Assignment 01: Understanding OpenAI Chat completion API Parameters

Kindly find the purpose and functionality of each term

| **Term** | **Purpose** | **Functionality** |
| --- | --- | --- |
| **Messages** | Input by user or AI generated response both are messages | Helps to generate relevant response and maintain history in a long conversations. |
| **Model** | Refers to the specific AI version used for a task (e.g., GPT-3.5, GPT-4). | Determines the capabilities, knowledge, and behavior of the AI in generating responses. |
| **Max Completion Tokens** | Limits the length of the AI response in terms of tokens (words/pieces of words). | Stops long responses by restricting the number of tokens the AI can generate. |
| **n** | reefers how many response the model should generate for a single prompt. | Allows generating multiple answers of to the same question. |
| **Stream** | Determines if the model’s output is generated and sent in real-time. | Enables real-time response generation as it’s being created (without waiting for the full response). |
| **Temperature** | Controls the randomness of the AI’s responses. | Affects creativity: low values (e.g., 0.1) make responses near reality, high values (e.g., 1.0) add randomness or imaginations. |
| **Top-p** | Controls the diversity of the model’s output by limiting the number of considered tokens. | Focuses on the smallest set of tokens that make up a cumulative probability, allowing diverse but coherent output. |
| **Tools** | Refers to external functions or systems the model can access. | Expands AI functionality by allowing it to interact with external tools like calendars, databases, or the internet. |

Bottom of Form