# Course: Applications of NL(X) and LLM

## Instructor: Anand Rao

## (7 weeks)

With the rapid rise in popularity and adoption of Generative AI, the world of NLP has expanded beyond traditional boundaries, offering unprecedented opportunities in business and industry. This comprehensive course is meticulously designed to provide an in-depth understanding of the vast landscape of Natural Language (NL) processing, understanding, generation, reasoning, planning, and optimization (X-for processing, understanding etc.,). Our course covers the evolution of NLP and delves into the intricacies of deep learning, transformer models, and the applications of large language models.

The course introduces the foundational concepts of NL-X, diving deep into techniques for sentiment analysis, named entity recognition, and question answering. As the course progresses, students will explore the world of chatbots, conversational AI, word embeddings, vector databases, and the revolutionary transformer models like BERT. The latter part of the course delves into the capabilities and applications of large language models, emphasizing their role in tasks like retrieval augmented generation and their use as generative agents for reasoning, planning, and optimization.

Several hands-on examples and exercises are integrated throughout the course, offering students practical experience in applying the learned techniques. Guest lectures and practical LLM applications in healthcare and financial services will demonstrate how enterprises are building and using these modern technologies. The course culminates in a final project presentation, allowing students to showcase their mastery of the content. This course is an invaluable resource for those aspiring to delve deep into the world of NL-X and Generative AI, offering practical insights and knowledge that can be immediately applied in the real world. Join us and equip yourself with the skills to navigate the exciting world of Generative AI using NL-X.

## Learning Objectives:
Upon completion of this course, students will be able to:

- **Understand the Evolution and Foundations of NLP:** Students will trace the historical milestones in NLP, appreciate its significance in the modern world, and grasp foundational concepts including neural networks and their architectures.
- **Master Practical NLP Techniques and Tools:** Students will be proficient in text mining techniques, understand popular NLP tools, analyze deep learning architectures (e.g., RNN, LSTM, Transformer) and apply advanced techniques like attention mechanisms to real-world scenarios.
- **Develop and implement NLP algorithms for specific industries:** Students will be able to develop and implement NLP algorithms for specific industries, such as financial services, healthcare, telecommunications, media, technology, retail, and manufacturing, to solve real-world problems.
- **Evaluate the capabilities and limitations of large language models:** Students will be able to evaluate the limitations and capabilities of large language models, including their ability to develop emergent abilities and their potential for propagating errors or biases.
- **Fine-tune and adapt large language models for specific generative AI tasks:** Students will be able to fine-tune and adapt large language models, such as GPT, LLaMA,

PaLM for specific generative AI tasks, including document retrieval, Q&A, language generation, etc.
- **Implement generative agent models for reasoning, planning, and optimization:** Students will be able to implement generative agent models for reasoning, planning, and optimization using large language models and agents to solve specific problems in various application areas, such as social analytics, report generation, chatbots, retrieval augmented generation, document processing, topic modeling, and named entity recognition.

This course requires a basic background in data science and/or Artificial Intelligence. Basic level of Python programming is required for completing the assignments.

# Course Topics

1. Week 1:
   a. Lecture 1: Introduction to NLP: History, Evolution, and Importance
      i. Overview of Natural Language Processing
      ii. Historical milestones in NLP
      iii. Importance and applications of NLP in today's world
   b. Lecture 2: Text Mining and NLP Tools: Practical Techniques and Applications
      i. Introduction to text mining techniques
      ii. Popular tools and libraries for NLP
      iii. Practical applications of text mining in various industries
2. Week 2:
   a. Lecture 3: Deep Learning Foundations for NLP: Neural Networks and Architectures
      i. Basics of neural networks and their relevance to NLP
      ii. Introduction to various neural network architectures
      iii. Deep learning challenges and solutions in NLP
   b. Lecture 4: Advanced Sequence Models: RNNs, LSTMs, GRUs, and Beyond
      i. Introduction to RNNs, LSTMs, and GRUs
      ii. Importance of sequence models in NLP
      iii. Applications beyond Named Entity Recognition (NER)
3. Week 3:
   a. Lecture 5: Introduction to Attention Mechanisms: Basics and Importance in NLP
      i. Fundamentals of attention mechanisms
      ii. Why attention is crucial in modern NLP tasks
      iii. Variants and evolution of attention mechanisms
   b. Lecture 6: Applications of Attention Models: Summarization, Question Answering, and Chatbots
      i. Using attention for text summarization
      ii. Building question-answering systems with attention
      iii. Designing chatbots using attention mechanisms
4. Week 4:
   a. Lecture 7: Generative AI and NLP: Use Cases and Model Pre-training
      i. Introduction to generative AI in the context of NLP: BERT, GPT, LLaMa, and PaLM
      ii. Importance and methods of model pre-training with examples like FinGPT and FinBERT for financial services
      iii. Real-world use cases of generative AI including domain-specific models
   b. Lecture 8: Fine-tuning Techniques: Prompt-engineering, Multi-task Learning, and Challenges
      i. Basics of prompt-engineering for NLP tasks with models like BERT and GPT
      ii. Introduction to multi-task learning in NLP using models like Llama and Palm

iii.   Challenges and solutions in fine-tuning models, with references to domain-specific models like BioBERT and BioMedLM
5.   Week 5:
   a.   Lecture 9: Evaluating and Benchmarking NLP Models: Techniques and Importance
      i.    Importance of model evaluation in NLP
      ii.   Common benchmarks and datasets for evaluation
      iii.  Challenges in evaluating large and complex models
   b.   Lecture 10: Advanced NLP Applications: Retrieval Augmented Generation and Enterprise Use Cases
      i.    Introduction to Retrieval Augmented Generation (RAG)
      ii.   Benefits of RAG in enterprise applications
      iii.  Real-world examples of RAG in action,
6.   Week 6:
   a.   Lecture 11: Reinforcement Learning in NLP: Basics, RLHF, RLAIF, and Practical Applications
      i.    Introduction to reinforcement learning in NLP.
      ii.   Overview of RLHF and RLAIF techniques
      iii.  Practical applications of reinforcement learning in NLP tasks
   b.   Lecture 12: LLM in Practice: Deployment, LLMOps, and Real-world Applications
      i.    Introduction to deploying Large Language Models (LLM) including popular models like BERT, GPT, Llama, and Palm
      ii.   Best practices in LLMOps for efficient deployment of domain-specific models in financial services and biomedicine
      iii.  Real-world applications and case studies of LLM, highlighting models like FinGPT, FinBERT, BioBERT, and BioMedLM
7.   Week 7:
   a.   Lecture 13: Ethics, Risks, and Agents: Understanding the Implications of Advanced NLP
      i.    Ethical considerations in NLP and potential risks
      ii.   Introduction to AI agents and architectures for agents (e.g., BabyAGI, AutoGPT)
      iii.  LLM agents for reasoning, planning, and optimization
   b.   Lecture 14: Course Recap and Future of NLP: Summarizing Key Takeaways and Looking Ahead
      i.    Recap of the major topics covered in the course.
      ii.   Discussion on the future trends in NLP
      iii.  Final project presentation

This course is designed for graduate-level students who have a programming and analytics background, exposed to systems thinking and are proficient in Python programming.

## Assessments
   The final course grade will be calculated using the following categories:

| Assessment | Percentage of Final Grade |
| --- | --- |
| Class Participation | 10% |
| Three Class Quizzes | 30% |
| Individual Assignment | 20% |
| Team Project Presentation | 40% |

| | |
|---|---|
| Total | 100% |

- **Class Participation:** Class participation would be based on (a) Coming prepared to the class having read the pre-reads; (b) Meaningful contributions to the case discussions and insightful questions during the lectures.
- **Individual Assignments:** Individual assignments will be based on answering the discussion questions in the industry-based synthetic scenarios and programming assignments. See note on the use of generative AI tools in the Generative AI Guidance section.
- **Team Project Presentation:** The final project will be a team presentation based on an industry-based synthetic scenario that will be distributed at the start of the course. The students will work as a team during the course and will make the final presentation to a panel of judges. There will be no final exam and the presentation will be conducted during the week of the exams.

Students will be assigned the following final letter grades, based on calculations coming from the course assessment section.

| Grade | Percentage Interval |
|---|---|
| A+ | 98.0-100% |
| A | 92.0-97.9% |
| A- | 90.0-91.9% |
| B+ | 88.0-89.9% |
| B | 82.0-87.9% |
| B- | 80.0-81.9% |
| C+ | 78.0-79.9% |
| C | 72.0-77.9% |
| C- | 70.0-71.9% |
| D | 50.0-69.9% |
| F | 00.0-49.9% |

## Grading Policies

- **Late-work policy**: To encourage timely submissions and ensure fair and prompt grading for all students, assignments should be submitted by 11:59 PM on the due date. For those facing unforeseen circumstances, assignments may be submitted up to 24 hours late for up to 90% of the original grade, with incremental reductions thereafter. No assignments will be marked after 10 days.
- **Make-up work policy**: To maintain the integrity of the grading process while offering flexibility, there will be no make-up assignments or quizzes.
- **Re-grade policy**: To uphold the integrity of the assessment process, regrading will not be

available. However, students are welcome to discuss the rationale for their grades during office hours to gain a better understanding of the assessment.
- **Attendance and/or participation policy**: To emphasize the value of class participation and active engagement in the learning process, attendance is mandatory and will be tracked via a sign-in sheet. Students have the flexibility to miss one class without affecting their class participation grade, as outlined in the Class Participation guidelines.

# Course Policies

- **Academic Integrity & Collaboration**: Students are expected to strictly follow Carnegie Mellon University rules of academic integrity in this course. This means that unless otherwise specified, Individual assignments are to be the work of the individual student using only permitted material and without any cooperation of other students or third parties. It also means that usage of work by others is only permitted in the form of quotations and any such quotation must be distinctively marked to enable identification of the student's own work and own ideas. All external sources used must be properly cited, including author name(s), publication title, year of publication, and a complete reference needed for retrieval. The same work may not be submitted for credit in multiple courses. Violations will be penalized to the full extent mandated by the CMU policies. There will be no exceptions.
- **Use of Generative AI Tools**: We encourage students to explore the use of generative artificial intelligence (AI) tools, such as ChatGPT, for all individual assignments. Any such use must be appropriately acknowledged and cited, following the guidelines established by the APA Style Guide, including the specific version of the tool used. Submitted work should include the exact prompt used to generate the content as well as the AI's full response in an Appendix. Because AI generated content is *not* necessarily accurate or appropriate, it is each student's responsibility to assess the validity and applicability of any generative AI output that is submitted. You may not earn full credit if inaccurate, invalid, or inappropriate information is found in your work. Deviations from these guidelines will be considered violations of CMU's academic integrity policy.
- **Disabilities**: If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.
- **Student wellness**: As a student, you may experience a range of challenges that can interfere with learning, such as strained relationships, increased anxiety, substance use, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may diminish your academic performance and/or reduce your ability to participate in daily activities. CMU services are available, and treatment does work. You can learn more about confidential mental health services available on campus at http://www.cmu.edu/counseling . Support is always available (24/7) from Counseling and Psychological Services: 412-268-2922.
- **Diversity:** It is my intent that students from all diverse backgrounds and perspectives be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength, and benefit. It is my intent to present materials and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups.