

Developing Applications in the Age of AI

Reacting to changing AI technology trends

The world of artificial intelligence (AI) is seeing rapid progress with AI Agents, large language models, machine learning, and computer vision. New AI tools, such as MCP Servers for AI Agents, Cursor, ChatGPT and Github CoPilot have emerged, sparking widespread interest with developers.

Whether starting with boiler plate code, configuration and implementation, or running inference models, AI has implications both in assisting in development as well as generating code or content and taking actions automatically.

When adopting new AI technologies, engineering teams can face a series of challenges including:

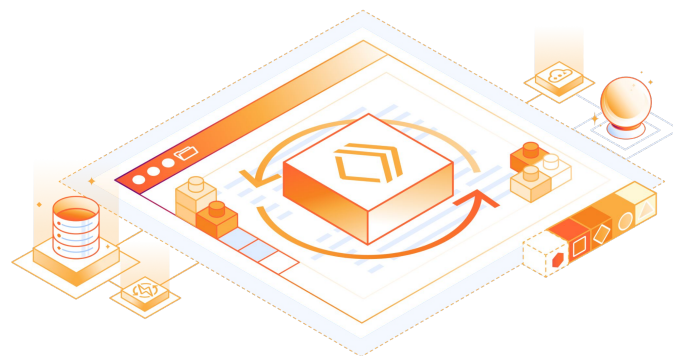


Figure 1: Cloudflare's developer platform has all the tools to build & deploy AI apps



Maintaining & scaling AI infrastructure

Managing infrastructure to support AI use cases is time consuming & costly.

Lack of AI talent is making it more difficult and expensive to build AI applications.



Data protection and security

Without the right data protection and security, integrating AI can create significant risks for business.

From mobile phones to cloud computing, new security threats such as prompt injection, harmful content, expose private data.



Managing excessive cloud spend & vendor lock-in

Excessive cloud costs are now slowing investments in development and speed of innovation.

Closed source AI models can lead to vendor lock-in & higher operational risk.



Legacy code & disparate tools slows time to market

Legacy code requires time-wasting deployment steps & lacks adaptability in integrating different tools (storage, AI models, APIs/services). Developers need a straight line from writing or updating code to production deployment.

Integrate AI Capabilities with Cloudflare's Developer Platform

Accelerate AI integrations faster and more efficiently

Write code and ship products faster than ever before

With Cloudflare's AI developer suite, developers can globally deploy AI applications with a few lines of code. Cloudflare's managed AI inference infrastructure has the latest open-source AI models available so developers can concentrate on shipping the latest technology without the hassles of maintaining AI infrastructure.

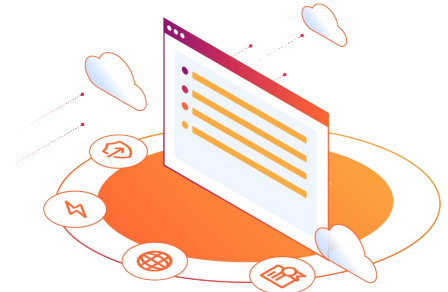


Figure 2: compute services



Workers AI For AI Inference-as-a-Service

Workers AI enables developers to run fast, low-latency Open Source AI models (i.e. Llama) and inference tasks on Cloudflare's globally distributed serverless infrastructure.



Vectorize & AutoRAG

Begin building for free in minutes. Vectorize enables fast & cost-effective vector storage. With AutoRAG, there is seamless integration with Workers AI, Vectorize & R2 so developers can build search & Retrieval Augmented Generation (RAG) applications in 4 easy steps.



AI Gateway

Cloudflare AI Gateway provides centralized visibility and control for your AI applications. Connect your apps with a single line of code to monitor usage, costs, and errors. Includes prompt/response guardrails to prevent private data, harmful content with AI applications.



R2 Object Storage Zero Egress

Manage storage costs and scalability with R2 object storage. Store everything from media files, data logs to application-specific metadata for AI/ML training. Consistent low latency, high durability, and limitless capacity.



Agent SDK & MCP Server

The Agent SDK lets you build AI-powered, stateful applications with built-in state management, real-time communication, and seamless deployment of remote MCP servers on Cloudflare's global network.



Zero Trust Security & Firewall for AI

Cloudflare for AI, a suite of Zero Trust security controls, enables enterprises to safely and securely use the latest generative AI tools without putting intellectual property and customer data at risk. Firewall for AI will detect shadow AI and PII leakage.

Cloudflare Developer Platform Product Suite

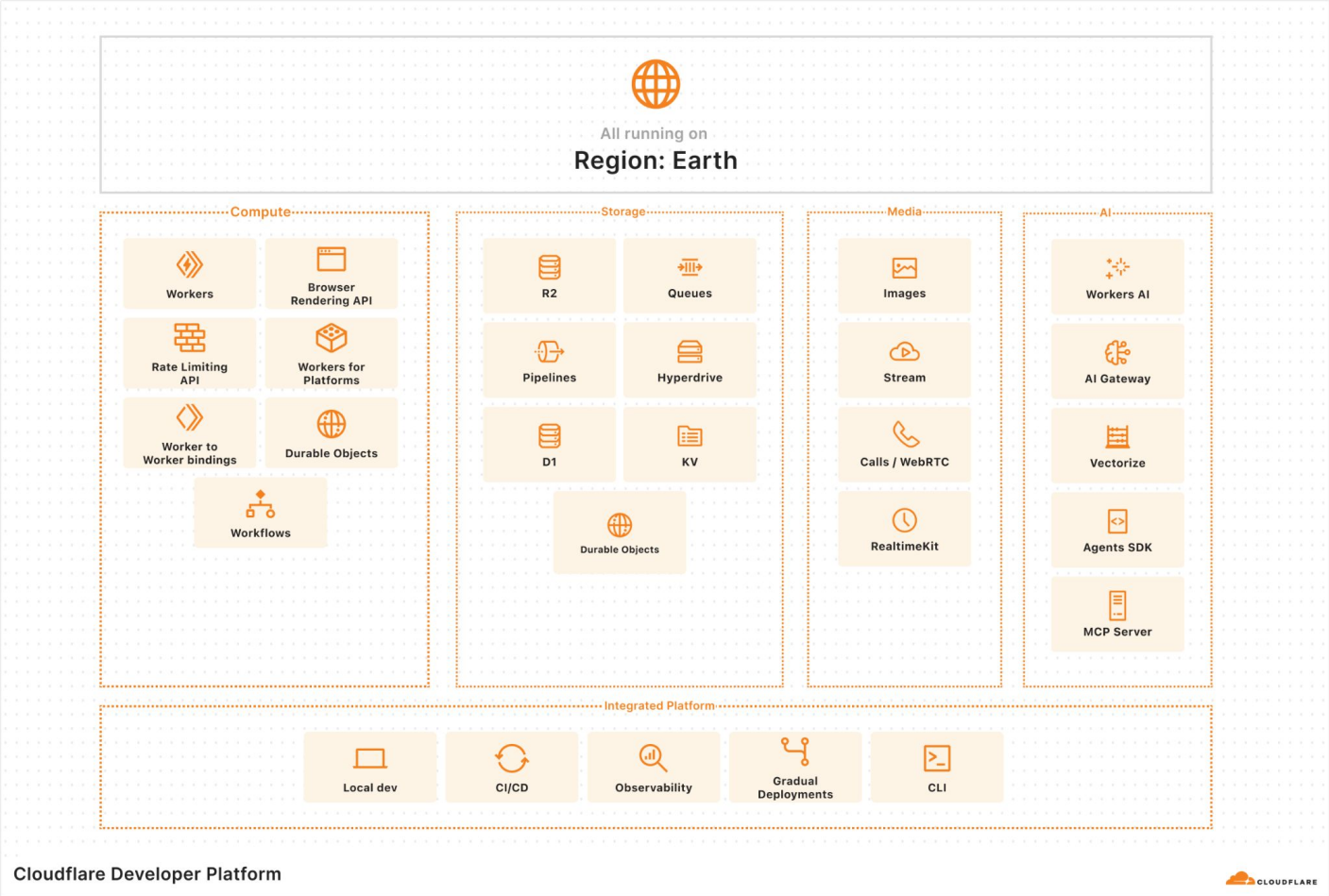


Figure 1: Unified Developer Platform

The platform for building AI Agents

Build agents on Cloudflare — the platform designed for durable execution, serverless inference, and pricing that scales up (and down). The whole tool set allows companies to deploy AI Agents and AI applications securely. With support for AI models, vector databases, object storage, external integrations, and seamless deployment, Cloudflare offers unmatched flexibility and speed for building modern, AI-driven experiences.