

Search Results

Searching for: Use Case: Code Generation

All (40) Models (40) Blueprints (0)



microsoft / **phi-4-mini-instruct**

Lightweight multilingual LLM powering AI applications in latency bound, memory/compute constrained environments

code generation

chat

+3



tiiuae / **falcon3-7b-instruct**

Instruction tuned LLM achieving SoTA performance on reasoning, math and general knowledge capabilities

coding

code generation

+5



nvidia / **usdcode**

State-of-the-art LLM that answers OpenUSD knowledge queries and generates USD-Python code.

openusd

synthetic data generation

+5



meta / llama-3.3-70b-instruct

Advanced LLM for reasoning, math, general knowledge, and function calling

[reasoning](#)

[code generation](#)

+4



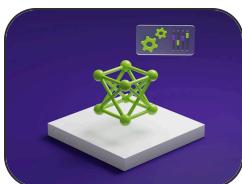
nvidia / llama-3.1-nemotron-70b-instruct

Llama-3.1-Nemotron-70B-Instruct is a large language model customized by NVIDIA in order to improve the helpfulness of LLM generated responses.

[code generation](#)

[chat](#)

+3



nvidia / mistral-nemo-minitron-8b-8k-instruct

State-of-the-art small language model delivering superior accuracy for chatbot, virtual assistants, and content generation.

[small language model](#)

[code generation](#)

+4



meta / llama-3.2-3b-instruct

Advanced state-of-the-art small language model with language

[Filter by text](#)

[Sort By: Most Recent](#)

[Publisher](#)

[Use Case](#)

(1)

[Use Case: Code Generation](#)

[Clear Filters](#)



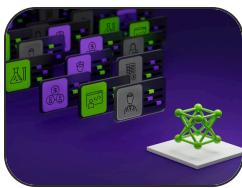
yentinglin / llama-3-taiwan-70b-instruct

Sovereign AI model finetuned on Traditional Mandarin and English data using the Llama-3 architecture.

[regional language generation](#)

[code generation](#)

+3

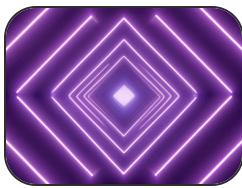


microsoft / phi-3.5-moe-instruct

Advanced LLM based on Mixture of Experts architecture to deliver compute efficient content generation

[moe](#)[code generation](#)

+4



microsoft / phi-3.5-mini-instruct

Lightweight multilingual LLM powering AI applications in latency bound, memory/compute constrained environments

[code generation](#)[chat](#)

+4



google / gemma-2-2b-it

Advanced small language generative AI model for edge applications

[code generation](#)[chat](#)

+3



mistralai / mamba-codestral-7b-v0.1

Model for writing and interacting with code across a wide range of programming languages and tasks.

[code completion](#)[code generation](#)

+2



meta / llama-3.1-405b-instruct

Advanced LLM for synthetic data generation, distillation, and inference for chatbots, coding, and domain-specific tasks.

[chat](#)[code generation](#)

+2

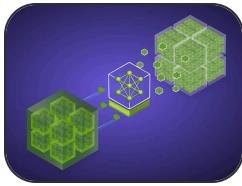


meta / llama-3.1-70b-instruct

Powers complex conversations with superior contextual understanding, reasoning and text generation.

[code generation](#)[chat](#)

+3



meta / llama-3.1-8b-instruct

Advanced state-of-the-art model with language understanding, superior reasoning, and text generation.

[run on rtx](#)[code generation](#)

+4



nv-mistralai / mistral-nemo-12b-instruct

Most advanced language model for reasoning, code, multilingual tasks; runs on a single GPU.

[code generation](#)[chat](#)

+3



microsoft / phi-3-medium-128k-instruct

Cutting-edge lightweight open language model excelling in high-quality reasoning.

[code generation](#)[chat](#)

+4



bigcode / starcoder2-7b

Advanced programming model for code completion, summarization, and generation

[code completion](#)[code generation](#)

+2



bigcode / starcoder2-15b

Advanced programming model for code completion, summarization, and generation

[code completion](#)[code generation](#)

+2

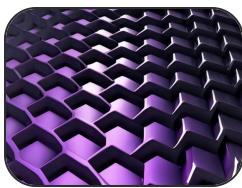


google / gemma-2-27b-it

Cutting-edge text generation model text understanding, transformation, and code generation.

[code generation](#)[chat](#)

+3

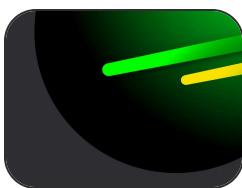


google / **gemma-2-9b-it**

Cutting-edge text generation model text understanding, transformation, and code generation.

[code generation](#)[chat](#)

+3



01-ai / **yi-large**

Powerful model trained on English and Chinese for diverse tasks including chatbot and creative writing.

[code generation](#)[chat](#)

+3

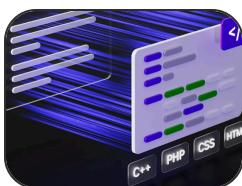


google / **codegemma-1.1-7b**

Advanced programming model for code generation, completion, reasoning, and instruction following.

[code generation](#)[code completion](#)

+1



ibm / **granite-34b-code-instruct**

Software programming LLM for code generation, completion, explanation, and multi-turn conversion.

[code generation](#)[chat](#)

+3



ibm / **granite-8b-code-instruct**

Software programming LLM for code generation, completion, explanation, and multi-turn conversion.

[code generation](#)[chat](#)

+3



microsoft / **phi-3-small-8k-instruct**

Cutting-edge lightweight open language model excelling in high-quality reasoning.

[code generation](#)[chat](#)

+4



microsoft / phi-3-small-128k-instruct

Long context cutting-edge lightweight open language model exceling in high-quality reasoning.

[code generation](#)[chat](#)

+4



microsoft / phi-3-medium-4k-instruct

Cutting-edge lightweight open language model exceling in high-quality reasoning.

[code generation](#)[chat](#)

+4



microsoft / phi-3-mini-4k-instruct

Lightweight, state-of-the-art open LLM with strong math and logical reasoning skills.

[code generation](#)[chat](#)

+4



microsoft / phi-3-mini-128k-instruct

Lightweight, state-of-the-art open LLM with strong math and logical reasoning skills.

[code generation](#)[chat](#)

+4



mistralai / mixtral-8x22b-instruct-v0.1

An MOE LLM that follows instructions, completes requests, and generates creative text.

[advanced reasoning](#)[code generation](#)

+4



meta / llama3-70b-instruct

Powers complex conversations with superior contextual understanding, reasoning and text generation.

[code generation](#)[chat](#)

+4



meta / llama3-8b-instruct

Advanced state-of-the-art LLM with language understanding, superior reasoning, and text generation.

[code generation](#)
[chat](#)

+4



google / recurrentgemma-2b

Novel recurrent architecture based language model for faster inference when generating long sequences.

[code generation](#)
[chat](#)

+3



google / codegemma-7b

Cutting-edge model built on Google's Gemma-7B specialized for code generation and code completion.

[code generation](#)
[chat](#)

+3



google / gemma-2b

Lightweight language model deployable on laptop, desktop or the cloud for summarization and reasoning.

[code generation](#)
[chat](#)

+3



google / gemma-7b

Cutting-edge text generation model text understanding, transformation, and code generation.

[code generation](#)
[chat](#)

+3

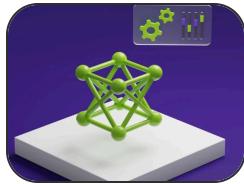


meta / codellama-70b

LLM capable of generating code from natural language and vice versa.

[code generation](#)
[chat](#)

+3



mistralai / mixtral-8x7b-instruct-v0.1

An MOE LLM that follows instructions, completes requests, and generates creative text.

[advanced reasoning](#)

[code generation](#)

+4

[Terms of Use](#) | [Privacy Policy](#) | [Manage My Privacy](#) | [Contact](#)

Copyright © 2025 NVIDIA Corporation