**Amrita School of Engineering, Amritapuri Campus,**

**Amrita Vishwa Vidyapeetham**

**Department of Computer Science and Engineering**

**Course Plan Jan-July 2025**

**22AIE315 NATURAL LANGUAGE PROCESSING 2-0-2-3**

**S6 BTech. CSE AI**

1. **Course Information**

|  |  |  |  |
| --- | --- | --- | --- |
| **Course Code** | 22AIE315 | **Title** | **NATURAL LANGUAGE PROCESSING** |
| **Academic Year** | 2023-2024 Even Semester | **Semester** | VI |
| **Program** | B.Tech CSE AI 2021-2025 | **L** – **T** – **P** – **C** | 2-0-2-3 |

1. **Course Faculty**: Dr. Veena
2. **Course Objectives**

The main objective of the course is to understand the leading trends and systems in Natural Language Processing. This course will help the students to understand the basic representations used in syntax, the semantics of Natural Language Processing. This course will help the students to understand and explore the models used for word/sentence representations for various NLP applications. This course will help the students to implement deep learning algorithms in Python and learn how to train deep networks for NLP applications.

1. **Course Outcomes (CO)**

|  |  |
| --- | --- |
| **CO#** | **Outcome** |
| CO1 | Apply modern tools for solving problems in computational linguistics |
| CO2 | Implement word representation models to solve NLP problems. |
| CO3 | Develop deep learning models for solving NLP applications |
| CO4 | Evaluate the performance of NLP models |

**CO-PO Affinity Map**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PO | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 | PSO1 | PSO2 | PSO3 |
| CO |
| CO1 | 3 | 3 | 2 | 3 | 3 | 1 | - | 1 | 3 | 3 | - | 3 | 3 | 3 | 3 |
| CO2 | 3 | 3 | 2 | 3 | 3 | 1 | - | 1 | 3 | 3 | - | 3 | 3 | 3 | 3 |
| CO3 | 3 | 3 | 2 | 3 | 3 | 1 | - | 1 | 3 | 3 | - | 3 | 3 | 3 | 3 |
| CO4 | - | - | 1 | 2 | 1 | 1 | - | 1 | 3 | 3 | - | 2 | - | 1 | 1 |

3-strong, 2-moderate, 1-weak

|  |  |  |  |
| --- | --- | --- | --- |
| **CO** | **PO** | **Affinity** | **CO-PO Justification** |
| CO1 | 5 |  | Modern tools are essential for computational linguistics, aligning with engineering tool usage. |
|  |  |  |
| CO2 | 1 |  | Word representation models rely on fundamental engineering and scientific principles. |
| 2 |  | Analyzing NLP problems requires problem-solving skills to derive meaningful solutions. |
|  |  |  |
| CO3 | 3 |  | Developing deep learning models involves designing solutions for real-world NLP challenges. |
| 5 |  | Implementing deep learning models requires modern tools, simulation, and analysis techniques. |
|  |  |  |
| CO4 | 4 |  | Evaluating NLP models requires conducting experiments and analyzing performance metrics. |
| 10 |  | Model evaluation results must be effectively communicated through reports and presentations. |
|  |  |  |
|  |  |  |

**Course Syllabus**

**Unit 1**

Computational linguistics- Introduction, syntax, semantics, morphology, collocation and other NLP problems.

**Unit 2**

Word representation: One-hot encoding, Bag-of-Words (BoW) Dictionary: Term Frequency – Inverse Document Frequency (TF-IDF), Language Model-n-gram – Neural Network-based word embedding algorithms

Unit 3

Sequences and sequential data: Machine learning and deep learning for NLP, Recurrent Neural Network, Long Short-Term Memory networks, Gated Recurrent Unit - Sequence to sequence modelling - Encoder decoder - Attention mechanism, Transformer Networks – BERT, GPT, Graph NLP, Hidden Markov Model, Conditional Random Field, Topic modelling

Unit 4

Applications of NLP: Part-of-Speech tagging, Named Entity recognition, Dependency parsing, - Sentiment Analysis, Machine translation, Question answering, Text summarization, Evaluation metrics for NLP models and Visualization

### **Text Books / References**

Daniel Jurafsky, James H Martin, Speech & language processing, preparation [cited 2020 June 1] Available from: https://web. stanford. edu/~ jurafsky/slp3 (2018).

Christopher Manning and Hinrich Schütze, F oundations of Statistical Natural Language Processing, MIT press, 1999.

Steven Bird, Ewan Klein and Edward Loper, Natural Language Processing with Python, O'Reilly Media, Inc., 2009.

Jason Browlee, Deep Learning for Natural Language Processing: Develop Deep Learning Models for your Natural Language Problems (Ebook), Machine Learning Mastery, 2017

1. **Evaluation Policy 70:30**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Assessment** | **Weightage (%)** |
| **1** | * **Mid Term Examination (50)** | **20** |
| 2 | * **CA (50 marks)**   + Offline Class Test1(**10 marks**)   + Online Quiz (**10 marks**)   + Report Submission : Literature Review &Research Gap Analysis, Functionalities of the Term project (**10 marks**)   + Lab Assignment submission + Lab Quiz (**10 Marks**)   + Report on the Term Project (Paper Format) -(**10 marks**) | 50 |
| 3 | * Term Project   + Phase I : Abstract Submission (Appln domain, Problem statement) - **5 marks**   + Phase 2: Mid Review – Methodology, Block diagram, Novelty (**10 marks)**   + Phase 3: Final Presentation(PPT,Demo,Code Review)- **15 marks** | 30 |

1. **Direct Assessment Tools**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sl. No** | **Direct Assessment Tools** | **Weightage** | **Max Exam Marks** | **CO wise mark distribution** | | | |
| **CO1** | **CO2** | **CO3** | **CO4** |
| 1 | Offline Class Test1 | 10 | 30 |  |  |  |  |
| 2 | Online Quiz | 10 | 20 |  |  |  |  |
| 3 | Report Submission : Literature Gap Functionalities | 10 | 10 |  |  |  |  |
| 4 | Lab Assignment submission + Lab Quiz | 10 | 30 |  |  |  |  |
| 5 | Report on the Term Project | 10 | 10 |  |  |  |  |
| 6 | Midterm | 20 | 50 |  |  |  |  |
| 7 | Term Project Phase I | 5 | 5 |  |  |  |  |
| 8 | Term Project Phase 2 | 10 | 10 |  |  |  |  |
| 9 | Term Project Phase 3 | 15 | 15 |  |  |  |  |

1. **CO Attainment Levels**

60% of the maximum mark (100) is set as threshold for all course outcomes.

|  |  |  |
| --- | --- | --- |
| **CO** | **Threshold %** | **Target %** |
| CO1 | 60 | 70 |
| CO2 | 60 | 70 |
| CO3 | 60 | 70 |
| CO4 | 60 | 70 |

1. **Course Delivery Plan**

**Lecture Plan**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#L** | **Topic** | **Subtopics** | **Objective** | **BTL** |
| **1** | Introduction to NLPApplications | Evolution of NLP, Real-world applications | Understand the significance of modern NLP | Understand (BTL 2) |
| **2-3** | NLTK for NLP Processing | Tokenization, Stemming, Lemmatization, POS tagging, NER, Parsing | Learn text preprocessing techniques for NLP tasks | Apply (BTL 3) |
| **4** | Word Representation Techniques | One-hot encoding, Distributed representations | Compare different word representation methods | Analyze (BTL 4) |
| **5** | TF-IDF | Concepts, calculation | Implement TF-IDF for text ranking | Apply (BTL 3) |
| **6** | Word2Vec & Vector Space Models | CBOW, Skip-gram, Cosine similarity, Word arithmetic | Understand and apply Word2Vec for semantic word representations | Apply (BTL 3) |
| **7-8** | RNNs & LSTMs | RNN architecture, Vanishing gradient problem, LSTM/GRU | Learn sequential modeling techniques for NLP | Understand (BTL 2) |
| **9-11** | Transformer Architecture | Self-attention, Multi-head attention, Positional encoding | Compare RNNs, LSTMs, and Transformers for NLP | Analyze (BTL 4) |
| **12-14** | BERT: Pretraining & Fine-Tuning |  | Understand BERT’s training and fine-tuning processes | Understand (BTL 2) |
| **15-16** | Decoder-Only Models (GPT) | Transformer decoder, GPT vs. BERT, Applications of GPT | Learn about autoregressive models for text generation | Apply (BTL 3) |
| **17-18** | Prompt Engineering & Optimization | Zero-shot, Few-shot learning, Effective prompt strategies | Design and optimize prompts for NLP tasks | Apply (BTL 3) |
| **19-20** | Parameter-Efficient Fine-Tuning (PEFT) | Fine-tuning challenges, LoRA, Efficient adaptation | Explore lightweight fine-tuning approaches for LLMs | Evaluate (BTL 5) |
|  | | | | |
| **21-22** | Fine-tuning | BERT/GPT, Implementing LoRA, Practical applications | Implement AI models in real-world scenarios | Create (BTL 6) |
| **MIDTERM** | | | | |
| **23-24** | LLM Applications | Practical Applications of LLMs : case studies Summarisation, Translation, chatbots in healthcare and Code generation, latest advancements in LLMs.  Comparison: GPT-4, Claude, Mistral, LLaMA, Falcon, Gemma, Deepseek R1 | Learn to apply LLM in different application domain | Apply(BTL3)  Analyze (BTL 4) |
| **Start with Term Project implementation** | | | | |
| **25-26** | Prompt Chainer patterns | Chain of Thought Learning, Tree of Thought Pattern, prompt Chainer Pattern with case study | Learn to use tree of thought for LLM applications | Apply(BTL3) |
|  |  |  |  |  |
| **27-30** | LLM Quantization | Intrinsic SAID LoRA QLoRA: Quantized LoRA LLaMA , |  | Apply(BTL3) |
| **31-34** | RAG | Memory-Augmented LLMs & Retrieval-Augmented Generation (RAG) | Analyze the impact of memory-augmented mechanisms in LLMs and their integration with external knowledge sources for efficient retrieval-based generation. | Apply(BTL3) |
| **35-37** | RLHF | Training LLMs using reinforcement Learning. Reinforcement Learning from Human Feedback (RLHF) |  | Apply(BTL3) |
| **38** | Evaluating LLMS | Evaluating LLMs: Benchmarks, evaluation frameworks and popular leaderboards; |  | Apply(BTL3) |
| **39-40** | Small Language Models | Architectures and Training of SLMs,Model compression techniques for SLMs  Pretraining vs Fine-Tuning in resource-constrained settings | Explain the design principles of SLMs and their computational advantages in low-resource environments. | Understand (BTL 2) |
| **41-42** |  | Multimodal LLMs and Vision-Language Models ( CLIP,BLIP,Gemini) Future Directions and Cutting-Edge Research in LLMs-Continual Learning in LLMs (Meta-Learning,Adaptive LLMs for Personalized AI Assistants,Neuro-Symbolic Hybrid AI (LLM + Logical Reasoning),Conscious AI and Autonomous Agents | Explore the integration of vision and language in multimodal models Explore cutting edge research in LLM and future directions | Understand (BTL 2) |

|  |  |
| --- | --- |
|  | **Lab Sheet Plan**  Lab Sheet #1 – Application of NLTK for various NLP tasks  Lab Sheet #2- Word Vectorization techniques(tf-idf, Word2vec,Fastext)  Lab Sheet #3-Transformer Architecture, BERT Applications  Lab Sheet #4-Using LLMs in Local Devices (GP4ALL & LM Studio) and LLM frameworks  Lab Sheet #5 - Using LLMs via Hugging Face API, Fine-Tuning an LLM using LoRA  Lab Sheet 6 - Retrieval-Augmented Generation (RAG)  Lab Sheet 7 - Building a Chatbot with LangChain  Lab Sheet 8 - Quantizing & Optimizing LLMs for Efficiency  Lab Sheet 9 - Fine-Tuning with QLoRA for Cost-Efficient Training  Lab Sheet 10 - LLM Agents & Tool Usage |

**Faculty Information**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No** | **Faculty Name** | **Class** | **Signature** |
| 1 | Dr Veena | S6CSE |  |