

init(temperature:topP:topK:candidateCount:maxOutputTokens:presencePenalty:frequencyPenalty:stopSequences:responseMimeType:responseSchema:)

Creates a new `GenerationConfig` value.

iOS 15.0+ iPadOS 15.0+ Mac Catalyst 15.0+ macOS 12.0+ tvOS 15.0+ watchOS 8.0+

```
init(  
    temperature: Float? = nil,  
    topP: Float? = nil,  
    topK: Int? = nil,  
    candidateCount: Int? = nil,  
    maxOutputTokens: Int? = nil,  
    presencePenalty: Float? = nil,  
    frequencyPenalty: Float? = nil,  
    stopSequences: [String]? = nil,  
    responseMimeType: String? = nil,  
    responseSchema: Schema? = nil  
)
```

Parameters

temperature

Controls the randomness of the language model's output. Higher values (for example, 1.0) make the text more random and creative, while lower values (for example, 0.1) make it more focused and deterministic.

Note

A temperature of 0 means that the highest probability tokens are always selected. In this case, responses for a given prompt are mostly deterministic, but a small amount of variation is still possible.

Important

The range of supported temperature values depends on the model; see the [Cloud documentation](#) for more details.

topP

Controls diversity of generated text. Higher values (e.g., 0.9) produce more diverse text, while lower values (e.g., 0.5) make the

output more focused.

The supported range is 0.0 to 1.0.

Important

The default topP value depends on the model; see the [Cloud documentation](#) for more details.

topK

Limits the number of highest probability words the model considers when generating text. For example, a topK of 40 means only the 40 most likely words are considered for the next token. A higher value increases diversity, while a lower value makes the output more deterministic.

The supported range is 1 to 40.

Important

Support for topK and the default value depends on the model; see the [Cloud documentation](#) for more details.

candidateCount

The number of response variations to return; defaults to 1 if not set. Support for multiple candidates depends on the model; see the [Cloud documentation](#) for more details.

maxOutputTokens

Maximum number of tokens that can be generated in the response. See the configure model parameters [documentation](#) for more details.

presencePenalty

Controls the likelihood of repeating the same words or phrases already generated in the text. Higher values increase the penalty of repetition, resulting in more diverse output.

Note

While both presencePenalty and frequencyPenalty discourage repetition, presencePenalty applies the same penalty regardless of how many times the word/phrase has already appeared, whereas frequencyPenalty increases the penalty for *each* repetition of a word/phrase.

Important

The range of supported presencePenalty values depends on the model; see the [Cloud documentation](#) for more details

frequencyPenalty

Controls the likelihood of repeating words or phrases, with the penalty increasing for each repetition. Higher values increase the penalty of repetition, resulting in more diverse output.

Note

While both `frequencyPenalty` and `presencePenalty` discourage repetition, `frequencyPenalty` increases the penalty for *each* repetition of a word/phrase, whereas `presencePenalty` applies the same penalty regardless of how many times the word/phrase has already appeared.

Important

The range of supported `frequencyPenalty` values depends on the model; see the [Cloud documentation](#) for more details

stopSequences

A set of up to 5 Strings that will stop output generation. If specified, the API will stop at the first appearance of a stop sequence. The stop sequence will not be included as part of the response. See the [Cloud documentation](#) for more details.

responseMimeType

Output response MIME type of the generated candidate text.

Supported MIME types:

- `text/plain`: Text output; the default behavior if unspecified.
- `application/json`: JSON response in the candidates.
- `text/x.enum`: For classification tasks, output an enum value as defined in the `responseSchema`.

responseSchema

Output schema of the generated candidate text. If set, a compatible `responseMimeType` must also be set.

Compatible MIME types:

- `application/json`: Schema for JSON response.

Refer to the [Control generated output](#) guide for more details.

Discussion

See the [Configure model parameters](#) guide and the [Cloud documentation](#) for more details.