**1. Messages**

The 'messages' parameter provides a structured format for the conversation's context. It consists of roles such as 'user', 'assistant', and 'system', along with their respective content. For example:  
  
[  
 {"role": "system", "content": "You are a supportive assistant."},  
 {"role": "user", "content": "What is the capital of Spain?"}  
]  
  
This structure ensures the model understands the flow and context, enabling coherent responses.

**2. Model**

The 'model' parameter determines which AI model to use. For instance:

* 'GPT-4': Best for highly complex and nuanced tasks.
* 'GPT -3.5-turbo': Cost-effective and suitable for general-purpose tasks.  
    
  Selecting the right model depends on factors such as complexity, budget, and desired response quality.

**3. Max Completion Tokens**

This parameter sets a limit on the number of tokens (words, punctuation, or fragments) in a response. For example:

* Setting it to 50 ensures brevity in responses.
* Higher values allow detailed and longer outputs.  
    
  It is crucial for managing token usage effectively within the API's constraints.

**4. n**

The 'n' parameter specifies the number of response variations to generate. For example, setting 'n' to 3 produces three different outputs for the same query. This is beneficial for comparing options and selecting the most fitting response.

**5. Stream**

When enabled, the 'stream' parameter delivers the response incrementally, token by token. This feature is particularly valuable for real-time applications, such as live chat interfaces, where immediate feedback is essential.

**6. Temperature**

The 'temperature' parameter controls the randomness and creativity of responses:

* A higher value (e.g., 1.0) results in diverse and imaginative outputs.
* A lower value (e.g., 0.2) ensures more focused and predictable results.  
    
  It is an excellent way to balance between creativity and precision.

**7.** **Top\_p**

Also known as 'nucleus sampling', the 'Top\_p ' parameter determines response diversity. Instead of selecting tokens based solely on probability, it considers a subset whose cumulative probability is within the 'Top\_p ' value:

* Top\_p: 0.1 restricts responses to the top 10% of token probabilities.
* Top\_p: 1.0 includes all token probabilities, similar to using temperature.

**8. Tools**

The 'tools' parameter enables the model to integrate with external functionalities, such as APIs, databases, or calculators. For example, it allows the model to fetch live data or perform computations, enhancing its capabilities beyond static responses.