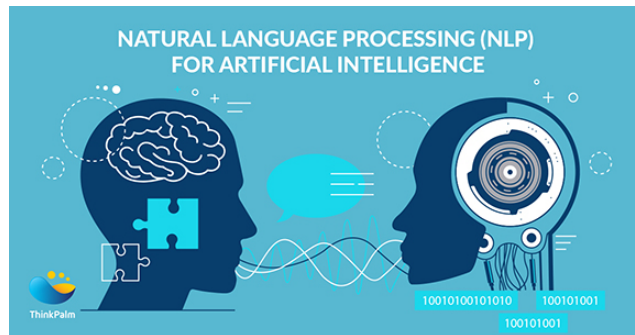


# Introduction to Natural Language Processing

## Description

This course is composed of 12 lectures and will systematically introduce (1) basics and mathematical foundations of Natural Language Processing (NLP), and (2) cutting-edge models for various NLP tasks and their applications. This course will include both theoretical knowledge and practical projects. This course is best suited for students with strong interests on Computer Science and eager to learn more on Artificial Intelligence and Deep Learning.



## What the students will learn:

- (1) To learn fundamentals of NLP: word embeddings, language modeling, topic modeling, Seq2Seq, question answering, text classification, etc.;
- (2) To learn state-of-the-art deep learning models, including convolutional neural network, recurrent neural network, transformer, BERT, etc.;
- (3) To acquire practical coding skills to build cutting-edge NLP models for solving real-world problems;
- (4) To learn and exercise how to perform scientific literature search and how to read and comprehend scientific paper in the most effective way;
- (5) To practice scientific paper writing and potentially how to publish a scientific paper (or research competitions).

## Textbook:

This course does not rely on specific textbook, but the following books are highly recommended to students for a better understanding of the contents:

*Jurafsky, Dan. Speech & language processing. Pearson Education India, 2000.*

*Bird, Steven, Ewan Klein, and Edward Loper. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.", 2009.*

*Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.*

### **Format of the course:**

The lectures will be given online via a Zoom conference of one hour once or twice a week. At the end of each lesson, there will be 15 minutes for questions!

Students will have assignments, most likely essay writing after certain literature search. Rubric are shown in the following sections.

This course is designed for a small group of learners (12-15) who are self-motivated to establish a solid mathematics and AI background.

The instructor believes the following groups of students will benefit from this course:

- 1) Those interested in computer science, AI, deep learning, and NLP;
- 2) Those interested in careers in the computer science or in STEAM;
- 3) Those intending on studying in an American high school or university;
- 4) Those intending to improve their science writing or writing in general;
- 5) Students aged 15-22.

**Instructor:** Di Jin, BS in Precision Instruments from Tsinghua University; Ph.D. in Computer Science from MIT; now Research Scientist at Amazon Alexa AI.

**Course leader:** Howard Shen, Masters of Education with a concentration in Technology from Harvard University in 1997. Mr. Shen was a former staff officer and lecturer at Harvard University. He was the IT lecturer and online course manager of Program for Global Leadership from 1997 to 2001 at Harvard Business School.



### **Course schedule (tentative and can be adjusted with students' interests):**

Classes begins in the mid-Sep, specific times TBD.

**Course 1: Introduction and Word Vectors**

**Course 2: Word Vectors 2 and Word Senses**

**Course 3: Word Window Classification, Neural Networks, and PyTorch**

**Course 4: Matrix Calculus, Loss Function, and Backpropagation**

**Course 5: The probability of a sentence? Recurrent Neural Networks and Language Models**

**Course 6: Vanishing Gradients and Fancy RNNs**

**Course 7: Machine Translation, Seq2Seq and Attention**

**Course 8: Convolutional Neural Network for NLP**

**Course 9: Introduction and Practical Tips for Final Projects**

**Course 10: Question Answering and an introduction to Transformer architectures**

**Course 11: Modeling contexts of use: Contextual Representations and Pretraining. ELMo and BERT.**

**Course 12: Closing remarks and final project presentation**

The tuition for the entire course, including all course materials: \$500

Please send payment via PayPal [hshen@ginsengedu.com](mailto:hshen@ginsengedu.com) or Alipay [caiwu@50edu.cn](mailto:caiwu@50edu.cn)

The following rubric is for optional essay:

Item	1-2	3-5	6-8	9-10
<b>Grammar</b>	Obvious grammatical errors, punctuation mistakes, poor verb conjugation that should have been caught by spellcheck	Basic grammatical errors, punctuation mistakes, poor verb conjugation that should have been caught by proofreading	Minor grammatical errors, punctuation, etc.	Few or no grammatical mistakes; sentences read fluently, punctuation is appropriate and not overused or underused;
<b>Spelling</b>	Multiple obvious spelling errors that should be addressed with spellcheck; spelling errors affect comprehension of text	Basic spelling errors, the spelling does not affect comprehension but it stands in the way of easy reading	Minor spelling errors, confusion of words, minor mistakes when spelling difficult words, or mistakes in conjugation	Few or no spelling errors
<b>Vocabulary</b>	Careless or inaccurate word choice that obscures meaning	Language that is trite, vague, or flat	Shows some variety of word choice, there is purposeful use of words	Effective and engaging use of word choice
<b>Conciseness</b>	Frequent run-ons or fragments, poor punctuation, poor understanding of the material	Writer inserts unnecessary sentences or opinions that are out of place; essay is not focused on topic	Writer uses clear point of view and writes with an understanding of the audience; sentences are of appropriate length for topic	Writer has a strong sense of style and voice and communicates concepts and ideas clearly and succinctly and with style
<b>Organization</b>	Writing is disorganized and underdeveloped with no transitions or closure	Writing is brief and underdeveloped with weak transitions and closure	Uses correct writing format, incorporates a coherent closure to ideas and makes transitions between ideas	Writing includes a strong and clear beginning, middle, and end with appropriate transitions and good closure
<b>Logic</b>	Writing is extremely limited in communicating knowledge with no central theme	Writing is limited in communicating knowledge; length may not be appropriate	Writes related ideas, quality paragraphs, coherent concepts, and does not diverge from topic	Writing is purposeful and focused; ideas contain examples and details

Item	1-2	3-5	6-8	9-10
<b>Data</b>	Writer uses no relevant information from the resources	Writer uses limited data from resources, may not apply concepts appropriately	Writer uses data from resources, applies the data in relevant ways	Writer uses data from resources with a keen understanding of implications and addresses those implications
<b>Beginning</b>	Has no introduction and/or does not have a thesis statement	Has a weak introduction and a weak thesis statement that does not introduce topics discussed in the essay	Has an introduction and a thesis statement that are relevant to the essay	Has an introduction and thesis statement that are both relevant in addition to providing an outline for the contents of the essay
<b>Conclusion</b>	Has no conclusion or conclusion is not related to essay	Conclusion is present but it does not provide a relevant summary of the essay and does not draw conclusions	Conclusion summarizes essay and outlines ideas presented	Conclusion is concise, draws clear, logical ideas and implications from the material included

**Additional remarks:**

Please be sure to check the link for the NYTimes Student Essay Competition, where you will find information on the STEM essay competition. Parents are encouraged to help their student research and formulate the essay.

New York Times Student Contest:

<https://www.nytimes.com/2020/07/15/learning/our-2020-21-student-contest-calendar.html>

Please reference the link above to view the calendar of upcoming events and contests. The STEM writing contest spans from January 19—March 2, 2021. There are other contests that you may wish to investigate.

**Registration:**

Please register with the link: <https://forms.gle/HZgje5LhkdtZm9V79>, or QR code below. You can also register by emailing [hshen@ginsengedu.com](mailto:hshen@ginsengedu.com), or visiting [www.ginsengedu.com](http://www.ginsengedu.com).