# **Assignment 01: Understanding OpenAI Chat Completion API Parameters**

|  |  |
| --- | --- |
| **Parameter** | **Description** |
| **Messages** | Message is a conversation between user and the agent. |
| **Model** | Model specifies that which version of LLM we are using like ChatGPT 3.5, 4.0, o1-mini etc. |
| **Max Completion Tokens** | Token could be a word or comprises of multiple words of characters. It varies with respect to the prompt and model of LLM. |
| **n** | Specifies how many different completions (responses) to generate for a given prompt. Useful for getting multiple options or responses. |
| **Stream** | When set to true, it enables real-time streaming of partial results as they are generated, rather than waiting for the full completion. |
| **Temperature** | It controls the randomness and response of the LLM to be included in the generated response. For example,  Controls the randomness of the output. Lower values (0) make the model more deterministic, while higher values (1) increase creativity and diversity in responses. |
| **Top\_p** | Determines the diversity of the model’s output by considering the smallest set of possible tokens whose cumulative probability is greater than p (nucleus sampling). |
| **Tools** | Refers to external tools or APIs that can be invoked by the model for tasks requiring additional functionality beyond language generation (e.g., code interpreter). |