

Virtus AI Platform

Add AI Model - Form Field Guide

This guide explains each field in the Add Model form. Use this form to configure AI models that will be available in your marketplace for users to subscribe to.

Basic Information

Display Name

The name shown to users in the marketplace. This should be descriptive and memorable, reflecting the model's purpose or specialty.

Example: GPT-4 Model

Category

Groups the model in the marketplace for easier discovery. Categories are defined in Admin > Categories and can be customized. Users can filter models by category.

Example: AI Language Models

Description

Marketing text explaining what the model does, who it's for, and its key capabilities. This appears on the model's detail page in the marketplace.

Example: A powerful AI language model for generating human-like text across various topics.

Provider Configuration

Provider

The AI service that powers this model. Choose based on where the model is hosted:

- OpenAI: GPT-4, GPT-3.5, and other OpenAI models
- Anthropic: Claude 3, Claude 2 models
- Ollama: Self-hosted open-source models (Llama, Mistral)
- vLLM: High-performance inference server for custom deployments
- Custom: Any API that uses OpenAI-compatible format

Model ID

The exact identifier used by the provider's API. This must match the provider's naming convention exactly.

Example: gpt-4

Base URL / API Endpoint

For self-hosted models (Ollama, vLLM, Custom): The URL where the model API is accessible. Not needed for OpenAI or Anthropic as they use their default endpoints.

Example: https://api.ollama.com/v1/llama/gpt-4

API Key

Authentication key for the provider. Required for OpenAI and Anthropic. Optional for self-hosted solutions if they require authentication. Keep this secure.

Example: 1234567890abcdef1234567890abcdef

Virtus AI Platform

Add AI Model - Form Field Guide

Model Parameters

Context Length

Maximum number of input tokens the model can process in a single request. This determines how much conversation history and context can be included. Higher values allow longer conversations but may increase costs.

Example

Max Output Tokens

Limit on the length of model responses. Prevents runaway generation and helps control costs. Set based on your use case - shorter for chat, longer for content generation.

Example

Fine-tuned Model (Toggle)

Enable if this is a custom-trained model based on a foundation model. When enabled, shows the Base Model field to specify the original model it was trained from.

Base Model

Only shown when Fine-tuned is enabled. Specify the foundation model that was fine-tuned to create this custom model. Helps users understand the model's capabilities.

Example

Data Sources (RAG)

Default Data Sources

Select data sources that will automatically be used for Retrieval-Augmented Generation (RAG) when users chat with this model. These provide domain-specific knowledge to the model.

Data sources can include:

- Document collections (PDFs, Word docs)
- Database connections
- Website content
- API endpoints

Users who subscribe can also add their own additional data sources.

Virtus AI Platform

Add AI Model - Form Field Guide

Pricing Configuration

Pricing Type

How users will be charged for using this model:

- Per Token: Charge based on input/output tokens used
- Per Request: Fixed price for each API call
- Per User: Monthly fee per active user
- Subscription: Fixed monthly fee with included usage
- Hybrid: Base subscription plus usage-based overage charges

Monthly Subscription (\$)

Fixed monthly fee for access to the model. Set to 0 for pay-per-use models. This amount is charged at the beginning of each billing period.

Example

Price per 1K Input Tokens (\$)

Cost charged for processing user messages. Input tokens are counted from the user's messages and any context/system prompts. Typically lower than output token pricing.

Example

Price per 1K Output Tokens (\$)

Cost charged for model responses. Output tokens are the generated text from the AI. Usually priced higher than input tokens as generation is more compute-intensive.

Example

Price per Request (\$)

Flat fee charged for each API call, regardless of token count. Useful for simple pricing or when token counting is impractical.

Example

Price per User (\$)

Monthly charge per unique user who accesses the model. Good for team/enterprise pricing models.

Example

Visibility

Public (Toggle)

Controls whether the model appears in the public marketplace:

- ON: Model is visible to all users and can be discovered in the marketplace
- OFF: Model is hidden and only accessible via direct link or admin assignment

Use OFF for testing new models or creating private/enterprise-only models.

Virtus AI Platform

Add AI Model - Form Field Guide

Tips for Configuring Models

1. Start with Public OFF while testing, then enable once ready for users.
2. Set reasonable Max Output Tokens to prevent excessive costs.
3. Link relevant data sources for domain-specific knowledge.
4. Choose pricing that reflects your costs plus desired margin.
5. For self-hosted models, ensure the Base URL is accessible from your server.
6. Use descriptive names that help users understand the model's purpose.
7. Write detailed descriptions including use cases and limitations.
8. Test the model thoroughly before making it public.
9. Monitor usage through Admin > Analytics after launch.
10. Consider offering a free tier or trial for new users.