

MODULE 5 : NATURAL LANGUAGE PROCESSING (NLP)

Lesson 1: Why NLP Matters Today

Natural Language Processing (NLP) is no longer just a research topic; it's the backbone of modern technology. Every time you use voice assistants, search engines, chatbots, or real-time translation, NLP is at work.

Think of NLP as teaching computers to become “language partners.” It allows machines not only to process words but also to grasp context, intent, and meaning — enabling smoother interaction between humans and technology.

Key terms: NLP, context, communication

Lesson 2: Text as Data

Unlike numbers in spreadsheets, language is messy and full of ambiguity. NLP transforms raw words into structured formats that machines can analyze.

- **Text Cleaning:** removing noise like punctuation, extra spaces, or emojis.
- **Tokenization:** splitting text into smaller units (words, sentences, or even characters).
- **Vectorization:** turning words into numbers using methods like embeddings.

Imagine trying to teach a robot to cook. Before it can follow the recipe, you must first clean the handwriting, break down steps into smaller instructions, and then encode each step into signals it can understand. That's what NLP does with text.

Key terms: cleaning, tokenization, vectorization

Lesson 3: Sentiment and Intent Detection

One of the most practical uses of NLP is understanding how people feel or what they want.

- **Sentiment Analysis:** determines if a review, tweet, or message is positive, negative, or neutral.
- **Intent Recognition:** figures out what a user is trying to achieve (e.g., “book a flight” vs. “check flight status”).

These tools are crucial in customer service, politics, and social media monitoring. For example, companies use sentiment analysis to detect unhappy customers early and act before problems escalate.

Key terms: sentiment, intent, classification

Lesson 4: NLP in Action — Chatbots and Virtual Assistants

Chatbots and assistants like Siri or Alexa combine multiple NLP tasks: speech recognition, intent detection, and dialogue management. They don't just answer questions — they maintain context across conversations.

Think of them like personal secretaries: they remember your preferences, schedule tasks, and respond naturally to your requests. Good chatbots rely on both rule-based logic and advanced models like transformers.

Key terms: chatbot, virtual assistant, dialogue

Lesson 5: Machine Translation and Multilingual NLP

NLP also powers translation tools that break language barriers. Modern systems use deep learning architectures like Transformers to produce fluent, context-aware translations.

Imagine having a friend who can instantly translate not just words but also tone and cultural context. That's what modern machine translation strives for, making global communication seamless.

Key terms: multilingual, translation, transformer

Lesson 6: Summarization and Information Extraction

Another key task is helping humans digest information faster.

- **Summarization** condenses long reports into key points.
- **Information Extraction** pulls out names, dates, or facts from documents.

Picture it as having an assistant who reads a 50-page report and hands you a 1-page summary with all the critical details. That's the practical power of NLP in business and research.

Key terms: summarization, extraction, knowledge

Lesson 7: NLP and Ethics

As NLP gets stronger, ethical concerns grow. Issues like biased datasets, harmful outputs, and misuse of chatbots for scams remind us that technology is not neutral.

Developers must design systems with **fairness, transparency, and responsibility** in mind. Understanding ethics is just as important as building better models.

Key terms: bias, fairness, responsibility

Lesson 8: Key Takeaways

NLP is about transforming unstructured human language into structured knowledge. From cleaning text to detecting intent, translating languages, and summarizing reports, it touches nearly every digital interaction today.

The future of NLP lies in balancing **innovation** with **ethics**, making systems that are powerful, trustworthy, and accessible to everyone.

Key terms: structure, innovation, ethics

Lesson 9: Text Summarization

Text summarization is the task of condensing a long document into a shorter, more information-dense version while retaining its key points, relationships, and factual accuracy. It is widely used in news aggregation, meeting minutes, legal and medical documents, customer service ticket summaries, and research reviews.

Two main types:

- **Extractive Summarization:** This method selects and concatenates key sentences or fragments from the original text. The advantage is high fidelity and relatively simple implementation; the downside is limited coherence, expressiveness, and compression rate.
- **Abstractive Summarization:** This method rephrases the key points of the original text into new sentences. The advantage is more concise, natural summaries that can integrate information across sentences; the downside is more complex training and control, with the risk of generating “hallucinations” (inaccurate summaries).

Practical workflow (from raw text to usable summaries):

1. **Preprocessing and Segmentation:** Split the text into paragraphs or thematic blocks; for long documents, use chunking with overlap to preserve context.
2. **Importance Modeling:** Evaluate the "information value" of sections using keywords, topic modeling, or attention scores.
3. **Generation and Compression:** Extractive methods focus on sentence ranking and redundancy removal; abstractive methods focus on key point extraction, semantic rewriting, and structured output.
4. **Fact-checking and Deduplication:** Trace entities, numbers, and dates in the summary to ensure accuracy; merge redundant expressions and avoid bias towards summarizing the beginning.
5. **Evaluation and Iteration:** Use automatic metrics (like ROUGE, BERTScore) and human review (fact consistency, coverage, readability, and safety) to optimize.

Tips to reduce "hallucinations":

- **Grounding summaries:** Referencing the source sentences or blocks within the summary, or using question-answering for fact-checking.
- **Controlled decoding:** Limit new entity generation, set length/style constraints, and use a fixed vocabulary list.
- **Layered summarization:** First generate paragraph-level summaries, then summarize them again in a second layer (map-reduce style).

Key terms: extractive, abstractive, ROUGE, factual consistency, chunking, grounding

Lesson Quick Check (MCQs)

Lesson 1 — Why NLP Matters Today

Q: What role does NLP play in modern technology?

- A. It only processes numbers
 - B. It enables machines to understand and respond to human language ☒
 - C. It is mainly used for computer graphics
 - D. It replaces all human teachers
-

Lesson 2 — Text as Data

Q: What is the purpose of tokenization?

- A. To delete all text
 - B. To split text into manageable units ☒
 - C. To translate text into another language
 - D. To check grammar mistakes
-

Lesson 3 — Sentiment and Intent Detection

Q: What is intent recognition used for?

- A. Identifying user goals behind their words ☒
 - B. Translating text
 - C. Generating new music
 - D. Removing punctuation from text
-

Lesson 4 — Chatbots and Assistants

Q: Which of these is a function of chatbots?

- A. Image recognition
 - B. Maintaining context in conversations ☒
 - C. Predicting weather patterns
 - D. Compressing files
-

Lesson 5 — Machine Translation

Q: Which architecture powers modern translation systems like Google Translate?

- A. Rule-based models
 - B. Transformers ☒
 - C. Decision trees
 - D. Linear regression
-

Lesson 6 — Summarization and Information Extraction

Q: What is the goal of summarization in NLP?

- A. To expand short texts into longer versions
 - B. To condense long documents into shorter, informative ones ☒
 - C. To translate documents
 - D. To detect sentiment only
-

Lesson 7 — NLP and Ethics

Q: Why is ethics important in NLP?

- A. Because NLP models can be biased or misused ☒
 - B. Because NLP should replace humans
 - C. Because NLP cannot process numbers
 - D. Because NLP ignores human interaction
-

Lesson 8 — Key Takeaways

Q: What is the overall goal of NLP?

- A. To make computers capable of understanding and interacting with human language ☒
 - B. To train models without any data
 - C. To replace all human teachers
 - D. To only process structured datasets
-

Lesson 9 — Quick Check (MCQ)

Q: Which type of summarization generates new sentences based on the original text?

- A) Extractive Summarization
- B) Abstractive Summarization ☒
- C) Keyword extraction
- D) Text deduplication