



Natural Language Processing

Introduction

2026–27

Ion Androutsopoulos

<http://www.aueb.gr/users/ion/>

Contents

- **What** this course is about.
- **Examples** of applications.
- **Topics** to be covered and **prerequisites**.
- **Organization** of the course:
 - Lectures, labs.
 - Assignments, exams, grading policy.

Large Language Models (LLMs)



Αυτόματη λειτουργία ▾

Νέο σχέδιο  

 Νέα συνομιλία

 Αναζήτηση

 Βιβλιοθήκη

 Διδασκαλία

Παράγοντες

 Μελέτη και εκμάθηση

 Νέος παράγοντας

... Περισσότεροι παράγοντες

Συνομιλίες

You are a top Natural Language...

You are an expert in Natural Lan...

Ανάφερε πέντε εκπληκτικά πρ...

 IOANNIS ANDROUTSOP...
Copilot Chat (Basic) ⓘ

...

You are a top Natural Language Processing (NLP) Instructor. I need to design a graduate NLP course. What should it cover by week?

For a **graduate NLP course in 2026**, I would strongly recommend organizing it around the evolution from **classical NLP** → **neural NLP** → **large language models (LLMs)** → **responsible and efficient NLP systems**. Most students should still learn tokenization, HMMs, CRFs, and parsing, but the emphasis today should clearly be on transformers, prompting, retrieval-augmented generation (RAG), evaluation, and agentic systems.

Course Learning Objectives

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

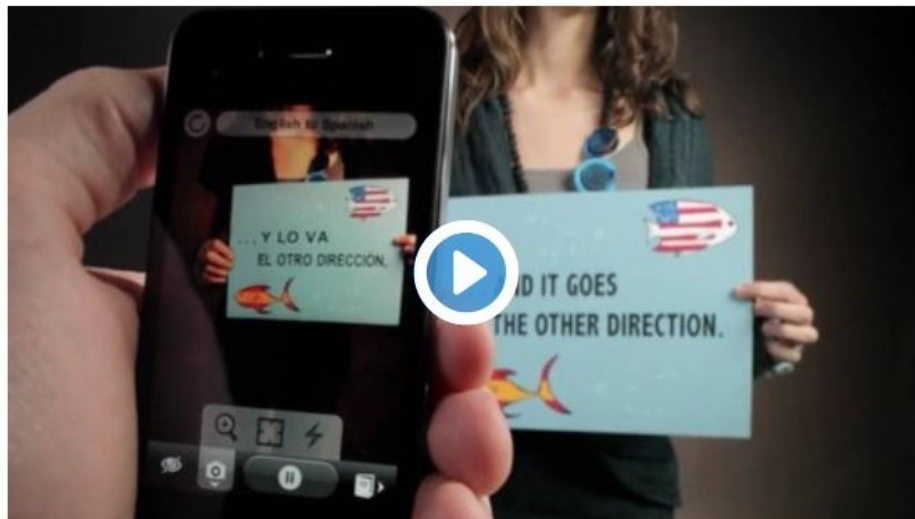
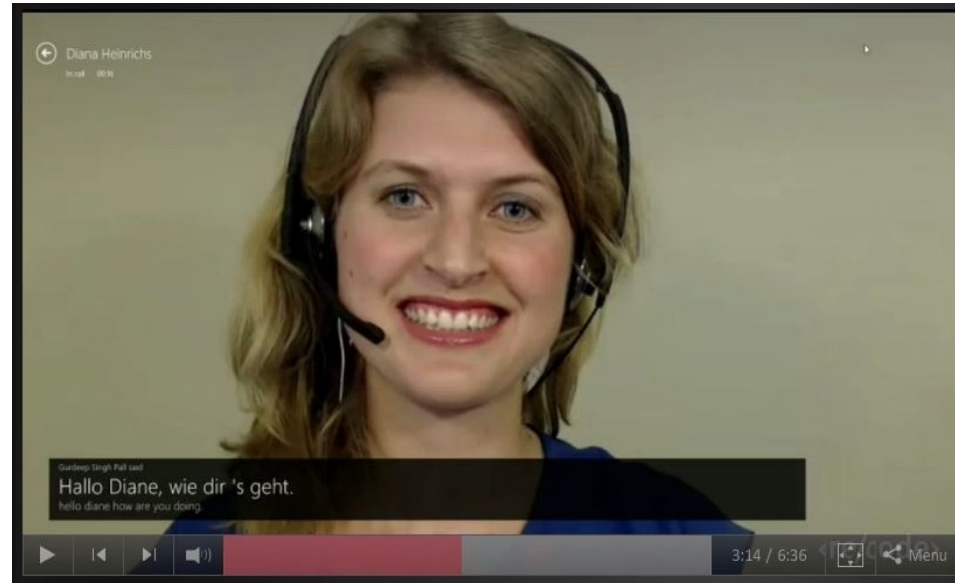
1. Explain major NLP paradigms from symbolic to neural approaches.
2. Build and evaluate NLP systems using modern toolkits.
3. Fine-tune transformer models for downstream tasks.
4. Design and evaluate LLM-based applications.

+ Αποστολή μηνύματος στο Copilot



Το περιεχόμενο που δημιουργείται από AI μπορεί να μην είναι σωστό

Machine translation



<https://www.microsoft.com/en-us/research/video/speech-recognition-breakthrough-for-the-spoken-translated-word-short/>
<https://www.youtube.com/watch?v=RuAp92wW9bg>
<https://www.youtube.com/watch?v=h2OfQdYrHRs>

Opinion mining

Apple Mac mini	GO
money, price, cost, ...	☆☆☆☆☆
ram, memory, ...	☆☆☆
design, color, feeling, ...	☆☆☆☆
extras, keyboard, screen, ...	☆☆

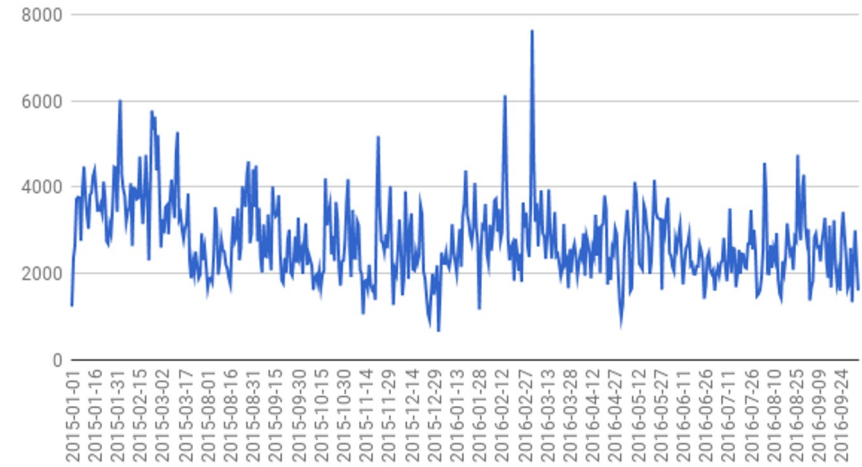
- From **customer reviews** (<http://alt.qcri.org/semeval2016/task5/>).
- From **tweets** (<http://alt.qcri.org/semeval2016/task4/>), **FB updates**, **blogs**, **newspapers**, **radio/TV broadcasts**, ...

User comment moderation

A moderation panel assists the moderators to detect abusive comments, and leads to quicker publication of non-abusive comments.

Highlighting suspicious words using an RNN with self-attention.

Number of comments per day



Moderation Panel

Go	and	hang	yourself	!					85%		
You	are	ignorant	and	vandal	!	Stop	it	!	88%		
Hello	there	try	to	relax					0%		
Thanks	.	Please	go	f#\$@	yourself	.	Ty	!	85%		

J. Pavlopoulos, P. Malakasiotis and I. Androutsopoulos, “Deeper Attention to Abusive User Content Moderation”, EMNLP 2017, <http://nlp.cs.aueb.gr/pubs/emnlp2017.pdf>.

Extracting Contract Elements

THIS AGREEMENT is made the 15th day of October 2009
(The “Effective Date”) BETWEEN:

- (1) **Sugar 13 Inc.**, a corporation whose office is at James House,
42-50 Bond Street, London, EW2H TL (“**Sugar**”);
- (2) **E2 UK Limited**, a limited company whose registered office is
at 260 Bathurst Road, Yorkshire, SL3 4SA (“**Provider**”).

RECITALS:

- A. The Parties wish to enter into a framework agreement which
will enable Sugar, from time to time, to [...]
- B. [...]

NO THEREFORE IT IS AGREED AS FOLLOWS:

ARTICLE I - DEFINITIONS

- “**Sugar**” shall mean: Sugar 13 Inc.
- “**Provider**” shall mean: E2 UK Limited
- “**1933 Act**” shall mean: Securities Act of 1933

ARTICLE II - TERMINATION

The Service Period will be for five (5) years from the Effective Date
(The “**Initial Term**”). The agreement is considered to be terminated
in October 16, 2014.

ARTICLE III - PAYMENT - FEES

During the service period monthly payments should occur. The
estimated fees for the Initial Term are £154,800.

ARTICLE IV - GOVERNING LAW

This agreement shall be governed and construed in accordance with
the Laws of England & Wales. Each party hereby irrevocably
submits to the exclusive jurisdiction of the courts sitting in Northern
London.

IN WITNESS WHEREOF, the parties have caused their
respective duly authorized officers to execute this Agreement.

BY: George Fake
Authorized Officer
Sugar 13 Inc.

BY: Olivier Giroux
CEO
E2 UK LIMITED

Extract start/end dates,
durations, contractors,
amount, legal references,
jurisdiction etc.

- I. Chalkidis, I. Androutsopoulos and A. Michos, “Extracting Contract Elements”, ICAIL 2017,
<http://nlp.cs.aueb.gr/pubs/icail2017.pdf>.
- I. Chalkidis and I. Androutsopoulos, “A Deep Learning Approach to Contract Element
Extraction”, JURIX 2017, <http://nlp.cs.aueb.gr/pubs/jurix2017.pdf>.

Dialogue Systems

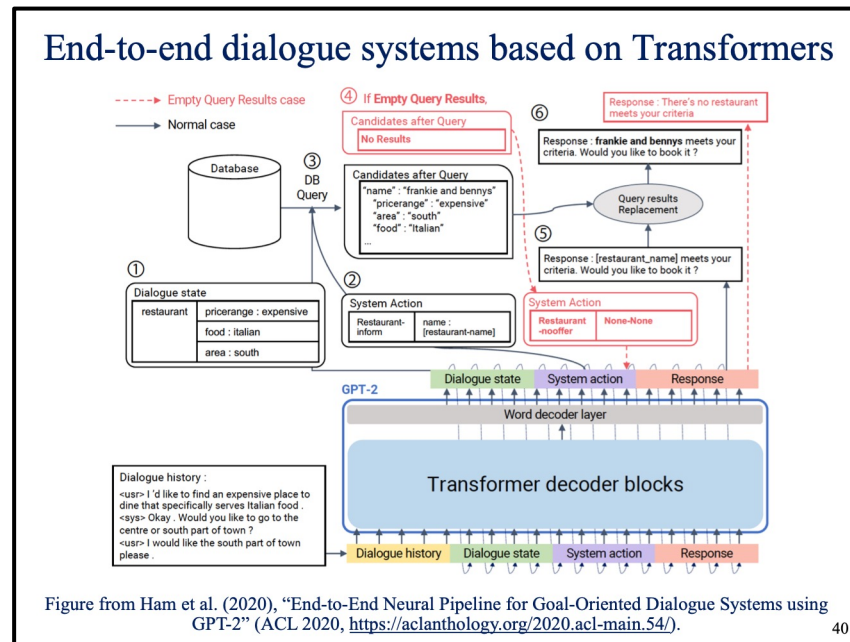
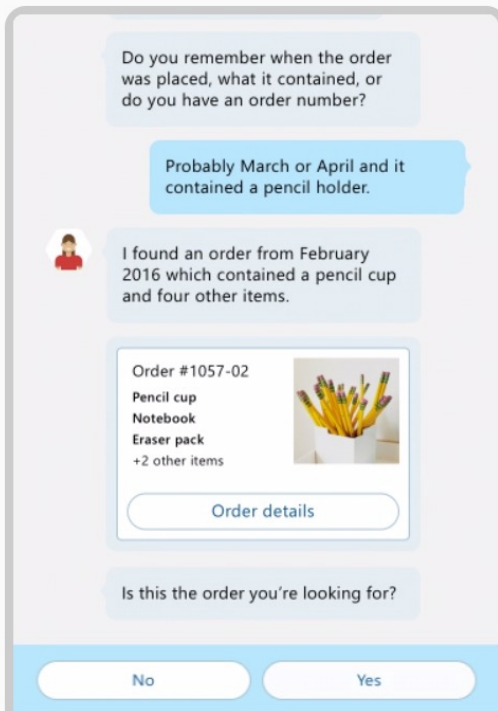
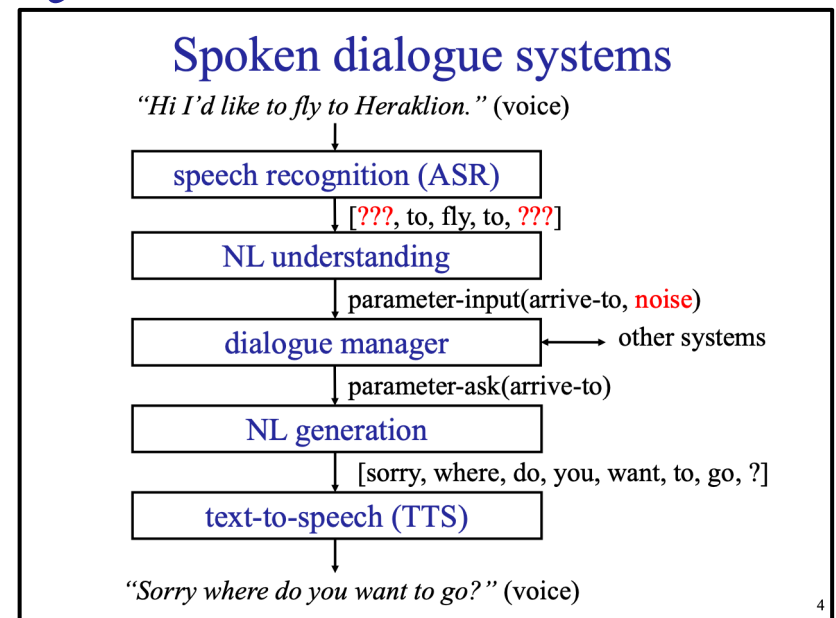
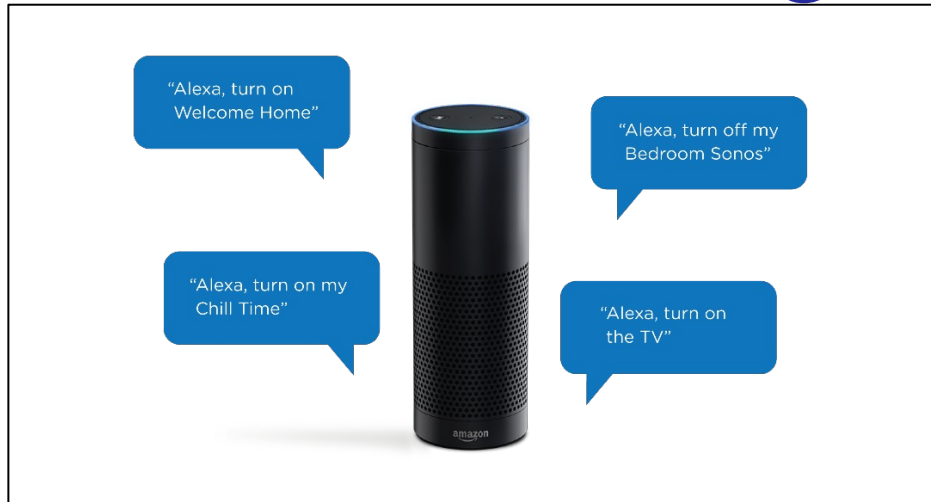
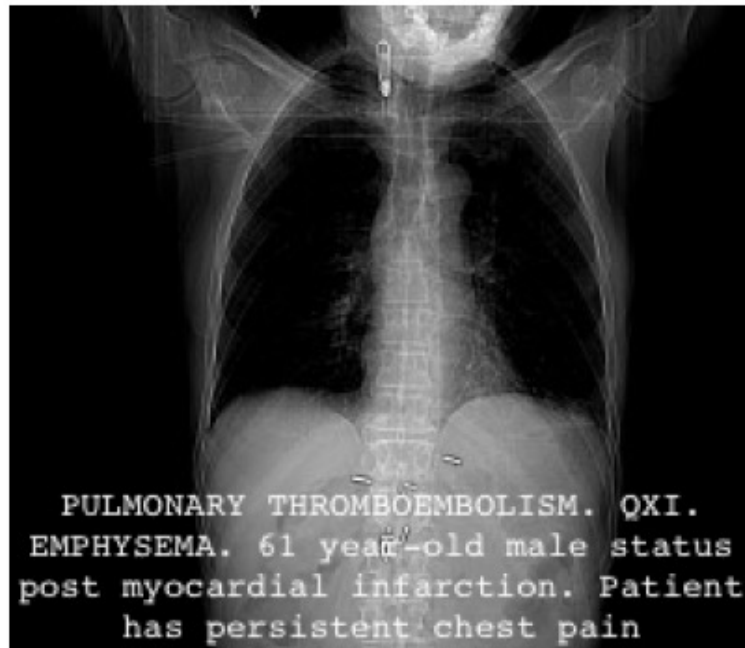


Image captioning

- Image retrieval via captions.
- Eyesight problems.
- Drafting medical reports.



(a) General



(b) Biomedical



Figure 1: Example of a caption produced by the model of Vinyals et al. (2017) for a non-biomedical image (1a) and an example of a PEIR Radiology image with its associated caption (1b).

From I. Pavlopoulos, V. Kougia, I. Androutsopoulos, "A Survey on Biomedical Image Captioning".
<https://www.aclweb.org/anthology/W19-1803/>

Many related terms

- **Natural Language Processing (NLP):**
 - Usually considered a sub-field of **Artificial Intelligence**.
 - **Algorithms, models, systems** to “understand” or generate **mostly written** natural language.
- **Computational Linguistics:**
 - Also mostly about **written** NL, originally more emphasis on computational *linguistic* theories, but now in effect a **synonym** of NLP, and (like NLP) probably closer to CS than linguistics.
 - The *Association for Computational Linguistics* (ACL) organizes the top NLP conferences and publishes the top journals.
- **(Human) Language Technology/Engineering:**
 - Less established term, often **includes speech processing**, more emphasis on **systems/engineering**.
- **Text (and Speech) Analytics:**
 - Business term, usually meaning information extraction, sentiment analysis, opinion mining, getting value from lots of text data.

Topics to be covered (hopefully)

- **W1–2:** Introduction, n-gram LMs, decoding strategies (e.g., beam search, top- k , top- p), application in spelling correction, sub-word tokenizers.
- **W3–4:** Text classification/regression with (mostly) linear models (e.g., logistic regression), stochastic gradient descent (SGD), evaluation measures.
- **W5–6:** Intro to deep learning, text/token classification with MLPs, word embeddings, computation graphs, backpropagation.
- **W7–9 :** NLP with RNNs and CNNs, including text/token classification, attention, encoders-decoders, machine translation, image-to-text.
- **W10–12:** NLP with Transformers, encoder-only models (e.g., BERT), encoder-decoder (e.g., T5, BART), decoder-only models and LLMs (e.g., GPT-x), prompting, fine-tuning, RLHF, reasoning, RAG, tools, agents, ...
- **W13:** Adding speech and vision input to Transformers, dialogs (?).

Prerequisites, recommended courses

- **Prerequisites:**
 - **Calculus** (e.g., derivatives), **linear algebra** (e.g., vector and matrix operations), **probability theory** (e.g., conditional probabilities).
 - **Python** for the programming assignments.
- Other **recommended** companion **courses**:
 - **Deep Learning**. The NLP course is also an introduction to DL.
 - **Reinforcement Learning**. We will cover some RL algorithms that are used in LLMs (e.g., DPO, GRPO), but the RL course covers more extensively the RL area.

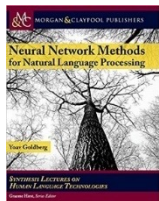
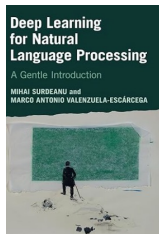
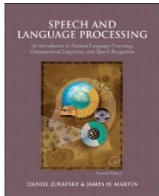
Lectures, demos, assignments

- **13 lectures & demos:**
 - **Lectures** (3 hours/week, key ideas, theory). Slides, study exercises (most solved), references provided.
 - **Labs** (2 hours/week, in ~5 weeks). Related code/tools (e.g., NLTK, spaCy, scikit-learn, PyTorch, HuggingFace, LangChain, LangGraph).
- **6 programming assignments (personal):**
 - Approx. every 2 weeks. Included in the study exercises, but **require code, are handed in, are examined.**
 - The rest of the study exercises are not handed in.
 - (50% if exam ≥ 5 , else 0%)
- **Final exam** (50% if exam ≥ 5 , else 100%).

Recommended books

- **General NLP textbooks:**

- *Speech & Language Processing* by D. Jurafsky and J.H. Martin, 2ⁿ ed., Pearson, 2009. But continuously updated. (Free updated draft: <http://web.stanford.edu/~jurafsky/slp3>.)
- *Deep Learning for Natural Language Processing: A Gentle Introduction*, by M. Surdeanu and M.A. Valenzuela-Escarcega, Cambridge Univ. Press, 2024. (Free draft: <https://clulab.org/gentlenlp/text.html>.)
- *Neural Network Models for Natural Language Processing*, by Y. Goldberg, Morgan & Claypool, 2017. (Available at AUEB's library. Does not cover Transformers, but very well written for MLPs, RNNs, CNNs.)



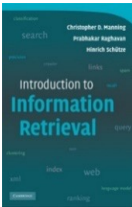
Recommended books

- **Older but still useful NLP/IR textbooks (optional):**

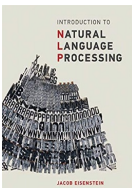
- *Foundations of Statistical Natural Language Processing* by C.D. Manning and H. Schutze, MIT Press, 1999. (Available at AUEB's library.)



- *An Introduction to Information Retrieval* by C.D. Manning, P. Raghavan and H. Schütze, Cambridge University Press, 2008. (Freely available at: <http://nlp.stanford.edu/IR-book/information-retrieval-book.html>.)



- *Introduction to Natural Language Processing* by J. Eisenstein, MIT Press, 2019. (Free draft: <https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>)



Recommended books

- **DL textbooks (optional):**
 - *Deep Learning with Python*, by F. Chollet, Manning Publications, 3rd edition, 2025.
(<https://deeplearningwithpython.io/>)
 - *Understanding Deep Learning*, by S.J.D. Prince, MIT Press, 2023. (Free pre-print and updates:
<https://udlbook.github.io/udlbook/>)
 - *Introduction to Deep Learning*, by E. Charniak, MIT Press, 2019. (Available at AUEB's library.)
 - *Dive into Deep Learning*, by Zhang et al. (Freely available: <https://d2l.ai/>)

