# ***Amazon Comprehend***

* Amazon Comprehend uses natural language processing (NLP) to extract insights about the content of documents.
* It develops insights by recognizing the entities, key phrases, language, sentiments, and other common elements in a document.
* Amazon Comprehend gathers the following types of insights:

- Entities

- Key Phrases

- Personally Identifiable Info

- Language

- Sentiment

- Targeted Sentiment

- Syntax

* Amazon Comprehend uses a pre-trained model to gather insights about a document or a set of documents. This model is continuously trained on a large body of text so that there is no need for you to provide training data.
* **Entities** - An entity is a textual reference to the unique name of a real-world object such as people, places, and commercial items, and to precise references to measures such as dates and quantities. Each entity also has a score that indicates the level of confidence that Amazon Comprehend has correctly detected the entity type.
* **Events** - Use event detection to analyze text documents for speciﬁc types of events and their related entities. Amazon Comprehend supports event detection across large collections of documents using asynchronous analysis jobs.
* From the input text, Amazon Comprehend extracts a list of entities that are related to the detected event. Each entity associated with a supported event type returns with Mentions
* Amazon Comprehend returns the list of events (of supported event types) that it detects in the input text
* Each event returns with the following related details:
  + Type
  + Arguments
  + Triggers
* **Key Phrases** - A key phrase is a string containing a noun phrase that describes a particular thing. It generally consists of a noun and the modifiers that distinguish it.
* **Dominant Language** - Uses Amazon Comprehend to examine text to determine the dominant language. Amazon Comprehend identifies the language using identifiers from RFC 5646
* if there is a 2-letter ISO 639-1 identifier, with a regional subtag if necessary, it uses that
* otherwise, it uses the ISO 639-2 3-letter code
* **Sentiment** - Use Amazon Comprehend to determine the sentiment of a document
* Sentiment determination returns the following values:
  + Positive
  + Negative
  + Neutral
  + Mixed
* **Targeted Sentiment** - Targeted sentiment provides a granular understanding of the sentiments associated with specific entities (such as brands or products) in your input documents.
* Targeted sentiment analysis determines the entity-level sentiment for specific entities in each input document. You can analyze the output data to determine the specific products and services that get positive or negative feedback.
* **Syntax** - Use syntax analysis to parse the words from the document and return the part of speech, or syntactic function, for each word in the document. You can identify the nouns, verbs, adjectives and so on in your document. Use this information to gain a richer understanding of the content of your documents, and to understand the relationship of the words in the document.
* **Personally Identifiable Information** - use the Amazon Comprehend console or APIs to detect personally identifiable information (PII) in English text documents. PII is a textual reference to personal data that could be used to identify an individual.

## **Use Cases**

* Customer Analytics(Sentiment Analysis)
  + Emails
  + Social Media
  + Online Comments
* Search Engines(Key Phrases Extractions)
* Content Recommendations(Topic Modeling)
  + Movies Recommendations
  + Readers Recommendations
* Customer Support Ticket Handling(Custom Classification/Sentiment Analysis/Key Phrases Extractions)
  + Feedback Forms
  + Product Reviews
  + Support Ticket
* Medical Cohort Analysis(Amazon Comprehend Medical)

## **Advantages**

* Easy to use
* Low cost & affordable
* Seamless integration with other AWS
* Highly Scalable
* Customizable

## **Disadvantages**

* Relatively New
* Not a complete AI solution
* Outputs files are zipped
* Models are too generic
* Billings and usage are not transparent