

RoostGPT AI Provider

Support for Unit, API and Other Tests

Overview

RoostGPT supports multiple AI providers for comprehensive test generation across unit tests, API tests, Integration Tests, Functional Tests and more. This document outlines the supported providers and their available models.

Supported Providers:

- OpenAI
- Claude AI (Anthropic)
- Azure OpenAI
- Vertex AI (Google Cloud)
- AWS Bedrock

OpenAI Models

Model Name	Context Window (Tokens)	Best For	Key Features
GPT-5	4 Million Input: 272k Output: 128k	Complex test scenarios, enterprise	Expert-level intelligence, built-in reasoning
GPT-4.5	Large	Standard test generation	Balanced performance, cost-effective
GPT-4o	Large	Fast test generation	Multimodal, real-time processing
GPT-4o mini	Standard	High-volume generation	Cost-efficient, fast

Model Name	Context Window (Tokens)	Best For	Key Features
GPT-4 Turbo	128K tokens	Production environments	Proven reliability
o3	Extended	Complex reasoning tasks	Advanced chain-of-thought
o4-mini	Standard	Cost-efficient reasoning	Faster reasoning model

Claude AI (Anthropic) Models

Model Name	Context Window (Tokens)	Best For	Key Features
Claude Opus 4.1	Input: 200K Output: 32k	Complex coding tasks, unit tests	Best coding model, 74.5% on SWE-bench
Claude Sonnet 4.5	Output: 64k	Coding, Finance and CyberSecurity	Long running tasks, enhanced domain knowledge and 77.2% on SWE-bench
Claude Sonnet 4	200K-1M Output: 64k	General test generation	Superior coding and reasoning, instruction following
Claude Sonnet 3.7	200K tokens Output: 64k	Hybrid reasoning tasks	Fast and extended thinking modes
Claude Sonnet 3.5	200K tokens Output: 8k	Legacy support	Strong coding capabilities
Claude Haiku 3.5	200K tokens Output: 8k	Quick, simple tests	Fast, cost-effective

Azure OpenAI Models

Azure OpenAI provides the same OpenAI models with enterprise features.

Model Name	Context Window	Additional Features
GPT-5 Series	4 Million Input: 272k Output: 128k	VNET integration, Private endpoints
GPT-4.5	Large	Managed Identity, Regional deployment

Model Name	Context Window	Additional Features
GPT-4o Series	Large	SLA guarantees, Azure Monitor
GPT-4 Turbo	128K tokens	Standard and Provisioned deployments
o3, o4-mini	Extended/Standard	Azure AD authentication

Vertex AI (Google Cloud) Models

Model Name	Context Window (Tokens)	Best For	Key Features
Gemini 2.5 Pro	Input: 1024k Output: 64k	Complex API testing, large codebases	Deep Think mode, long context

AWS Bedrock Models

AWS Bedrock provides access to multiple foundation models from various providers in a fully managed service.

Available Model Families

Provider	Models	Context Window	Best For
Anthropic	Claude Opus 4.1, Sonnet 4, Sonnet 3.7	Output Token - 4K	Code generation, unit tests

Recommended Models

OpenAI

- GPT-4o, GPT-5

Azure OpenAI

- GPT-4o, GPT-5

Claude

- Claude Sonnet 4.1

Support Resources

- **OpenAI:** platform.openai.com/docs
- **Anthropic (Claude):** docs.claude.com
- **Azure OpenAI:** learn.microsoft.com/azure/ai-services/openai
- **Vertex AI:** cloud.google.com/vertex-ai/docs
- **AWS Bedrock:** docs.aws.amazon.com/bedrock

Revision #7

Created 29 September 2025 04:55:03 by Divyesh

Updated 30 September 2025 15:28:13 by Harish