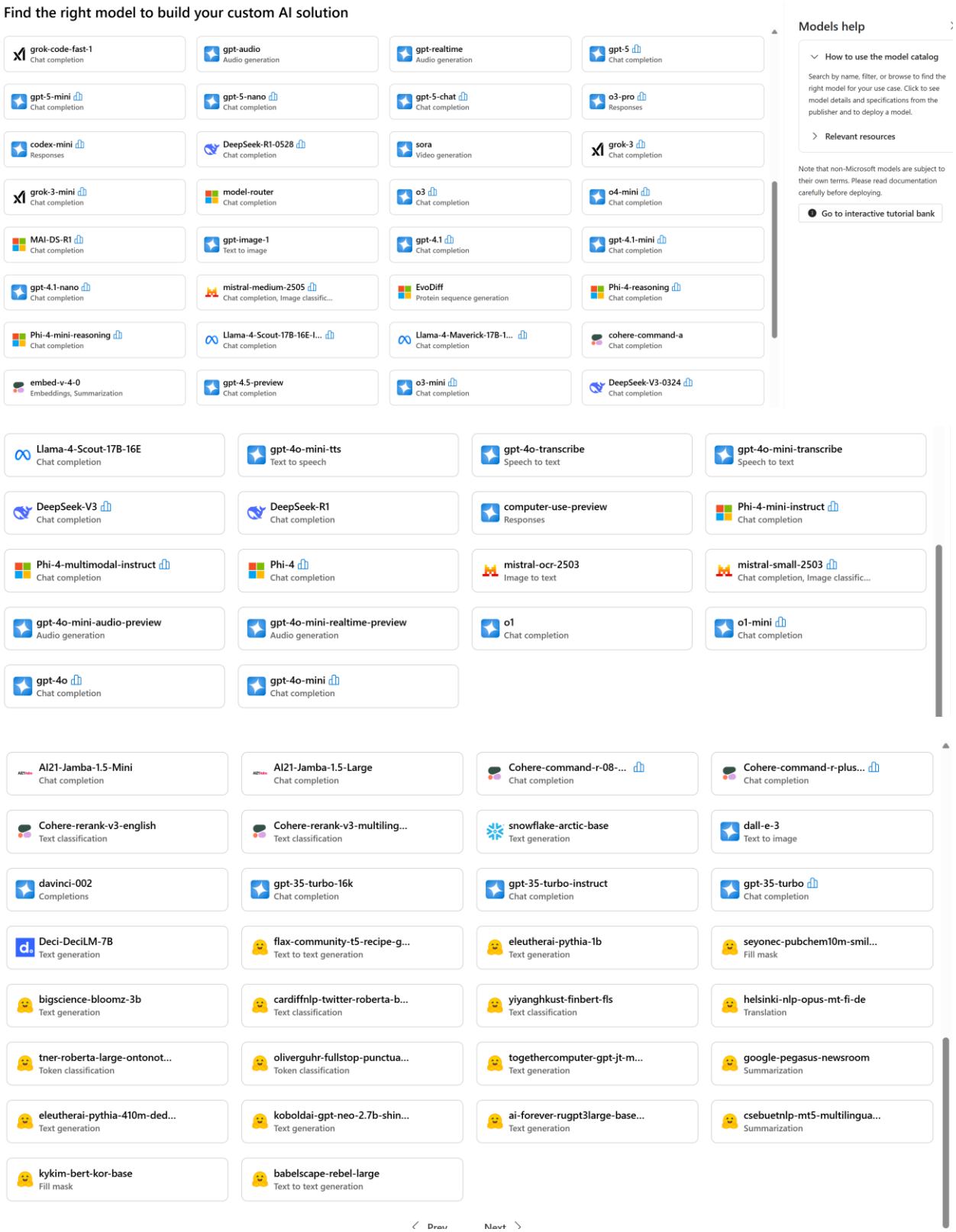
Find the right model to build your custom AI solution
grok-code-fast-1
Chat completion
gpt-audio
Audio generation
gpt-realtime
Audio generation
gpt-5
Chat completion
gpt-5-mini
Chat completion
gpt-5-nano
Chat completion
gpt-5-chat
Chat completion
o3-pro
Responses
codex-mini
Responses
DeepSeek-R1-0528
Chat completion
sora
Video generation
grok-3
Chat completion
grok-3-mini
Chat completion
model-router
Chat completion
o3
Chat completion
o4-mini
Chat completion
MAI-DS-R1
Chat completion
gpt-image-1
Text to image
gpt-4.1
Chat completion
gpt-4.1-mini
Chat completion
gpt-4.1-nano
Chat completion
mistral-medium-2505
Chat completion, Image classific...
EvoDiff
Protein sequence generation
Phi-4-reasoning
Chat completion
Phi-4-mini-reasoning
Chat completion
Llama-4-Scout-17B-16E-I...
Chat completion
Llama-4-Maverick-17B-1...
Chat completion
cohere-command-a
Chat completion
embed-v-4-0
Embeddings, Summarization
gpt-4.5-preview
Chat completion
o3-mini
Chat completion
DeepSeek-V3-0324
Chat completion
Models help
How to use the model catalog
Search by name, filter, or browse to find the right model for your use case. Click to see model details and specifications from the publisher and to deploy a model.
Relevant resources
Note that non-Microsoft models are subject to their own terms. Please read documentation carefully before deploying.
Go to interactive tutorial bank
Llama-4-Scout-17B-16E
Chat completion
gpt-4o-mini-tts
Text to speech
gpt-4o-transcribe
Speech to text
gpt-4o-mini-transcribe
Speech to text
DeepSeek-V3
Chat completion
DeepSeek-R1
Chat completion
computer-use-preview
Responses
Phi-4-mini-instruct
Chat completion
Phi-4-multimodal-instruct
Chat completion
Phi-4
Chat completion
mistral-ocr-2503
Image to text
mistral-small-2503
Chat completion, Image classific...
gpt-4o-mini-audio-preview
Audio generation
gpt-4o-mini-realtime-preview
Audio generation
o1
Chat completion
o1-mini
Chat completion
gpt-4o
Chat completion
gpt-4o-mini
Chat completion
AI21-Jamba-1.5-Mini
Chat completion
AI21-Jamba-1.5-Large
Chat completion
Cohere-command-r-08-...
Chat completion
Cohere-command-r-plus...
Chat completion
Cohere-rerank-v3-english
Text classification
Cohere-rerank-v3-multiling...
Text classification
snowflake-arctic-base
Text generation
dall-e-3
Text to image
davinci-002
Completions
gpt-35-turbo-16k
Chat completion
gpt-35-turbo-instruct
Chat completion
gpt-35-turbo
Chat completion
Deci-DeciLM-7B
Text generation
flax-community-t5-recipe-g...
Text to text generation
eleutherai-pythia-1b
Text generation
seyonec-pubchem10m-smil...
Fill mask
bigscience-bloomz-3b
Text generation
cardiffnlp-twitter-roberta-b...
Text classification
yiyanghkust-finbert-fls
Text classification
helsinki-nlp-opus-mt-fi-de
Translation
tner-roberta-large-ontonot...
Token classification
oliverguhr-fullstop-punctua...
Token classification
togethercomputer-gpt-jt-m...
Text generation
google-pegasus-newsroom
Summarization
eleutherai-pythia-410m-ded...
Text generation
koboldai-gpt-neo-2.7b-shin...
Text generation
ai-forever-rugpt3large-base...
Text generation
csebuetnlp-mt5-multilingua...
Summarization
kykim-bert-kor-base
Fill mask
babelscape-rebel-large
Text to text generation
Prev
Next