

Local AI Installation Complete!

Installation Summary

✓ Ollama Successfully Installed

- Version: 0.12.3
- Location: /usr/local/bin/ollama
- Service: Running (API available at 127.0.0.1:11434)

✓ AI Models Installed

- **llama2** (3.8 GB) - Balanced performance and quality
- **mistral** (4.4 GB) - Faster responses, great for quick queries

Quick Start

Test Ollama Installation

```
./test-ollama.sh
```

Chat with AI Models

```
# Chat with llama2
ollama run llama2

# Chat with mistral (faster)
ollama run mistral
```

Useful Commands

```
ollama list      # List installed models
ollama ps       # Show running models
ollama pull [model] # Download additional models
ollama serve    # Start Ollama service
```

Using Local AI in Sports Bar AI Assistant

1. Configure AI Keys

1. Start your Sports Bar AI Assistant application
2. Navigate to `/ai-keys` page
3. Select **“Local AI”** as your provider
4. Choose your preferred model:
 - **llama2** - Better for complex queries
 - **mistral** - Faster for simple queries

2. Test AI Features

Once configured, you can use local AI for:

- **Device Configuration** - AI-powered insights and optimization
- **Log Analysis** - Intelligent log analysis without cloud APIs
- **Bartender Remote** - AI-enhanced channel recommendations
- **Sports Guide** - Team name fuzzy search with local AI

Benefits of Local AI

- ✓ **No API Keys Required** - No need for Claude, ChatGPT, or Grok API keys
- ✓ **Privacy** - All AI processing happens on your local machine
- ✓ **No Rate Limits** - Use AI as much as you want
- ✓ **Cost-Free** - No per-request charges
- ✓ **Offline Capable** - Works without internet connection

Troubleshooting

Ollama Service Not Running

```
ollama serve > /dev/null 2>&1 &
```

Check Service Status

```
ps aux | grep ollama
```

Re-install Ollama

```
./install-local-ai.sh
```

Additional Models

You can install more models from [Ollama Library](https://ollama.com/library) (<https://ollama.com/library>):

```
ollama pull codellama    # Code-focused model
ollama pull phi          # Small, fast model (2.7B)
ollama pull llama3       # Latest Llama version
```

System Requirements






- **CPU**: Multi-core processor (4+ cores recommended)
- **RAM**: 8GB minimum, 16GB+ recommended
- **Disk**: 10GB+ free space per model
- **OS**: Linux (Ubuntu 22.04+ recommended)

Performance Tips

1. **First run is slow** - Models load into memory on first use

2. **Keep service running** - Faster responses when service is persistent
3. **Use mistral for speed** - About 30% faster than llama2
4. **Close unused models** - Free up RAM: `ollama stop [model]`

Next Steps

1.  Ollama installed and tested
2.  Configure AI provider in your Sports Bar AI Assistant
3.  Test AI features in device configuration
4.  Try AI-powered log analysis
5.  Use AI features in the bartender interface

Scripts

- `install-local-ai.sh` - Install/update Ollama and models
 - `test-ollama.sh` - Quick test of Ollama installation
 - `update_from_github.sh` - Auto-installs Ollama during updates
-

Last Updated: October 1, 2025

Ollama Version: 0.12.3

Models: llama2 (3.8GB), mistral (4.4GB)