

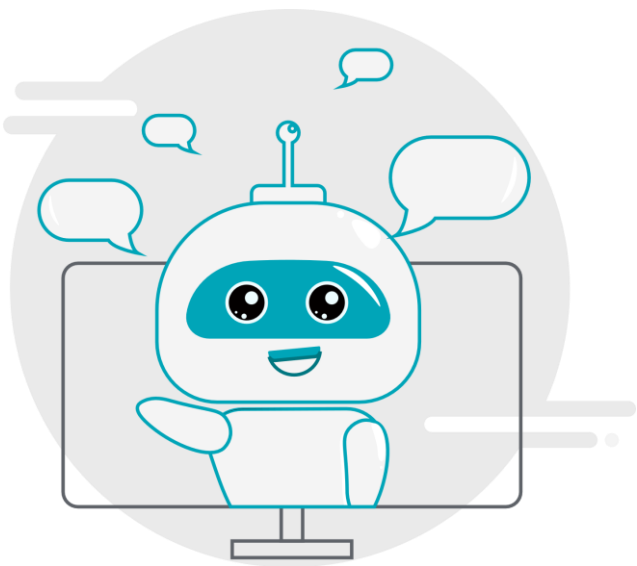
## Natural language processing

Rapid advances in natural language processing technology are opening the door to possibilities that seem like science fiction just a short while ago. We can speak to devices and they understand us and respond back; translation software allow us to cross language barriers; and the technology holds promise in helping to make the world more trustworthy, resilient, and secure.

### What is natural language processing (NLP)?

NLP is a domain of **artificial intelligence** (AI) that enables computer systems to process, analyze, and generate **human language** to facilitate interaction with humans through natural language text or speech<sup>1</sup>.

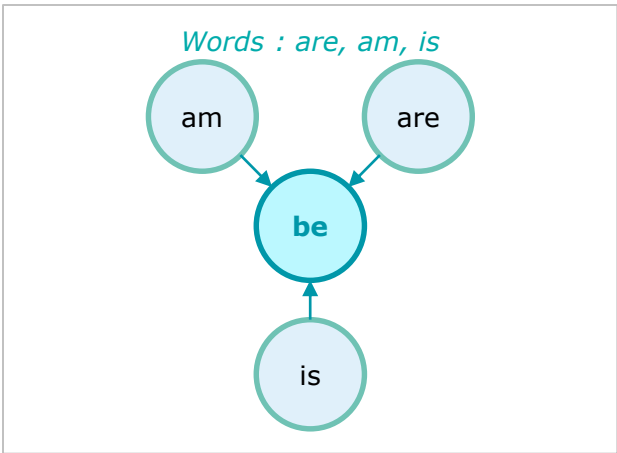
An important and challenging subset of NLP is **natural language understanding** (NLU). While both interpret human language, NLU **goes beyond recognizing words and interprets their underlying meaning**.



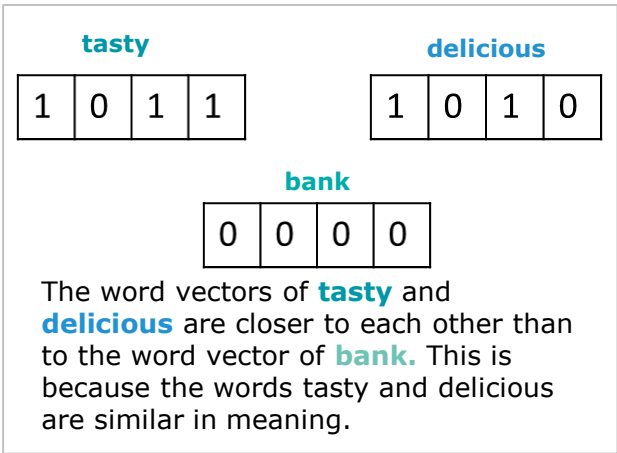
### NLP techniques

NLP uses three broad categories of techniques to process language. Most applications of NLP leverage a combination of techniques.

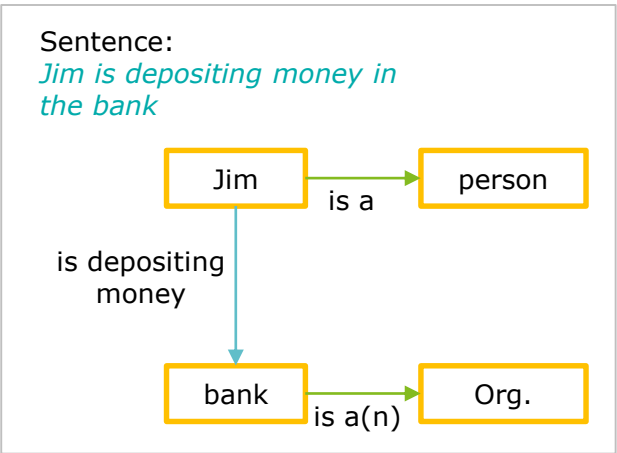
#### Parse text to identify root forms of words



#### Derive the meaning of words by converting them to their numerical vectors



#### Identify relationships between words and phrases to deduce context










### Capabilities powered by NLP

The capabilities listed below are presented in increasing order of complexity, and can be used in conjunction with each other to build solutions for specific use cases, such as chatbots and intelligent e-discovery systems.




*With cloud vendors investing in automated machine learning for NLP, sophisticated NLP models can be built by professionals with limited programming expertise.*

Capability	Description	Example
 <b>Entity extraction</b>	Classifies entities that are present in a piece of text into predefined categories like “individuals”, “places”, “organization”, “cities”, “dates”, etc.	News media and publishing houses use entity extraction to extract the most important entities from a news article to reveal the major people, organizations, and places featured
 <b>Sentiment analysis</b>	Identifies and categorizes opinions expressed in text or speech, especially in order to determine whether the person’s attitude toward a particular topic, product, etc., is positive, negative, or neutral.	Sentiment analysis can be used in customer care to parse through tickets and identify the order of execution based on priority of the issue mentioned
 <b>Language detection and translation</b>	Identifies the language of the text or speech content and translates it to a target language	Language detection and translation tools can be leveraged by software companies to document product information and frequently asked questions in multiple languages suited to different markets
 <b>Topic or document classification</b>	Discovers the abstract “topics” that documents contain to classify documents with similar content	Law firms can use topic classification to automatically classify contracts into classes, such as lease agreements, royalty contracts, and service agreements.
 <b>Natural language querying</b>	Uses NLP and deep learning to index available data and build a comprehensive knowledge base to find relevant information in real time	Natural language querying can be used to build enterprise chatbots for employee query resolution. The chatbot can sift through product documentations, discussion forums, etc., to find answers
 <b>Language generation*</b>	Creates text narratives from structured data using tools with rules-based linguistic capabilities, enabling translation of data in spreadsheets to written narratives	Language generation can be used to write investor reports on the performance of mutual funds, as well as generate individual client financial portfolio summaries based on existing data
 <b>Text summarization</b>	Generates insights from a long piece of text, with the aim of creating a coherent and fluent summary of key points outlined in the document	Risk sensing applications can use automatic text summarization to parse through millions of articles and summarize them to generate insights about potential risks

<sup>1</sup>Natural language generation (NLG), or the use of AI to produce a written or spoken narrative from structured data, is sometimes considered to be separate from NLP, which is viewed as the ability of a machine to read and process natural language. Others consider NLG to be a capability under NLP. For completeness, we have included NLG in this discussion.


Opportunities to leverage NLP for risk management

NLP can be leveraged to prevent, manage and mitigate risks across different domains




Internal audit and assurance

**Document review of audit evidences**  
Automating the digitization and retrieval of key facts and figures from unstructured evidences gathered for controls testing using NLP. This can reduce manual effort and increase accuracy in sifting through thousands of documents.




Financial risk and transaction restructuring

**Automated lease abstraction**  
Using NLP to automate the process of lease abstraction to produce a summary of specific and crucial information from an extensive lease document. This makes useful data easily accessible for enhanced review of key points and values.



Risk intelligence


**Brand and reputation management**  
Providing near real-time insights on key issues and customer concerns for a company by leveraging NLP on data collected from a variety of sources, such as surveys, social media, and enterprise databases.



Forensic

**Third-party due diligence**  
Filtering unstructured data from multiple media sources and databases to identify suspicious people or organizations. NLP can help screen entities against thousands of sanctions, embargos, politically-exposed persons, and government watch lists through entity and relationship extraction from aggregated information.


**Fraud investigation**  
Natural language querying can be used to analyze communication patterns, financial information, and corporate documents to identify relevant evidence and reduce the cost and time required for investigations. Companies can also use NLP to proactively identify potential fraud.



Cyber risk

**Threat detection**  
Using NLP with pattern detection to proactively monitor and identify threats such as phishing and malicious domains, and also conduct source code vulnerability analysis and malware analysis.

**Data protection**  
Automating monitoring and detection of sensitive data, such as personally identifiable information, that requires extra protection, along with real-time alert generation using NLP and pattern recognition.




Regulatory and operational Risk

**Operational risk policy guidance**  
Providing answers to queries on operational risk and compliance policies and procedures using NLP (e.g., "What should I do if I have lost my laptop?"), which can lead to faster resolution and higher compliance.

**Regulatory requirements and controls management**  
Identifying and extracting regulatory changes and updating the enterprise regulatory compliance requirements using NLP, thus reducing time, labor, and costs of managing regulatory compliance.

Vectors of progress for NLP

With the immense amount of investment into NLP research and development, it is likely that NLP technology will continue to become more sophisticated over time, especially in the following domains:



**Rise of explainable AI:** With advancements in NLP, many researchers are actively investing effort in building self-explainable NLP models to address the problem of opaque algorithms, and provide explanation of what aspects of source text was used to come to the conclusion provided.

**Advancement in natural language generation:** NLG is moving beyond rule-based text generation, with computers gaining the ability to mine huge amount of quantitative and qualitative data from diverse sources, gain insights, and present them in a way that is easily understood by humans.




**Enabling cognitive search:** Cognitive search is an advancement of natural language querying powered by deep learning, enabling enterprises to powerfully index data from a deep corpus of knowledge base to feed relevant information to users in real time.


**Advancement in emotion detection:** Sentiment analysis is beginning to move from broad categories of sentiments (e.g., positive/negative/neutral) to identifying nuanced human emotions.

**Advancement in natural language understanding:** In the future, NLU will likely be able to develop a increasingly human-like contextual and semantic understanding of speech and text, which can be useful across various applications.

Potential risk associated with NLP

Although NLP offers significant capabilities for smarter and more efficient risk management, it can introduce novel risks to the organization. Below are some risk considerations along with potential mitigation approaches.

Potential risks	Mitigation approaches
 <div>Heightened cyber risks: Business operations can be crippled by cyber attacks if the attackers can re-train the NLP-enabled tools (such as chatbots and contract extraction tools) for malicious use.</div>	Implement tamper detection controls and validate NLP algorithm results as part of ongoing monitoring activities
 <div>Privacy and data confidentiality implications: Smart assistants and smart speakers raise a number of confidentiality and privacy concerns due to their ability to record anything being spoken around them, without the awareness or approval of those being recorded.</div>	Protecting sensitive data via data anonymization and governance on usage of data for downstream processing and analysis.
 <div>Critical implications of training bias: NLP tools are powered by machine learning algorithms trained on data. Using biased data may result in ineffective and inaccurate models leading to poor performance and quality of NLP applications.</div>	Thorough evaluation of the training data and algorithm outputs to identify potential biases should be a key step in the development of NLP applications.



Key contacts

**Nancy Albinson**  
Managing Director  
Deloitte Risk & Financial Advisory  
Deloitte & Touche LLP  
[nalbinson@deloitte.com](mailto:nalbinson@deloitte.com)

**Yang Chu**  
Senior Manager  
Deloitte Risk & Financial Advisory  
Deloitte & Touche LLP  
[yangchu@deloitte.com](mailto:yangchu@deloitte.com)

This document contains general information only and Deloitte Risk & Financial Advisory is not, by means of this document, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This document is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor. Deloitte Risk & Financial Advisory shall not be responsible for any loss sustained by any person who relies on this document.

As used in this document, "Deloitte Risk & Financial Advisory" means Deloitte & Touche LLP, which provides audit, assurance, and risk and financial advisory services; Deloitte Financial Advisory Services LLP, which provides forensic, dispute, and other consulting services; and its affiliate, Deloitte Transactions and Business Analytics LLP, which provides a wide range of advisory and analytics services. These entities are separate subsidiaries of Deloitte LLP. Please see [www.deloitte.com/us/about](http://www.deloitte.com/us/about) for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting.

Copyright © 2020 Deloitte Development LLC. All rights reserved.