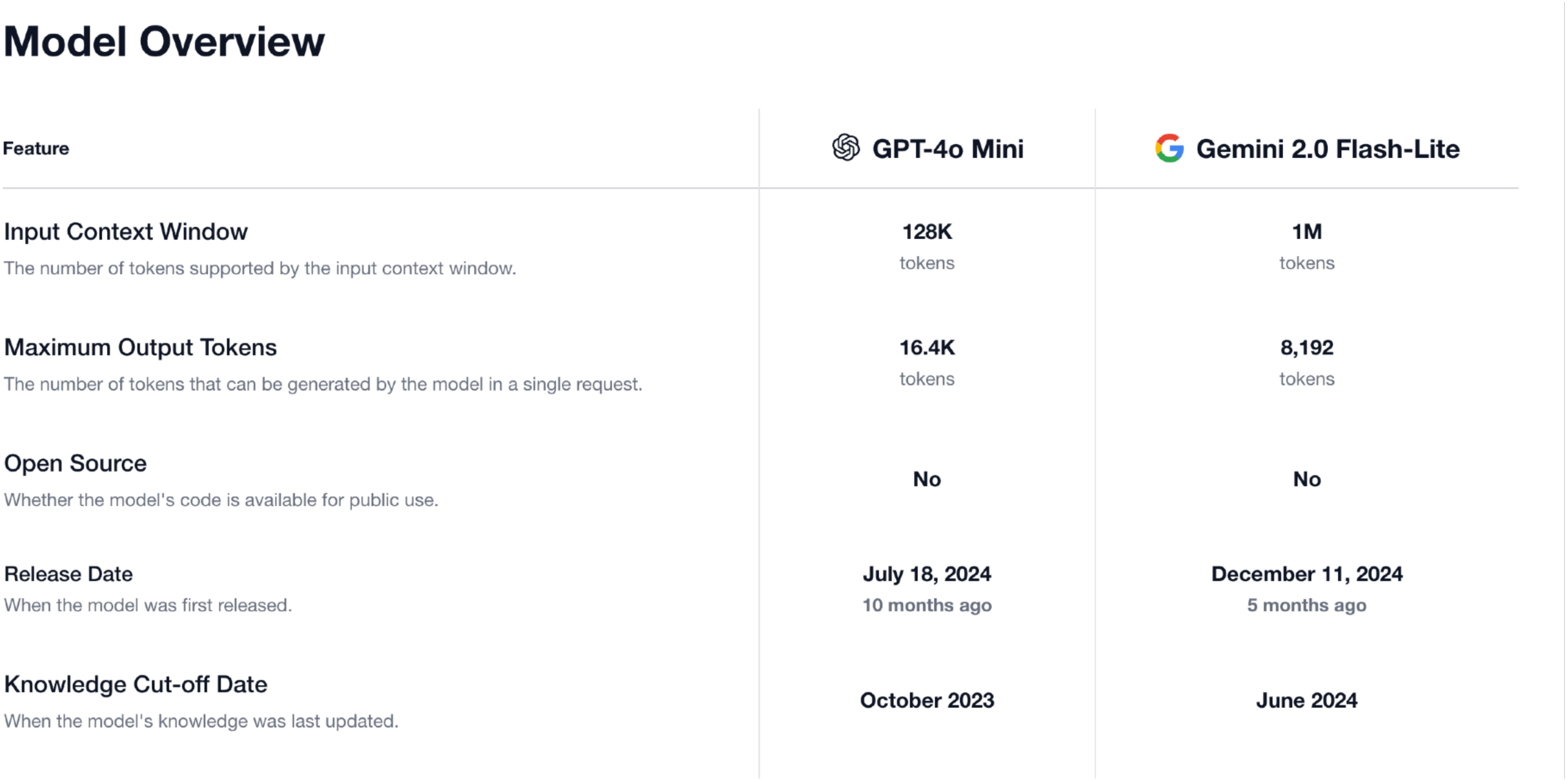

Model Overview

Feature	GPT-4o Mini	Gemini 2.0 Flash-Lite
Input Context Window The number of tokens supported by the input context window.	**128K** tokens	**1M** tokens
Maximum Output Tokens The number of tokens that can be generated by the model in a single request.	**16.4K** tokens	**8,192** tokens
Open Source Whether the model's code is available for public use.	**No**	**No**
Release Date When the model was first released.	**July 18, 2024** **10 months ago**	**December 11, 2024** **5 months ago**
Knowledge Cut-off Date When the model's knowledge was last updated.	**October 2023**	**June 2024**

Pricing Comparison

Compare costs for input and output tokens between GPT-4o Mini and Gemini 2.0 Flash-Lite.

Price Type	GPT-4o Mini	Gemini 2.0 Flash-Lite
Input Cost for processing tokens in your prompts	**$0.15** per million tokens	**$0.07** per million tokens
Output Cost for tokens generated by the model	**$0.60** per million tokens	**$0.30** per million tokens