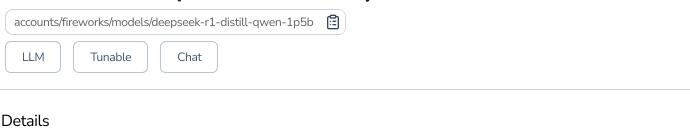






## Fireworks / Deepseek R1 Distill Qwen 1.5B



Qwen 1.5B distilled with reasoning from Deepseek R1

## Fine-tuning

**Deepseek R1 Distill Qwen 1.5B** can be fine-tuned on your data to create a model with better response quality. Fireworks uses low-rank adaptation (LoRA) to train a model that can be served efficiently at inference time.

See the Fine-tuning Guide for details.

Fine-tune this model

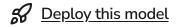
## **On-demand deployments**

On-demand deployments allow you to use **Deepseek R1 Distill Qwen 1.5B** on dedicated GPUs with Fireworks' high-performance serving stack with high reliability and no rate limits.

See the On-demand deployments  $\ ^{\square}$  guide for details.

Deploy this Base model





**Pages** 

## Model Details

Created by	yingliu@fireworks.ai
Created	1/22/2025
Visibility	Public
Kind	Base model
Model size	1B parameters
Fine-tuning	Supported
Serverless LoRA Deployment	Not supported
Function Calling	Not supported
Provider	Fireworks























Home Blog Trust Center ☐
Pricing Careers ☐ Terms

Models Privacy

Docs ☐ Licenses

Legal

Company

© 2024 Fireworks AI, Inc. All rights reserved.