
Professional Certificate in AI-Driven Architectural Innovation

Natural Language Processing in Architecture

Artificial Intelligence (AI): a branch of computer science that aims to create systems capable of performing tasks that would typically require human intelligence, such as understanding natural language, recognizing patterns, and making decisions.

Architectural Innovation: the application of new technologies, processes, or methods to improve the design, construction, and operation of buildings and other physical structures.

Natural Language Processing (NLP): a subfield of AI that focuses on the interaction between computers and human (natural) languages. NLP enables machines to understand, interpret, and generate human language in a valuable way.

Machine Learning (ML): a type of AI that allows systems to automatically learn and improve from experience without being explicitly programmed.

Deep Learning (DL): a subset of ML that uses artificial neural networks with many layers to analyze data and make decisions. DL is particularly effective for processing large amounts of unstructured data, such as text, images, and audio.

Named Entity Recognition (NER): a process in NLP that identifies and categorizes key information (entities) in text, such as names of people, organizations, locations, and expressions of times, quantities, and monetary values.

Part-of-Speech (POS) Tagging: the process of identifying the grammatical role of each word in a sentence, such as noun, verb, adjective, or adverb.

Sentiment Analysis: a technique in NLP that determines the emotional tone behind words to gain an understanding of the attitudes, opinions, and emotions expressed within an online mention.

Topic Modeling: a type of statistical model used in NLP to uncover the abstract "topics" that occur in a collection of documents.

Word Embeddings: a type of word representation that allows words with similar meaning to have a similar representation. Word embeddings are often learned from large amounts of text using deep learning models.

Syntax: the arrangement of words to create meaningful sentences in a language.

Semantics: the meaning of words and the way they combine to form meaningful phrases and sentences.

Pragmatics: the branch of linguistics that studies how context influences the interpretation of meaning.

Chatbot: a software application used to conduct an online chat conversation via text or text-to-speech, in lieu of providing direct contact with a live human agent.

Speech Recognition: the ability of a machine or program to identify and respond to human speech.

Text-to-Speech (TTS): a type of speech synthesis that converts written text into spoken words.

Question Answering (QA): a task in NLP that involves automatically answering questions posed by humans in natural language.

Information Extraction (IE): the process of automatically extracting structured information from unstructured text data.

Semantic Role Labeling (SRL): the process of identifying the semantic roles of words or phrases in a sentence, such as agent, patient, and instrument.

Coreference Resolution: the process of identifying when two or more expressions in a text refer to the same entity.

Dialogue Systems: systems that conduct an spoken or textual conversation with a user, allowing for a more natural interaction between humans and machines.

Machine Translation (MT): the process of automatically translating text from one language to another.

Syntax-based Machine Translation: a type of MT that relies on the syntactic structure of sentences to translate text.

Statistical Machine Translation: a type of MT that uses statistical models to translate text based on patterns learned from large amounts of bilingual text.

Neural Machine Translation: a type of MT that uses deep learning models to translate text based on the meaning of sentences, rather than their syntactic structure.

Speech Synthesis: the process of converting written text into spoken words, also known as text-to-speech (TTS).

Natural Language Understanding (NLU): the ability of a machine or program to understand and interpret human language in a meaningful way.

Natural Language Generation (NLG): the process of automatically producing human-like text from a machine or program.

Transfer Learning: a technique in ML where a pre-trained model is used as the starting point for a new task,

allowing for faster and more accurate learning.

Fine-tuning: a type of transfer learning where a pre-trained model is further trained on a new task with a smaller dataset.

Data Augmentation: the process of artificially increasing the size of a training dataset by applying transformations to the existing data.

Named Entity Recognition (NER): a process in NLP that identifies and categorizes key information (entities) in text, such as names of people, organizations, locations, and expressions of times, quantities, and monetary values.

Relation Extraction: the process of identifying semantic relationships between entities mentioned in text.

Dependency Parsing: the process of analyzing the grammatical structure of a sentence to determine the relationships between its words.

constituency Parsing: a type of syntactic parsing that represents sentences as tree structures, where each node represents a constituent, or group of words, that functions as a single unit in the sentence.

Semantic Parsing: the process of converting natural language sentences into formal representations that can be executed by a computer.

Word Sense Disambiguation (WSD): the process of determining the meaning of a word based on its context in a sentence.

Word Embeddings: a type of word representation that allows words with similar meaning to have a similar representation. Word embeddings are often learned from large amounts of text using deep learning models.

Contextual Word Embeddings: a type of word representation that takes into account the context in which a word appears, allowing for a more nuanced representation of word meaning.

Pre-trained Word Embeddings: word embeddings that have been trained on large amounts of text data in advance and can be used as the starting point for other NLP tasks.

GloVe: an unsupervised learning algorithm for obtaining vector representations for words. Global Vectors for Word Representation (GloVe) is a word embedding technique that combines the global statistical information of word co-occurrence probabilities with the local context information of individual words.

Word2Vec: a group of related models that are used to generate numerical representations of words, also known as word embeddings. Word2Vec models are trained on large amounts of text data and can be used for various NLP tasks, such as text classification, named entity recognition, and machine translation.

FastText: an open-source, free, lightweight library that allows users to learn text representations and text classifiers. FastText is a word embedding technique that improves upon Word2Vec by representing each word as an n-gram of characters, rather than a single vector.

BERT: Bidirectional Encoder Representations from Transformers (BERT) is a transformer-based machine learning technique for NLP pre-training developed by Google. BERT is designed to better understand the context of words in a sentence by considering the words that come before and after it.

ELMo: Embeddings from Language Models (ELMo) is a deep contextualized word representation that models both complex characteristics of word use and how these uses vary across linguistic contexts.

UseRUnet: a Russian search engine that uses NLP techniques to understand user queries and provide more relevant results.

Sberbank: a Russian bank that uses NLP techniques for customer service, fraud detection, and risk management.

Yandex: a Russian technology company that provides internet-related products and services, including search, transportation, and e-commerce. Yandex uses NLP techniques for its search engine, chatbots, and voice assistants.

Tinkoff Bank: a Russian bank that uses NLP techniques for customer service, fraud detection, and risk management.

Rosbank: </