

Read AI & Machine Learning Blogs

Read about the latest AWS AI & ML product news and best practices

What is NLP?

Natural language processing (NLP) is a machine learning technology that gives computers the ability to interpret, manipulate, and comprehend human language. Organizations today have large volumes of voice and text data from various communication channels like emails, text messages, social media newsfeeds, video, audio, and more. They use NLP software to automatically process this data, analyze the intent or sentiment in the message, and respond in real time to human communication.

Why is NLP important?

Natural language processing (NLP) is critical to fully and efficiently analyze text and speech data. It can work through the differences in dialects, slang, and grammatical irregularities typical in day-to-day conversations.

Companies use it for several automated tasks, such as to:

- Process, analyze, and archive large documents
- Analyze customer feedback or call center recordings
- Run [chatbots](#) for automated customer service
- Answer who-what-when-where questions
- Classify and extract text

You can also integrate NLP in customer-facing applications to communicate more effectively with customers. For example, a chatbot analyzes and sorts customer queries, responding automatically to common questions and redirecting complex queries to customer support. This automation helps reduce costs, saves agents from spending time on redundant queries, and improves customer satisfaction.

Sensitive data redaction

Businesses in the insurance, legal, and healthcare sectors process, sort, and retrieve large volumes of sensitive documents like medical records, financial data, and private data. Instead of reviewing manually, companies use NLP technology to redact personally identifiable information and protect sensitive data. For example, [Chisel AI](#) helps insurance carriers extract policy numbers, expiration dates, and other personal customer attributes from unstructured documents with [Amazon Comprehend](#).

Customer engagement

NLP technologies allow chat and voice bots to be more human-like when conversing with customers. Businesses use chatbots to scale customer service capability and quality while keeping operational costs to a minimum. [PubNub](#), which builds chatbot software, uses Amazon Comprehend to introduce localized chat functionality for its global customers. [T-Mobile uses NLP](#) to identify specific keywords in customers' text messages and offer personalized recommendations. Oklahoma State University deploys a [Q&A chatbot solution](#) to address student questions using machine learning technology.

Business analytics

Marketers use NLP tools like Amazon Comprehend and [Amazon Lex](#) to gain an educated perception of what customers feel toward a company's product or services. By scanning for specific phrases, they can gauge the customers' moods and emotions in written feedback. For example, [Success KPI](#) provides natural language processing solutions that help businesses focus on targeted areas in sentiment analysis and help contact centers derive actionable insights from call analytics.

How does NLP work?

Natural language processing (NLP) combines computational linguistics, machine learning, and [deep learning](#) models to process human language.

Computational linguistics

Computational linguistics is the science of understanding and constructing human language models with computers and software tools. Researchers use computational linguistics, such as syntactic and semantic analysis, to create frameworks that help machines understand

[Machine learning](#) is a technology that trains a computer with sample data to improve its efficiency. Human language has several features like sarcasm, metaphors, variations in sentence structure, plus grammar and usage exceptions that take humans years to learn. Programmers use machine learning methods to teach NLP applications to recognize and accurately understand these features from the start.

Deep learning

Deep learning is a specific field of machine learning which teaches computers to learn and think like humans. It involves a [neural network](#) that consists of data processing nodes structured to resemble the human brain. With deep learning, computers recognize, classify, and co-relate complex patterns in the input data.

NLP implementation steps

Typically, NLP implementation begins by gathering and preparing unstructured text or speech data from sources like cloud data warehouses, surveys, emails, or internal business process applications.

Pre-processing

The NLP software uses pre-processing techniques such as tokenization, stemming, lemmatization, and stop word removal to prepare the data for various applications.

Here's a description of these techniques:

- Tokenization breaks a sentence into individual units of words or phrases.
- Stemming and lemmatization simplify words into their root form. For example, these processes turn "starting" into "start."
- Stop word removal ensures that words that do not add significant meaning to a sentence, such as "for" and "with," are removed.

Training

Researchers use the pre-processed data and machine learning to train NLP models to perform specific applications based on the provided textual information. Training NLP algorithms requires feeding the software with large data samples to increase the algorithms' accuracy.

Deployment and inference

What are NLP tasks?

Natural language processing (NLP) techniques, or NLP tasks, break down human text or speech into smaller parts that computer programs can easily understand. Common text processing and analyzing capabilities in NLP are given below.

Part-of-speech tagging

This is a process where NLP software tags individual words in a sentence according to contextual usages, such as nouns, verbs, adjectives, or adverbs. It helps the computer understand how words form meaningful relationships with each other.

Word-sense disambiguation

Some words may hold different meanings when used in different scenarios. For example, the word "bat" means different things in these sentences:

- A bat is a nocturnal creature.
- Baseball players use a bat to hit the ball.

With word sense disambiguation, NLP software identifies a word's intended meaning, either by training its language model or referring to dictionary definitions.

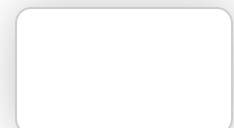
Speech recognition

Speech recognition turns voice data into text. The process involves breaking words into smaller parts and understanding accents, slurs, intonation, and nonstandard grammar usage in everyday conversation. A key application of speech recognition is transcription, which can be done using speech-to-text services like [Amazon Transcribe](#).

Machine translation

Machine translation software uses natural language processing to convert text or speech from one language to another while retaining contextual accuracy. The AWS service that supports machine translation is [Amazon Translate](#).

Named-entity recognition



local cuisines."

The NLP software will pick "Jane" and "France" as the special entities in the sentence. This can be further expanded by co-reference resolution, determining if different words are used to describe the same entity. In the above example, both "Jane" and "she" pointed to the same person.

Sentiment analysis

Sentiment analysis is an artificial intelligence-based approach to interpreting the emotion conveyed by textual data. NLP software analyzes the text for words or phrases that show dissatisfaction, happiness, doubt, regret, and other hidden emotions.

What are the approaches to natural language processing?

We give some common approaches to natural language processing (NLP) below.

Supervised NLP

Supervised NLP methods train the software with a set of labeled or known input and output. The program first processes large volumes of known data and learns how to produce the correct output from any unknown input. For example, companies train NLP tools to categorize documents according to specific labels.

Unsupervised NLP

Unsupervised NLP uses a statistical language model to predict the pattern that occurs when it is fed a non-labeled input. For example, the autocomplete feature in text messaging suggests relevant words that make sense for the sentence by monitoring the user's response.

Natural language understanding

Natural language understanding (NLU) is a subset of NLP that focuses on analyzing the meaning behind sentences. NLU allows the software to find similar meanings in different sentences or to process words that have different meanings.

Natural language generation

How can AWS help with your NLP tasks?

AWS provides the broadest and most complete set of [artificial intelligence](#) and [machine learning](#) (AI/ML) services for customers of all levels of expertise. These services are connected to a comprehensive set of data sources.

For customers that lack ML skills, need faster time to market, or want to add intelligence to an existing process or an application, AWS offers a range of [ML-based language services](#). These allow companies to easily add intelligence to their AI applications through pre-trained APIs for speech, transcription, translation, text analysis, and chatbot functionality.

Here's a list of AWS ML-based language services:

- [Amazon Comprehend](#) helps discover insights and relationships in text
- [Amazon Transcribe](#) performs automatic speech recognition
- [Amazon Translate](#) fluently translates text
- [Amazon Polly](#) turns text into natural-sounding speech
- [Amazon Lex](#) helps build chatbots to engage with customers
- [Amazon Kendra](#) does an intelligent search of enterprise systems to quickly find the content one is looking for

For customers who want to create a standard natural language processing (NLP) solution across their business, consider [Amazon SageMaker](#). SageMaker makes it easy to prepare data and build, train, and deploy ML models for any use case with fully managed infrastructure, tools, and workflows, including no-code offerings for business analysts.

With [Hugging Face on Amazon SageMaker](#), you can deploy and fine-tune pre-trained models from Hugging Face, an open-source provider of NLP models known as Transformers. This reduces the time it takes to set up and use these NLP models from weeks to minutes.

Get started with NLP by creating an [AWS account](#) today.

AWS Natural Language Processing next steps