Graphic Era
Hill University

End Term (Odd) Semester Examination November 2025

Roll no.....

Name of the Course and semester: BTech-CSE(AI and ML Specialization) 5th Semester
Name of the Paper: Natural Language Processing and Computer Vision
Paper Code: TCS 564

Time: 3 hour Maximum Marks: 100

Note:

(i) All the questions are compulsory.
(ii) Answer any two sub questions from a, b and c in each main question.
(iii) Total marks for each question is 20 (twenty).
(iv) Each sub-question carries 10 marks.

Q1. CO1 (2X10=20 Marks)

a. Tokenization is a fundamental step in NLP, but informal texts challenge standard tokenizers. Examples are: Strong emotions like "goooood", Emojis, Hashtags (#MondayMotivation).
Suggest how will you improve tokenization for such social media text. Then explain how your improvement will affect the entire NLP pipeline.
b. Chunking in multi-language Hinglish text (mixing of Hindi and English) requires different approach. Propose how will you handle such multi language while performing chunking.
c. Write a Python program that compares recall between stemming and no stemming in a simple retrieval task.

Q2. CO2 (2X10=20 Marks)

a. Define and Discuss Cosine Similarity. Find the cosine similarity between the following documents:
Doc1: Antarctica is the fifth largest continent, roughly twice the size of Australia.
Doc2: The largest continent is Asia, more than five times bigger than Australia and three times than Antarctica.
b. What are Word Embeddings? Differentiate between CBOW and Skip-gram using appropriate example.
c. Using the example sentences below, compute the TF-IDF weight of each word in each document:
S1: The cat loves to chase the laser pointer.
S2: Dogs bark loudly at the cat every night.
S3: The laser pointer is red and fun for cats.

Q3. CO3 (2X10=20 Marks)

a. What is Computer Vision and why do we use Computer Vision? Discuss some applications of Computer Vision. "Images also have ambiguity". – Explain this statement very briefly.
b. Define Hue, Brightness and Saturation. Explain the JPEG algorithm in detail.
c. Write complete code to perform sentiment analysis using NLTK or spaCy. Assume that you have a dataset of reviews of Restaurants in a TSV file. The dataset contains two columns, one contains the reviews and the other a value (1 or 0), 1 indicates positive review and 0 indicates negative review. There are total 1500 reviews. Perform all necessary preprocessing work. Make any other assumptions you require.

Q4. CO4 (2X10=20 Marks)

a. "Image Segmentation is subjective in nature for humans"- What are the different criteria based on which humans find segments in images? Explain either the k-means segmentation or mean shift segmentation technique using an example.
b. What is a Linear Shift Invariant System (LSIS)? Given that f(x) convolves with h(x) to produce g(x), how can you obtain h(x)? Give a brief idea as to how the idea of Convolution is used in Image Processing.
c. Differentiate between Sobel and Canny edge detection. Name an edge detection technique that uses double integration in place of single integration.
Write code in OpenCV for Canny and Sobel Edge detection.

Graphic Era HILL UNIVERSITY

Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)
University under section 2(f) of UGC Act, 1956

End Term (Odd) Semester Examination November 2025

Q5. CO5 (2X10=20 Marks)

a. What is the working principle of YOLO? Explain the YOLO architecture. List three improvements made in the current YOLO version.

b. List some popular object detection frameworks. Write code in OpenCV for detecting a face in an image using Haar Cascades.

c. Provide the working principles of the following cues: Relative Size, Texture Gradient, Interposition, Linear Perspective, Height in Plane, Light and Shadow, Binocular Disparity, Binocular Convergence