CURRICULUM & SYLLABUS

B. Tech. (Honours) *in*

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

2024 SCHEME

CURRICULUM

B.Tech (Honours) is an enhanced version of the Bachelor of Technology degree, offering the students the opportunity to undertake additional courses within their own discipline. This pathway allows students to deepen their knowledge in emerging or advanced areas of Engineering relevant to their field of study, providing a stronger foundation for specialized career paths or further academic pursuits.

For the award of B.Tech (Honours) in Artificial Intelligence and Data Science, the student shall fulfill all the curricular requirements for B.Tech in Artificial Intelligence and Data Science as per SJCET B.Tech Academic Regulations 2024 and shall earn 15 additional credits by undergoing the following courses, which shall be further governed by clause R16 of the Regulations.

Sl.No	Semester	Course Code	Course Name/Type	Weekly hours				Total Marks		Credits
				L	T	P	SS	CIE	ESE	
1	4	24SJHNADT409	Foundations of Natural Language Processing	3	1	0	5	40	60	4
2	5	24SJHNADT509	Deep Learning for NLP	3	1	0	5	40	60	4
		24SJHNADM5XX	Approved MOOC *							
3	6	24SJHNADT609	Transformer models and Advanced NLP	3	1	0	5	40	60	4
		24SJHNADM6XX	Approved MOOC *							
4	7	24SJHNADT709	Prompt Engineering	3	0	0	5	40	60	3
		24SJHNADM7XX	Approved MOOC *							
Total Credits										15

*MOOC to be approved by the Academic Council on recommendation of the Board of Studies.

SEMESTER S4

FOUNDATIONS OF NATURAL LANGUAGE PROCESSING

Course Code	24SJHNADT409	**CIE Marks**	40
Teaching Hours/Week (L: T: P: R)	3:1:0:0	**ESE Marks**	60
Credits	4	**Exam Hours**	2Hrs. 30Min.
Prerequisites (if any)	None	**Course Type**	Theory

Course Objectives:

1. Introduce the fundamental concepts, linguistic background, and computational approaches in Natural Language Processing (NLP).
2. Enable students to apply preprocessing and text representation techniques for effective text analysis and feature extraction.
3. Develop the ability to implement and analyze text mining and machine learning methods for tasks such as classification, sentiment analysis, and information extraction.
4. Familiarize learners with advanced NLP applications such as information retrieval, question answering, and machine translation using modern NLP tools and frameworks.

SYLLABUS

Module No.	Syllabus Description	Contact Hours
1	**Introduction to NLP** Natural Language Processing – Tasks, Applications, and Challenges. Linguistic background, spoken and written language processing. Machine Learning for NLP – Naïve Bayes, Logistic Regression, Support Vector Machines. Finite State and Statistical Models, Finite State Methods in NLP.NLP System Pipeline – Data acquisition, text extraction, pre-processing steps (tokenization, stemming, lemmatization, stop-word removal).	11
2	**Text Representation and Parsing** Text Representation Models: One-Hot Encoding, Bag-of-Words, Bag-of-n-Grams, TF-IDF, Word2Vec (CBOW and Skip-Gram). Part-of-Speech Tagging, Context-	11

	Free Grammars, Parsing Techniques, and Syntax Trees. Feature Engineering, Model Evaluation Metrics. Morphological Analysis and Semantic Interpretation.	
3	**Text Classification and Information Extraction** Text Classification Pipeline and Algorithms – Naïve Bayes, Decision Trees, SVM, KNN. Sentiment Analysis – Supervised and Unsupervised Learning Approaches. Information Extraction: Named Entity Recognition (NER), Relation Detection, Sequence Labeling, and Evaluation. Document Categorization, Clustering (Hierarchical and Flat), and Document Summarization using Lexical Chains.	**11**
4	**Advanced NLP Applications** Information Retrieval – Inverted Index, TF-IDF Scoring, Evaluation Metrics. Question Answering Systems – Factoid QA, Passage Retrieval, Answer Extraction. Machine Translation – Classical, Statistical, and Phrase-based Translation Models. Word Sense Disambiguation, Coreference Resolution, and Multilingual NLP. Case Studies using Python/NLTK/SpaCy for tokenization, tagging, and text classification.	**11**

Course Assessment Method

(CIE: 40 marks, ESE: 60 marks)

Continuous Internal Evaluation Marks (CIE):

Attendance	Assignment	Internal Examination-1 (Written)	Internal Examination-2 (Written)	Total
5	**15**	**10**	**10**	**40**

End Semester Examination Marks (ESE)

In Part A, all questions need to be answered and in Part B, each student can choose anyone full question out of two questions

Part A	Part B	Total
• 2 Questions from each module. • Total of 8 Questions, each carrying 3 marks **(8x3=24 marks)**	• Each question carries 9 marks. • Two questions will be given from each module, out of which 1 question should be answered. • Each question can have a maximum of 3 sub divisions. **(4x9=36 marks)**	**60**

Course Outcomes

At the end of the course students should be able to:

Course Outcome		Bloom's Knowledge Level (KL)
CO1	Explain the fundamental concepts, tasks, and challenges of Natural Language Processing, and describe the basic language structures used in text analysis.	**K2**
CO2	Apply text preprocessing, feature extraction, and representation models (such as TF-IDF, Word2Vec) for effective text processing and analysis.	**K3**
CO3	Implement and analyze machine learning and text mining techniques such as classification, sentiment analysis, and information extraction for real-world applications.	**K3**
CO4	Summarize and evaluate advanced NLP applications including information retrieval, question answering, and machine translation using modern tools and libraries.	**K3**

Note*: K1-Remember, K2-Understand, K3-Apply, K4-Analyse, K5-Evaluate, K6-Create*

CO-PO Mapping Table (Mapping of Course Outcomes to Program Outcomes)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11
CO1	3	2									1
CO2	3	3	2	2	2						1
CO3	3	3	3	3	3				2		1
CO4	3	2	3	3	3				2		2

Note:*1: Slight (Low),2: Moderate (Medium),3: Substantial (High)*

Text Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition And Year**
1	Speech and Language Processing	Daniel Jurafsky & James H. Martin	Pearson.	3rd Edition, 2023
2	Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems	Sowmya Vajjala et al.	O'Reilly	1st Edition 2020.

Reference Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition And Year**
1	Foundations of Statistical Natural Language Processing	Christopher D. Manning, Hinrich Schütze e	MIT Press	1st Edition 1999
2	Natural Language Understanding	James Allen	Pearson.	2nd Edition 1995

Video Links (NPTEL, SWAYAM…)	
Module No.	**Link ID**
1, 2, 3	https://onlinecourses.nptel.ac.in/noc23_cs45/preview

SEMESTER S5

DEEP LEARNING FOR NLP

Course Code	24SJHNADT509	**CIE Marks**	40
Teaching Hours/Week (L: T: P: R)	3:1:0:0	**ESE Marks**	60
Credits	4	**Exam Hours**	2 Hrs. 30 Min.
Prerequisites (if any)	24SJHNADT409	**Course Type**	Theory

Course Objectives:

1) To understand the fundamentals of natural language processing and neural network architectures.
2) To learn modern deep learning approaches for representing and processing text data.
3) To explore state-of-the-art models such as Transformers, BERT, GPT, and their applications.
4) To gain practical experience in implementing NLP models for real-world tasks.

SYLLABUS

Module No.	Syllabus Description	Contact Hours
1	**Foundations of NLP and Deep Learning** Introduction to NLP: Word-level and sentence-level tasks, Language Models – N-gram models, evaluation metrics (perplexity) – Neural Networks: shallow vs. deep networks, activation functions, loss functions – Representation Learning – Word Embeddings: Word2Vec, GloVe, fast Text, multilingual embeddings with emphasis on Indian languages.	11
2	**Recurrent Models and Attention Mechanisms** Recurrent Neural Networks (RNNs) – RNN Language Models, GRUs, LSTMs, Bi-LSTMs – Sequence-to-Sequence models for translation and tagging – Attention Mechanism: concept, implementation – Transformers: architecture, "Attention is All You Need", encoder-decoder models.	11
3	**Pretraining and Large-Scale NLP Models** Self-Supervised Learning (SSL): objectives and methods – Contextual Representations: ELMo, BERT, GPT, T5, BART – Fine-tuning methods and applications: Question Answering, Dialogue Systems, Text Summarization – Instruction Fine-Tuning: FLAN-T5, Reinforcement Learning through Human Feedback (RLHF) – In-context Learning, Chain-of-	11

	Thought prompting, Scaling Laws – Overview of major Large Language Models and their architectural variations.	
4	**Advanced Fine-Tuning, RAG, and Interpretability** Parameter-Efficient Fine-Tuning (PEFT): LoRA, QLoRA – Handling long context, Retrieval-Augmented Generation (RAG) – Model Analysis and Interpretability: probing, visualization, explainability – Ethical considerations: bias, fairness, and responsible AI in NLP systems.	**11**

Course Assessment Method

(CIE: 40 marks, ESE: 60 marks)

Continuous Internal Evaluation Marks (CIE):

Attendance	Assignment	Internal Examination-1 (Written)	Internal Examination- 2 (Written)	Total
5	**15**	**10**	**10**	**40**

End Semester Examination Marks (ESE)

In Part A, all questions need to be answered and in Part B, each student can choose any one full question out of two questions

Part A	Part B	Total
• 2 Questions from each module. • Total of 8 Questions, each carrying 3 marks **(8x3 =24 marks)**	• Each question carries 9 marks. • Two questions will be given from each module, out of which 1 question should be answered. • Each question can have a maximum of 3 subdivisions. **(4x9 = 36 marks)**	**60**

Course Outcomes (COs)

At the end of the course students should be able to:

Course Outcome		Bloom's Knowledge Level (KL)
CO1	Explain the fundamental concepts of NLP, language modeling, and neural network foundations.	**K2**
CO2	Apply recurrent and attention-based neural architectures for sequence modeling and text generation tasks.	**K3**
CO3	Utilize pretrained transformer-based language models for various NLP applications such as QA, summarization, and dialogue modeling.	**K3**
CO4	Implement and analyze advanced fine-tuning strategies (e.g., PEFT, RAG) and discuss interpretability and ethical considerations in NLP systems.	**K3**

Note*: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create*

CO-PO Mapping Table (Mapping of Course Outcomes to Program Outcomes)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11
CO1	3	3	3	3							2
CO2	3	3	3	3	3						2
CO3	3	3	3	3	3						2
CO4	3	3	3	3	3						2

Note: *1: Slight (Low), 2: Moderate (Medium), 3: Substantial (High)*

Text Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	Speech and Language Processing	Daniel Jurafsky, James H. Martin	Prentice Hall / Pearson	3/e (Draft), 2023
2	Neural Network Methods for Natural Language Processing	Yoav Goldberg	Morgan & Claypool Publishers	1/e, 2017

Reference Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	Deep Learning	Ian Goodfellow, Yoshua Bengio, Aaron Courville	MIT Press	1/e, 2016
2	Transformers for Natural Language Processing	Denis Rothman	Packt Publishing	2/e, 2022
3	Deep Learning for Natural Language Processing	Palash Goyal, Sumit Pandey, Karan Jain	Apress	1/e, 2018

Video Links (NPTEL, SWAYAM…)	
Module No.	**Link ID**
1, 2, 3, 4	https://onlinecourses.nptel.ac.in/noc25_cs22/preview

SEMESTER S6

TRANSFORMER MODELS AND ADVANCED NLP

Course Code	**24SJHNADT609**	**CIE Marks**	40
Teaching Hours/Week (L: T:P: R)	3:1:0:0	**ESE Marks**	60
Credits	4	**Exam Hours**	2 Hrs. 30 Min.
Prerequisites (if any)	24SJHNADT509	**Course Type**	Theory

Course Objectives

1. To provide a deep understanding of modern NLP architectures, focusing on Transformer-based models such as BERT, GPT, and T5.
2. To impart practical knowledge of tokenization, embeddings, attention mechanisms, and transfer learning in NLP.
3. To equip students with the ability to fine-tune, evaluate, and deploy transformer-based NLP applications for real-world tasks.

SYLLABUS

Module No.	Syllabus Description	Contact Hours
1	**Introduction to NLP and Deep Learning for Text** — NLP pipeline – Text preprocessing – Word embeddings: Word2Vec, GloVe, FastText – Recurrent and convolutional architectures for text – Limitations of RNNs and CNNs – Evolution of Transformer architectures.	11
2	Transformer Architecture – Self-Attention Mechanism – Scaled Dot-Product Attention – Multi-Head Attention – Positional Encoding – Encoder-Decoder Architecture – Layer Normalization and Residual Connections – Applications in Sequence Modeling and Machine Translation.	11

3	**Pretrained Language Models** – BERT: Architecture, Masked Language Modeling, Next Sentence Prediction – GPT Family: Generative Modeling, Causal Attention – T5 and Encoder-Decoder Paradigm – RoBERTa, XLNet, and DistilBERT – Fine-tuning Techniques and Transfer Learning in NLP.	**11**
4	**Advanced Topics and Applications** – Prompt Engineering and Instruction Tuning – Sentence Embeddings and Semantic Search – Question Answering, Text Summarization, and Chatbots – Evaluation Metrics for NLP – Ethical Issues and Bias in Transformer Models – Trends in Multimodal and Multilingual NLP.	**11**

Course Assessment Method
(CIE: 40 marks, ESE: 60 marks)

Continuous Internal Evaluation Marks (CIE):

Attendance	Assignment/ Microproject	Internal Examination-1 (Written)	Internal Examination- 2 (Written)	Total
5	**15**	**10**	**10**	**40**

End Semester Examination Marks (ESE)

In Part A, all questions need to be answered and in Part B, each student can choose any one full question out of two questions

Part A	Part B	Total
• 2 Questions from each module. • Total of 8 Questions, each carrying 3 marks **(8x3 =24 marks)**	• Each question carries 9 marks. • Two questions will be given from each module, out of which 1 question should be answered. • Each question can have a maximum of 3 subdivisions. **(4x9 = 36 marks)**	**60**

Course Outcomes (COs)

At the end of the course students should be able to:

Course Outcome		Bloom's Knowledge Level (KL)
CO1	Explain the architecture and working principles of Transformer-based NLP models.	**K2**
CO2	Apply attention mechanisms and transformer components for NLP tasks.	**K3**
CO3	Fine-tune and evaluate pretrained transformer models for specific applications.	**K3**
CO4	Analyze limitations, ethical aspects, and emerging trends in advanced NLP systems.	**K3**

Note*: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create*

CO-PO Mapping Table (Mapping of Course Outcomes to Program Outcomes)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11
CO1	3	3	2	3							3
CO2	3	3	3	2							3
CO3	3	3	3	3							3
CO4	3	3	2	2							3

Note: *1: Slight (Low), 2: Moderate (Medium), 3: Substantial (High)*

Text Books

Sl. No.	Title of the Book	Author(s)	Publisher	Edition and Year
1	Attention Is All You Need (Research Paper)	Vaswani et al.	NeurIPS	2017
2	Speech and Language Processing	Daniel Jurafsky, James H. Martin	Pearson	3rd Edition, 2023
3	Transformers for Natural Language Processing	Denis Rothman	Packt Publishing	2nd Edition, 2022

Reference Books

Sl. No.	Title of the Book	Author(s)	Publisher	Edition and Year
1	Deep Learning for Natural Language Processing	Palash Goyal et al.	Apress	2018
2	Natural Language Processing with Transformers	Lewis Tunstall, Leandro von Werra, Thomas Wolf	O'Reilly	2022

Video Links (NPTEL, SWAYAM…)	
Module No.	**Link ID**
1, 2, 3, 4	https://youtube.com/playlist?list=PLqGkIjcOyrGnjyBHl4GE2S9kX47X96FH-&si=nDrmpNfgmnTyf1yx

SEMESTER S7

PROMPT ENGINEERING

Course Code	**24SJHNADT709**	**CIE Marks**	40
Teaching Hours/Week (L: T: P: R)	3:0:0:0	**ESE Marks**	60
Credits	3	**Exam Hours**	2 Hrs. 30 Min.
Prerequisites (if any)	None	**Course Type**	Theory

Course Objectives:

1. To develop students' practical skills in applying prompt engineering techniques to real-world applications, while fostering an awareness of the ethical considerations and challenges in the field

2. To give an understanding of contextual cues to mitigating biases with techniques for seamless interaction with AI systems.

SYLLABUS

Module No.	Syllabus Description	Contact Hours
1	Introduction to Prompt Engineering and Language Models :- Fundamentals of Natural Language Processing (NLP) - Overview of Language Models: From Rule-Based Systems to Transformer Architectures (e.g., GPT, BERT) - Understanding Prompts: Definition, Importance, and Applications - Introduction to Prompt Engineering: Techniques and Use Cases – Ethical Considerations in Prompt Engineering.	**9**
2	Techniques and Strategies in Prompt Engineering :- Designing Effective Prompts - Best Practices and Common Pitfalls; Prompt Tuning and Fine-Tuning Language Model; Using Zero-Shot, Few-Shot, and Multi-Shot Learning in Prompts; Exploring the Role of Context, Repetition, and Specificity in Prompt Responses; Advanced Prompt Engineering Techniques: Prompt Chaining, Iterative Prompting.	**9**
3	Applications of Prompt Engineering :- Prompt Engineering in Chatbots and Conversational AI; Content Generation: Creative Writing, Code Generation, and Data Augmentation; Prompt	**9**

	Engineering for Sentiment Analysis, Classification, and Translation; Integration of Prompt Engineering with Other AI Technologies (e.g., Computer Vision, Data Science); Real-World Case Studies and Industry Applications	
4	Challenges, Future Trends, and Research in Prompt Engineering :- Challenges in Prompt Engineering: Ambiguity, Bias, and Misinterpretation; Evaluating and Improving Prompt Performance: Metrics and Benchmarks; Future Trends: Emerging Techniques and the Evolution of Language Models.	**9**

Course Assessment Method (CIE: 40 marks, ESE: 60 marks)

Continuous Internal Evaluation Marks (CIE):

Attendance	Assignment/ Microproject	Internal Examination-1 (Written)	Internal Examination- 2 (Written)	Total
5	**15**	**10**	**10**	**40**

End Semester Examination Marks (ESE)

In Part A, all questions need to be answered and in Part B, each student can choose any one full question out of two questions

Part A	Part B	Total
• 2 Questions from each module. • Total of 8 Questions, each carrying 3 marks **(8x3 =24 marks)**	• Each question carries 9 marks. • Two questions will be given from each module, out of which 1 question should be answered. • Each question can have a maximum of 3 subdivisions. **(4x9 = 36 marks)**	**60**

Course Outcomes (COs)

At the end of the course, students should be able to:

Course Outcome		Bloom's Knowledge Level (KL)
CO1	Explain the core principles of NLP, language models, and the role of prompts in influencing AI behavior.	**K2**
CO2	Demonstrate the ability to design and fine-tune prompts for specific tasks, optimizing language models for desired outputs	**K3**
CO3	Apply prompt engineering techniques to develop functional AI applications, such as chatbots, content generation tools, and automated systems.	**K3**
CO4	Compare the ethical implications of prompt engineering, addressing challenges such as bias, ambiguity, and misuse, and propose solutions to mitigate these issues.	**K3**

Note*: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create*

CO-PO Mapping Table

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11
CO1	3	2	3								2
CO2	3	2	3								2
CO3	3	2	3								2
CO4	3	2	3								2

Note: *1: Slight (Low), 2: Moderate (Medium), 3: Substantial (High)*

Text Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	Speech and Language Processing	Daniel Jurafsky, James H. Martin	Pearson	2/e, 2013
2	Unlocking the Secrets of Prompt Engineering	Gilbert Mizrahi	Packt	1/e, 2023
3	Prompt Engineering	Ian Khan	Wiley	1/e, 2024

Reference Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	Natural Language Processing with Python	Steven Bird, Ewan Klein, and Edward Loper	Oreilly	1/e, 2009
2	Transformers for Natural Language Processing	Denis Rothman	Packt	1/e, 2021

Video Links (NPTEL, SWAYAM…)	
No.	**Link ID**
1	https://youtu.be/gNOQ-K7QN4s