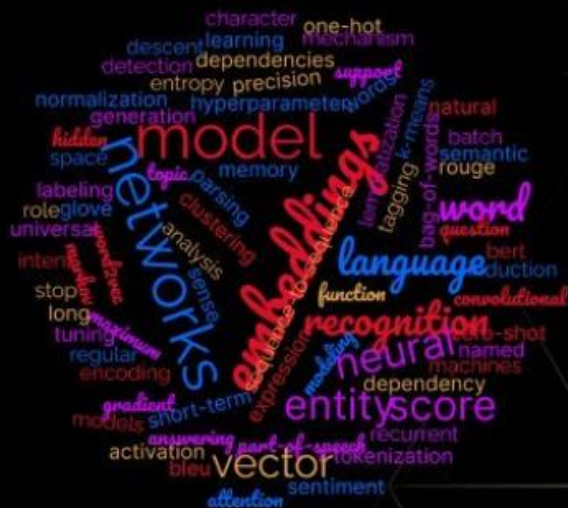# Mini NLP Dictionary

# Mini NLP Dictionary

## PART 1: A-F

- **Activation Function:** A mathematical function used in neural networks to introduce non-linearity, enabling the model to learn complex relationships between inputs and outputs. Examples include ReLU, Sigmoid, and Tanh.
- **Attention Mechanism:** A technique used in sequence-to-sequence models to focus on specific parts of the input data when generating the output.
- **Bag-of-Words (BoW):** A representation of text data where each document is represented as a bag, or set, of its word frequencies, ignoring word order.
- **Batch Normalization:** A technique used to normalize the input data for each layer in a neural network, improving stability and speeding up training.
- **BERT (Bidirectional Encoder Representations from Transformers):** A pre-trained language model developed by Google that has achieved state-of-the-art results in various NLP tasks.
- **BLEU (Bilingual Evaluation Understudy) Score:** A metric used to evaluate the quality of machine translation output by comparing it to a reference translation.
- **Character Embeddings:** Vector representations of individual characters in a word, used in character-level language models.
- **Convolutional Neural Networks (CNNs):** A type of neural network architecture commonly used for image processing, but also applied to NLP tasks like text classification and sentiment analysis.
- **Dependency Parsing:** A technique used to analyze sentence structure by identifying the grammatical dependencies between words.
- **Embeddings:** Vector representations of words, characters, or documents in a high-dimensional space, enabling semantic relationships to be captured.
- **Entity Recognition:** A task in NLP that involves identifying and categorizing named entities in unstructured text into categories like person, organization, or location.
- **F1 Score:** A metric used to evaluate the performance of a model, particularly in classification tasks, by calculating the harmonic mean of precision and recall.

# Mini NLP Dictionary

## PART 2: G-N

- **GloVe (Global Vectors for Word Representation):** A type of word embedding that represents words as vectors in a high-dimensional space, capturing semantic relationships.
- **Gradient Descent:** An optimization algorithm used to minimize the loss function in machine learning models, including NLP tasks.
- **Hidden Markov Model (HMM):** A statistical model used for modeling sequential data, such as speech or text, by assuming the underlying system is a Markov process.
- **Hyperparameter Tuning:** The process of adjusting model parameters to optimize performance, often using techniques like grid search, random search, or Bayesian optimization.
- **Intent Detection:** A task in NLP that involves identifying the intent behind a piece of text, such as determining whether a customer inquiry is about returns or product information.
- **K-Means Clustering:** An unsupervised learning algorithm used for clustering text data, such as grouping similar documents or customer feedback.
- **Language Model:** A type of AI model that predicts the next word in a sequence of text, given the context of the previous words.
- **Large Language Model (LLM):** Refers to extremely large-scale language models that are trained on vast amounts of text data, such as GPT-3 (Generative Pre-trained Transformer 3) developed by OpenAI. LLMs have significantly improved language understanding and generation capabilities.
- **Lemmatization:** The process of reducing words to their base or dictionary form, known as the lemma, to simplify text analysis.
- **Long Short-Term Memory (LSTM) Networks:** A type of recurrent neural network (RNN) designed to handle the vanishing gradient problem in sequence data.
- **Maximum Entropy (MaxEnt):** A machine learning model used for classification and tagging tasks, such as part-of-speech tagging or sentiment analysis.
- **Named Entity Recognition (NER):** A task in NLP that involves identifying and categorizing named entities in unstructured text into categories like person, organization, or location.
- **Natural Language Generation (NLG):** The process of generating human-like text or speech from structured data, such as generating product descriptions or chatbot responses.

# Mini NLP Dictionary

## PART 3: O-R

1. **One-Hot Encoding:** A technique used to convert categorical data, such as text labels, into a numerical representation that can be processed by machine learning algorithms.
2. **Part-of-Speech (POS) Tagging:** A task in NLP that involves identifying the grammatical category of each word in a sentence, such as noun, verb, or adjective.
3. **Precision:** A metric used to evaluate the performance of a model, calculated as the ratio of true positives to the sum of true positives and false positives.
4. **Question Answering:** A task in NLP that involves generating an answer to a question based on the content of a text or knowledge base.
5. **Recurrent Neural Networks (RNNs):** A type of neural network architecture designed to handle sequential data, such as text or speech, by maintaining a hidden state that captures information from previous time steps.
6. **Regular Expression (Regex):** A pattern-matching language used for string manipulation and text processing tasks, such as extracting specific patterns from text data.
7. **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** A metric used to evaluate the quality of automatic summarization systems by comparing the generated summary to a reference summary.

# Mini NLP Dictionary

## PART 4: S-U

1. **Semantic Role Labeling (SRL):** A task in NLP that involves identifying the roles played by entities in a sentence, such as "agent", "patient", or "theme".
2. **Sentiment Analysis: A** task in NLP that involves determining the emotional tone or attitude conveyed by a piece of text, such as positive, negative, or neutral.
3. **Sequence-to-Sequence (Seq2Seq) Models:** A type of neural network architecture used for tasks that involve generating a sequence of output tokens based on a sequence of input tokens, such as machine translation or text summarization.
4. **Stop Words:** Common words like "the", "and", or "a" that are often ignored in text analysis because they do not carry much meaningful information.
5. **Support Vector Machines (SVMs):** A type of machine learning algorithm used for classification and regression tasks, often used in NLP for tasks like text classification and sentiment analysis.
6. **Tokenization:** The process of breaking down text into individual words or tokens, which can then be processed by machine learning algorithms.
7. **Topic Modeling:** A technique used to extract underlying topics or themes from a large corpus of text data, such as Latent Dirichlet Allocation (LDA).
8. **Universal Dependencies (UD):** A framework for annotating sentence structure and dependencies, used for tasks like parsing and semantic role labeling.

# Mini NLP Dictionary

## PART 5: V-Z

1. **Vector Space Model:** A mathematical framework used to represent words or documents as vectors in a high-dimensional space, enabling semantic relationships to be captured.
2. **Word Embeddings:** Vector representations of words in a high-dimensional space, capturing semantic relationships and enabling word similarity and analogy calculations.
3. **Word Sense Induction (WSI):** A task in NLP that involves identifying the different senses or meanings of a word in a given context.
4. **Word2Vec:** A popular technique for learning word embeddings from large amounts of text data, using either the Continuous Bag of Words (CBOW) or Skip-Gram models.
5. **Zero-Shot Learning:** A machine learning paradigm where a model is trained on one set of tasks or classes and then applied to a new, unseen set of tasks or classes without additional training.