in the years to come. As NLP technology continues to advance and evolve, it holds immense potential for future innovations and improvements in areas like healthcare, education, and customer service. We hope that the set of scenario based NLP Interview Questions have helped you in understanding the concept in more depth.

