Glossary

Each of these technologies plays a distinct role in the broader field of language processing and contributes to the development of applications that interact with users through written or spoken language.

**HMM (Hidden Markov Model):** This is a statistical model that assumes the system being modeled is a Markov process with unobserved (hidden) states. In language processing, HMMs have been used for tasks like speech recognition and part-of-speech tagging, where each word or sound is treated as a state which the model transitions between.

**GPT (Generative Pre-trained Transformer):** This refers to a type of model architecture that uses transformers, which are deep learning models that rely on self-attention mechanisms to process input data. GPT models are pre-trained on large amounts of text and fine-tuned on specific tasks. They are used for generating coherent and contextually relevant text based on input prompts, making them effective for tasks like text completion, translation, and more.

**ASR (Automatic Speech Recognition):** ASR systems convert spoken language into text. These systems typically involve acoustic modeling (to capture the properties of spoken language) and language modeling (to understand linguistic structures). ASR is crucial for enabling voice-controlled applications like virtual assistants.

**RNN (Recurrent Neural Network):** RNNs are a class of neural networks that are effective at processing sequences, such as sequences of words or sounds. They are capable of maintaining a form of internal memory that captures information about previous elements in the sequence, making them suitable for tasks where context from earlier in the sequence is important, such as speech recognition or language modeling.

**CNN (Convolutional Neural Network):** Although primarily known for their use in image processing, CNNs are also used in language processing, particularly for tasks that can benefit from local feature detection, such as sentiment analysis or text classification. In speech recognition, CNNs can help process spectrograms or other visual representations of audio data.

**NLU (Natural Language Understanding**): NLU involves processing and understanding human language in a way that computers can effectively respond to. It includes interpreting the intents and entities in the user's input and often deals with complexities such as ambiguity and context. NLU is essential for systems that interact with humans in natural language, such as chatbots and virtual assistants.

**TTS (Text-to-Speech):** TTS technologies convert written text into spoken speech. This involves the synthesis of human-like speech from text, including the generation of correct pronunciations, intonations, and rhythms. TTS is crucial for accessibility features, audiobooks, and virtual assistants that need to communicate with users audibly.

**Interactive Voice Response (IVR):** IVR systems allow users to interact with a computer-operated phone system through the use of voice and DTMF tones input via a keypad. It's widely used in customer service and information retrieval systems.

**Voice Command and Control (VCC):** This type of technology is used to perform specific commands or control devices and software through voice commands. It's commonly seen in smartphones, smart home devices, and various types of consumer electronics.

**Speaker Identification and Authentication:** This type of voice recognition is used to identify and verify the identity of a speaker. It's commonly used in security applications where voice serves as a biometric identifier, similar to a fingerprint or facial recognition.

**Speech-to-Text (STT):** Similar to ASR, STT technology specifically focuses on converting spoken words into written text. It is widely used in transcription services, live event captioning, and as an accessibility tool for people with disabilities.

**Text-to-Speech (TTS):** The opposite of ASR and STT, TTS systems convert written text into spoken words. This is useful for assistive technologies, reading digital content aloud, and in interfaces where auditory feedback is necessary.

**Machine learning (ML):** A subset of AI, ML refers to systems that learn from experience and get “smarter” over time, without human intervention. ML is a method of training an algorithm so that it can learn how to perform a specific task. Training involves feeding the algorithm large amounts of data and allowing it to adjust and improve.

**Deep learning (DL):** Like ML, DL refers to systems that learn from experience. The difference is that DL is applied to much larger data sets.

**Neural networks:** This is a technique or approach to ML that refers to biologically inspired networks of artificial neurons that mimic the structure and capabilities of the animal brain. It is a framework or model designed to help machines learn.