

Relay server AI



AI Overview

A relay function in an AI server acts as an intermediary that manages communication between clients and AI services, enabling secure, efficient, and scalable interactions.

Purpose of a Relay in AI Systems

A relay server serves as a middle layer between client applications and AI servers or APIs. Its main purposes include:

- **Security:** Clients do not need direct access to sensitive API keys, as the relay handles authentication and forwards requests securely .
- **Scalability:** Relays can manage multiple client connections, queue messages, and distribute workloads efficiently .
- **Reliability:** They ensure messages are delivered even if the AI server connection is temporarily unavailable, using message queues to prevent data loss .
- **Protocol Translation:** Relays can convert between different communication protocols, such as HTTP requests from clients to WebSocket connections with AI servers .

Technical Implementation

For example, the OpenAI Workers Relay uses Cloudflare Workers to implement a WebSocket relay:

- **Fetch Handler:** The entry point that intercepts HTTP requests and upgrades them to WebSocket connections if needed

Article Content

Feb 14, 2026

The 10 best AI automation tools in 2026

The 10 best AI automation tools of 2026, ranked. Compare features, pricing, and ease of use to find the right fit for your

Jan 06, 2026

Claude Relay MCP Server

The server supports both host and client modes for cross-machine communication, featuring advanced capabilities like

Dec 25, 2025

Building serverless apps with the OpenAI Realtime API

In early October, OpenAI released the Realtime API, designed for building advanced multimodal conversational

Sep 11, 2025

Unity MCP Server: Connect Claude Code, Cursor, and Other AI

In today's article on the Unity AI Open Beta, learn how to connect Claude Code, Github Copilot, and other AI agents

Jul 14, 2026

Email Relay Server Setup

Technical AI Prompt Email Relay Server Setup Outline a plan to build an email relay server handling forwarding, routing, security,

May 21, 2026

Relay AI — Your Gateway to Every AI Model

Introducing Relay AI — One Platform, Every AI Model We built Relay AI so you can access GPT-4o, Claude, Gemini, Midjourney,

Dec 20, 2025

Relay: Local MCP Server for AI Agents & Data Workflows

Relay provides a unified local Multi-Agent Compute Platform (MCP) server, empowering AI agents with 40+ tools for file, image,

Jul 29, 2025

Relay (relay.app)

Relay.app is a platform to create AI agents that work for you across Gmail, Notion, HubSpot, and 100+ other apps.

Oct 29, 2025

OpenClaw Browser Relay Guide 2026: Master CDP Protocol,

OpenClaw Browser Relay enables AI agents to control real web browsers through the Chrome DevTools Protocol

Jul 23, 2026

How to Build a Serverless Voice AI Assistant for Telephony in AWS

Learn how to build a real-time, serverless Voice AI assistant for telephony using Twilio ConversationRelay and AWS

Aug 12, 2025

ai-relay · PyPI

WebSocket relay that bridges AI coding agent CLIs (Claude Code, Codex, Gemini CLI, Snowflake Cortex) to any web

Jun 05, 2026

China warns of security risks in relay services for foreign AI models

China's national security authority has warned of risks in using “ artificial intelligence relay services” that provide access

Mar 25, 2026

RelayPlane | Cut Your Agent API Costs by 90%

RelayPlane spreads load across multiple credentials, fails over before 429s, and stops runaway agents at the budget cap. One

Mar 11, 2026

Conversation Relay | Twilio

Conversation Relay is part of the Conversations layer of the Twilio platform. You can turn

Oct 26, 2025

Azure Relay documentation | Microsoft Learn

Azure Relay documentation Learn how Azure Relay service allows you to securely expose services that run in your corporate

Feb 07, 2026

What Are Relays Servers for Multiplayer Games? A Peer-to-Peer

Read all about What Are Relays Servers for Multiplayer Games? A Peer-to-Peer Networking Guide on Edgegap's blog.

Nov 30, 2025

Building serverless apps with the OpenAI Realtime API

The relay server runs the provided relay.js code and holds an OpenAI API key to handle Realtime API connections.

Aug 08, 2025

What is Azure Relay?

This article provides an overview of the Azure Relay service, which allows you to develop cloud applications that

Oct 09, 2025

Fabler Relay: Human-in-the-Loop Request Queue for AI Agents

Fabler Relay: a secure human-in-the-loop request queue for autonomous AI agents. Agents escalate tasks for human intervention;

Mar 05, 2026

Chat Relay: OpenAI-Compatible Relay for AI Chat Interfaces

Chat Relay is a system that allows Cline/RooCode to communicate with web-based AI chat interfaces (like Gemini, AI Studio,

Aug 07, 2025

Relay Server

Setting up a Relay Server Connecting directly to tensorstudio through SDK on browser exposes your api-key. If you are running your

Mar 27, 2026

RelayCloud — Remote AI Agents, Local Feel

Deploy OpenCode on powerful remote servers, manage all your AI assistants from one dashboard, and connect securely through our

Nov 23, 2025

relay-server · GitHub Topics · GitHub

A tiny relay server that bridges two WebSocket connections, allowing the clients to talk directly to each other.

Contact Us

For more information, pricing, or custom solutions, please contact us:

Website: <https://ntmmgroup.co.za>

Email: sales@ntmmgroup.co.za

Phone: +27 82 471 2596

Address: Unit 7, Optic Park, 24 High-Tech Lane, Century City, Cape Town, 7441, South Africa

This document is for informational purposes only. Specifications subject to change without notice.

