Natural Language Processing (NLP)

Natural language processing strives to build machines that understand and respond to text or voice data—and respond with text or speech of their own—in much the same way humans do.

What is natural language processing?

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of [artificial intelligence or AI](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence)—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

Artificial Intelligence (AI)

Artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind.

What is artificial intelligence?

While a number of definitions of artificial intelligence (AI) have surfaced over the last few decades, John McCarthy offers the following definition in this 2004 [paper](https://homes.di.unimi.it/borghese/Teaching/AdvancedIntelligentSystems/Old/IntelligentSystems_2008_2009/Old/IntelligentSystems_2005_2006/Documents/Symbolic/04_McCarthy_whatisai.pdf) (PDF, 106 KB) (link resides outside IBM), " It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."

However, decades before this definition, the birth of the artificial intelligence conversation was denoted by Alan Turing's seminal work, "[Computing Machinery and Intelligence](https://www.csee.umbc.edu/courses/471/papers/turing.pdf)" (PDF, 89.8 KB) (link resides outside of IBM), which was published in 1950. In this paper, Turing, often referred to as the "father of computer science", asks the following question, "Can machines think?"  From there, he offers a test, now famously known as the "Turing Test", where a human interrogator would try to distinguish between a computer and human text response. While this test has undergone much scrutiny since its publish, it remains an important part of the history of AI as well as an ongoing concept within philosophy as it utilizes ideas around linguistics.

Stuart Russell and Peter Norvig then proceeded to publish, [Artificial Intelligence: A Modern Approach](http://aima.cs.berkeley.edu/) (link resides outside IBM), becoming one of the leading textbooks in the study of AI. In it, they delve into four potential goals or definitions of AI, which differentiates computer systems on the basis of rationality and thinking vs. acting:

**Human approach:**

* Systems that think like humans
* Systems that act like humans

**Ideal approach:**

* Systems that think rationally
* Systems that act rationally

Alan Turing’s definition would have fallen under the category of “systems that act like humans.”

At its simplest form, artificial intelligence is a field, which combines computer science and robust datasets, to enable problem-solving. It also encompasses sub-fields of machine learning and deep learning, which are frequently mentioned in conjunction with artificial intelligence. These disciplines are comprised of AI algorithms which seek to create expert systems which make predictions or classifications based on input data.

Today, a lot of hype still surrounds AI development, which is expected of any new emerging technology in the market. As noted in [Gartner’s hype cycle](https://www.gartner.com/en/documents/3887767/understanding-gartner-s-hype-cycles) (link resides outside IBM), product innovations like, self-driving cars and personal assistants, follow “a typical progression of innovation, from overenthusiasm through a period of disillusionment to an eventual understanding of the innovation’s relevance and role in a market or domain.” As Lex Fridman notes [here](https://www.youtube.com/watch?v=O5xeyoRL95U) (01:08:15) (link resides outside IBM) in his MIT lecture in 2019, we are at the peak of inflated expectations, approaching the trough of disillusionment.

As conversations emerge around the ethics of AI, we can begin to see the initial glimpses of the trough of disillusionment. To read more on where IBM stands within the conversation around [AI ethics](https://www.ibm.com/cloud/learn/ai-ethics), read more [here](https://www.ibm.com/artificial-intelligence/ethics).

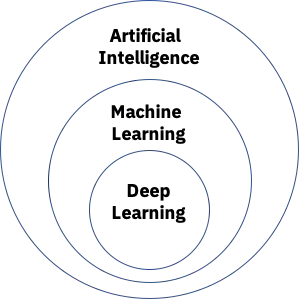
Types of artificial intelligence—weak AI vs. strong AI

Weak AI—also called Narrow AI or Artificial Narrow Intelligence (ANI)—is AI trained and focused to perform specific tasks. Weak AI drives most of the AI that surrounds us today. ‘Narrow’ might be a more accurate descriptor for this type of AI as it is anything but weak; it enables some very robust applications, such as Apple's Siri, Amazon's Alexa, IBM Watson, and autonomous vehicles.

Strong AI is made up of Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI). Artificial general intelligence (AGI), or general AI, is a theoretical form of AI where a machine would have an intelligence equaled to humans; it would have a self-aware consciousness that has the ability to solve problems, learn, and plan for the future. Artificial Super Intelligence (ASI)—also known as superintelligence—would surpass the intelligence and ability of the human brain. While strong AI is still entirely theoretical with no practical examples in use today, that doesn't mean AI researchers aren't also exploring its development. In the meantime, the best examples of ASI might be from science fiction, such as HAL, the superhuman, rogue computer assistant in *2001: A Space Odyssey.*

Deep learning vs. machine learning

Since deep learning and machine learning tend to be used interchangeably, it’s worth noting the nuances between the two. As mentioned above, both deep learning and machine learning are sub-fields of artificial intelligence, and deep learning is actually a sub-field of machine learning.



Deep learning is actually comprised of neural networks. “Deep” in deep learning refers to a neural network comprised of more than three layers—which would be inclusive of the inputs and the output—can be considered a deep learning algorithm. This is generally represented using the following diagram:



The way in which deep learning and machine learning differ is in how each algorithm learns. Deep learning automates much of the feature extraction piece of the process, eliminating some of the manual human intervention required and enabling the use of larger data sets. You can think of deep learning as "scalable machine learning" as Lex Fridman noted in same MIT lecture from above. Classical, or "non-deep", machine learning is more dependent on human intervention to learn. Human experts determine the hierarchy of features to understand the differences between data inputs, usually requiring more structured data to learn.

"Deep" machine learning can leverage labeled datasets, also known as supervised learning, to inform its algorithm, but it doesn’t necessarily require a labeled dataset. It can ingest unstructured data in its raw form (e.g. text, images), and it can automatically determine the hierarchy of features which distinguish different categories of data from one another. Unlike machine learning, it doesn't require human intervention to process data, allowing us to scale machine learning in more interesting ways.

Artificial intelligence applications

There are numerous, real-world applications of AI systems today. Below are some of the most common examples:

* **Speech recognition:** It is also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, and it is a capability which uses natural language processing (NLP) to process human speech into a written format. Many mobile devices incorporate speech recognition into their systems to conduct voice search—e.g. Siri—or provide more accessibility around texting.
* **Customer service:**  Online [virtual agents](https://www.ibm.com/products/watson-assistant) are replacing human agents along the customer journey. They answer frequently asked questions (FAQs) around topics, like shipping, or provide personalized advice, cross-selling products or suggesting sizes for users, changing the way we think about customer engagement across websites and social media platforms. Examples include messaging bots on e-commerce sites with [virtual agents](https://www.ibm.com/products/watson-assistant), messaging apps, such as Slack and Facebook Messenger, and tasks usually done by virtual assistants and [voice](https://www.ibm.com/products/watson-assistant/integrations/voice) assistants.
* **Computer vision:** This AI technology enables computers and systems to derive meaningful information from digital images, videos and other visual inputs, and based on those inputs, it can take action. This ability to provide recommendations distinguishes it from image recognition tasks. Powered by convolutional neural networks, computer vision has applications within photo tagging in social media, radiology imaging in healthcare, and self-driving cars within the automotive industry.
* **Recommendation engines:** Using past consumption behavior data, AI algorithms can help to discover data trends that can be used to develop more effective cross-selling strategies. This is used to make relevant add-on recommendations to customers during the checkout process for online retailers.
* **Automated stock trading:**Designed to optimize stock portfolios, AI-driven high-frequency trading platforms make thousands or even millions of trades per day without human intervention.

NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to ‘understand’ its full meaning, complete with the speaker or writer’s intent and sentiment.

NLP drives computer programs that translate text from one language to another, respond to spoken commands, and summarize large volumes of text rapidly—even in real time. There’s a good chance you’ve interacted with NLP in the form of voice-operated GPS systems, digital assistants, speech-to-text dictation software, customer service chatbots, and other consumer conveniences. But NLP also plays a growing role in enterprise solutions that help streamline business operations, increase employee productivity, and simplify mission-critical business processes.

NLP tasks

Human language is filled with ambiguities that make it incredibly difficult to write software that accurately determines the intended meaning of text or voice data. Homonyms, homophones, sarcasm, idioms, metaphors, grammar and usage exceptions, variations in sentence structure—these just a few of the irregularities of human language that take humans years to learn, but that programmers must teach natural language-driven applications to recognize and understand accurately from the start, if those applications are going to be useful.

Several NLP tasks break down human text and voice data in ways that help the computer make sense of what it's ingesting. Some of these tasks include the following:

* **Speech recognition**, also called speech-to-text, is the task of reliably converting voice data into text data. Speech recognition is required for any application that follows voice commands or answers spoken questions. What makes speech recognition especially challenging is the way people talk—quickly, slurring words together, with varying emphasis and intonation, in different accents, and often using incorrect grammar.
* **Part of speech tagging**, also called grammatical tagging, is the process of determining the part of speech of a particular word or piece of text based on its use and context. Part of speech identifies ‘make’ as a verb in ‘I can make a paper plane,’ and as a noun in ‘What make of car do you own?’
* **Word sense disambiguation** is the selection of the meaning of a word with multiple meanings  through a process of semantic analysis that determine the word that makes the most sense in the given context. For example, word sense disambiguation helps distinguish the meaning of the verb 'make' in ‘make the grade’ (achieve) vs. ‘make a bet’ (place).
* **Named entity recognition,**or NEM, identifies words or phrases as useful entities. NEM identifies ‘Kentucky’ as a location or ‘Fred’ as a man's name.
* **Co-reference resolution** is the task of identifying if and when two words refer to the same entity. The most common example is determining the person or object to which a certain pronoun refers (e.g., ‘she’ = ‘Mary’),  but it can also involve identifying a metaphor or an idiom in the text  (e.g., an instance in which 'bear' isn't an animal but a large hairy person).
* **Sentiment analysis**attempts to extract subjective qualities—attitudes, emotions, sarcasm, confusion, suspicion—from text.
* **Natural language generation**is sometimes described as the opposite of speech recognition or speech-to-text; it's the task of putting structured information into human language.

See the blog post “[NLP vs. NLU vs. NLG: the differences between three natural language processing concepts](https://www.ibm.com/blogs/watson/2020/11/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-language-processing-concepts/)” for a deeper look into how these concepts relate.

NLP tools and approaches

Python and the Natural Language Toolkit (NLTK)

The Python programing language provides a wide range of tools and libraries for attacking specific NLP tasks. Many of these are found in the Natural Language Toolkit, or NLTK, an open source collection of libraries, programs, and education resources for building NLP programs.

The NLTK includes libraries for many of the NLP tasks listed above, plus libraries for subtasks, such as sentence parsing, word segmentation, stemming and lemmatization (methods of trimming words down to their roots), and tokenization (for breaking phrases, sentences, paragraphs and passages into tokens that help the computer better understand the text). It also includes libraries for implementing capabilities such as semantic reasoning, the ability to reach logical conclusions based on facts extracted from text.

Statistical NLP, machine learning, and deep learning

The earliest NLP applications were hand-coded, rules-based systems that could perform certain NLP tasks, but couldn't easily scale to accommodate a seemingly endless stream of exceptions or the increasing volumes of text and voice data.

Enter statistical NLP, which combines computer algorithms with machine learning and [deep learning](https://www.ibm.com/cloud/learn/deep-learning) models to automatically extract, classify, and label elements of text and voice data and then assign a statistical likelihood to each possible meaning of those elements. Today, deep learning models and learning techniques based on convolutional neural networks (CNNs) and recurrent neural networks (RNNs) enable NLP systems that 'learn' as they work and extract ever more accurate meaning from huge volumes of raw, unstructured, and unlabeled text and voice data sets.

For a deeper dive into the nuances between these technologies and their learning approaches, see “[AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What’s the Difference?](https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks)”

NLP use cases

Natural language processing is the driving force behind machine intelligence in many modern real-world applications. Here are a few examples:

* **Spam detection:**You may not think of spam detection as an NLP solution, but the best spam detection technologies use NLP's text classification capabilities to scan emails for language that often indicates spam or phishing. These indicators can include overuse of financial terms, characteristic bad grammar, threatening language, inappropriate urgency, misspelled company names, and more. Spam detection is one of a handful of NLP problems that experts consider 'mostly solved' (although you may argue that this doesn’t match your email experience).
* **Machine translation:**Google Translate is an example of widely available NLP technology at work. Truly useful machine translation involves more than replacing words in one language with words of another.  Effective translation has to capture accurately the meaning and tone of the input language and translate it to text with the same meaning and desired impact in the output language. Machine translation tools are making good progress in terms of accuracy. A great way to test any machine translation tool is to translate text to one language and then back to the original. An oft-cited classic example: Not long ago, translating “*The spirit is willing but the flesh is weak”* from English to Russian and back yielded “*The vodka is good but the meat is rotten*.” Today, the result is “*The spirit desires, but the flesh is weak*,” which isn’t perfect, but inspires much more confidence in the English-to-Russian translation.
* **Virtual agents and chatbots:** [Virtual agents](https://www.ibm.com/products/watson-assistant) such as Apple's Siri and Amazon's Alexa use speech recognition to recognize patterns in voice commands and natural language generation to respond with appropriate action or helpful comments. [Chatbots](https://www.ibm.com/cloud/learn/chatbots-explained) perform the same magic in response to typed text entries. The best of these also learn to recognize contextual clues about human requests and use them to provide even better responses or options over time. The next enhancement for these applications is question answering, the ability to respond to our questions—anticipated or not—with relevant and helpful answers in their own words.
* **Social media sentiment analysis:**NLP has become an essential business tool for uncovering hidden data insights from social media channels. Sentiment analysis can analyze language used in social media posts, responses, reviews, and more to extract attitudes and emotions in response to products, promotions, and events–information companies can use in product designs, advertising campaigns, and more.
* **Text summarization:**Text summarization uses NLP techniques to digest huge volumes of digital text and create summaries and synopses for indexes, research databases, or busy readers who don't have time to read full text. The best text summarization applications use semantic reasoning and natural language generation (NLG) to add useful context and conclusions to summaries.