AI API Glossary and Generative AI Tools Overview

# Glossary of API Terminology

|  |  |
| --- | --- |
| Term | Definition |
| API (Application Programming Interface) | A set of rules and protocols allowing different software systems to communicate. In this context, used to send prompts to an AI model and receive responses programmatically. |
| API Key | A unique authentication code required to access an API securely, identifying the user or application making the request. |
| Endpoint | A specific URL path where an API request is sent to perform a particular function (e.g., generate a completion, retrieve data). |
| Prompt | The input message or instruction provided to the AI model, describing the task or question it should respond to. |
| Token | A unit of text used for processing in AI models, typically representing a word or subword. APIs often bill based on the number of tokens processed in a request and response. |
| Temperature | A parameter controlling the randomness of AI-generated outputs. Lower values result in more deterministic responses; higher values increase creativity and variability. |
| Max Tokens | A parameter specifying the maximum length of the AI’s response in tokens. Helps control output size. |
| Rate Limit | A restriction on how many API requests can be made within a certain time period, enforced to prevent overuse or server overload. |
| Latency | The delay between sending an API request and receiving a response, measured in milliseconds or seconds. |
| Payload | The body of an API request containing the prompt, parameters, and other necessary data. |
| Response Object | The structured data returned by the API, usually in JSON format, containing the generated output and related metadata. |

# General Overview of Generative AI Tools for Research and Analysis

|  |  |  |
| --- | --- | --- |
| AI Tool | Best Suited For | Limitations / Considerations |
| Claude (Anthropic) | High-quality summarization of long texts, nuanced qualitative analysis, generating structured narrative responses. | Performs well on interpretive tasks; may lack precision in numeric outputs or code generation. |
| ChatGPT (OpenAI) | Brainstorming, ideation, drafting, qualitative content analysis, exploratory synthesis across diverse topics. | Tends to hallucinate or oversimplify technical or numerical details; requires strong prompt guidance for accuracy. |
| Perplexity AI | Quick fact-finding, exploratory research, web search with citations, answering questions based on online sources. | Output reliability depends on real-time web data; not specialized for deep synthesis or data manipulation. |
| DeepSeek | Code generation, data parsing, script drafting, technical content generation. | Emerging platform; less mature than competitors; may lack robustness for complex programming tasks. |
| Gemini (Google AI) | Conversational queries blending web search with generative responses, general-purpose research assistance. | Output may prioritize accessible summaries over depth; limited parameter control compared to developer APIs. |
| OpenAI API (GPT models) | Programmatic integration for automating content generation, analysis, or batch processing. | Requires API management; token limits and costs must be monitored; prompt engineering needed for consistency. |
| Claude API | Similar programmatic applications as OpenAI API but optimized for longer-context inputs and collaborative-style outputs. | Same API usage considerations; model behaviors may differ in tone, interpretation style, and verbosity. |