

MCP Gateway for AI agents

Let AI agents act — with an identity, an allow list and a person for what matters

Zero trust for AI agents: every tool an agent calls through your MCP servers is checked against its allow list, approved by a person where it matters, and recorded.



TotalHeal
AgentGate

Contents

1	Executive summary	6	Security and data handling
2	The challenge	7	Where it fits
3	Architecture	8	Deployment and system requirements
4	How it works	9	Technical specifications
5	Capabilities in detail	10	Getting started

1 Executive summary

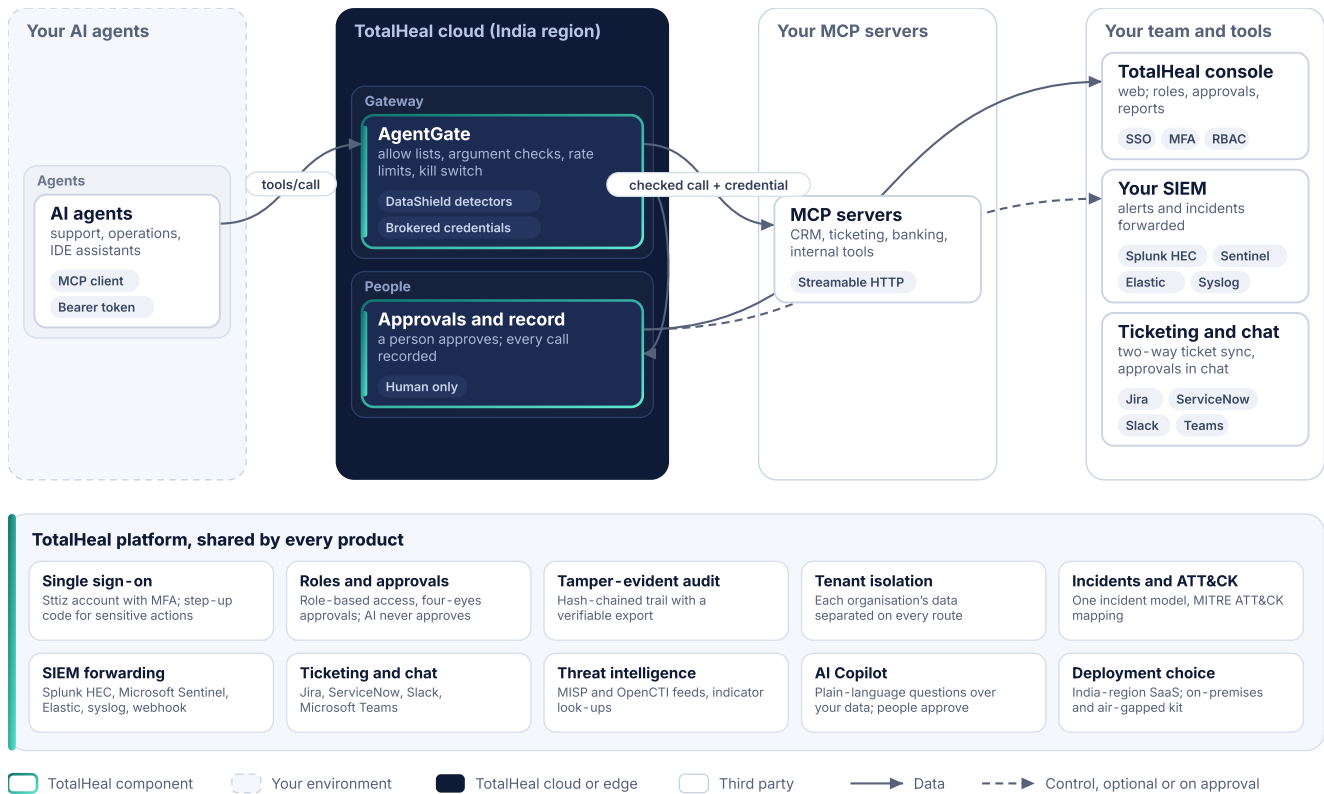
AgentGate puts every AI agent behind its own identity and every MCP tool call behind a check: is this agent allowed this tool, do the arguments carry sensitive data, does a person need to approve it? Server credentials stay with AgentGate, and every call is recorded.

2 The challenge

AI agents now act — they read customer records, send e-mails, issue refunds and change systems through MCP servers. Each agent usually holds a server token that can do everything, nobody can say which agent called what, and a single prompt injection can turn a helpful agent into one that closes accounts or mails out customer data.

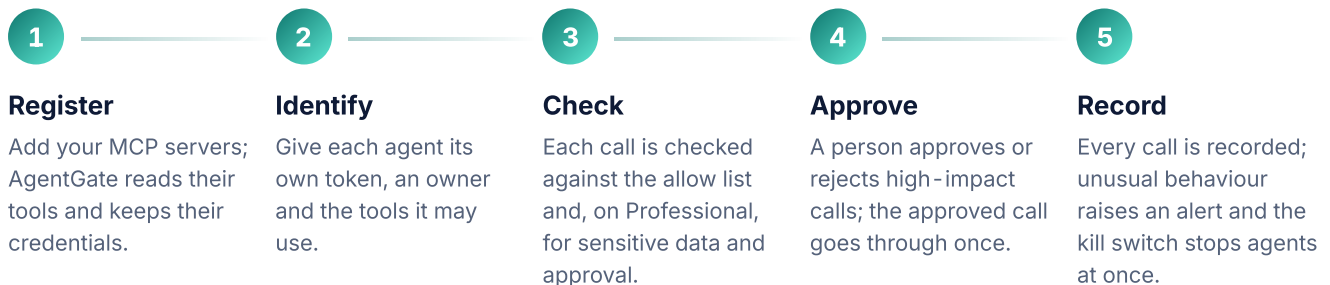
The TotalHeal approach. AgentGate sits between your AI agents and your MCP servers. Each agent gets its own identity and token and sees only the tools it is allowed; the servers' own credentials stay with AgentGate. Every tool call is checked against the agent's allow list — and, on the Professional tier, for sensitive data and for tools that need a person's approval — then forwarded and recorded.

3 Architecture



Agents call AgentGate with their own token; AgentGate checks each tool call, adds the server's credential and forwards it; people approve the calls that need it.

4 How it works



5 Capabilities in detail

- Registry of MCP servers**

Register each server's endpoint once; AgentGate reads its tools and how each declares itself (read only, or changes and deletes data). Turn any tool off for everyone, and pause a whole server.

Credentials agents never hold

The server's token is kept encrypted by AgentGate and added to each forwarded request. Agents authenticate to AgentGate with their own token; revoking it cuts the agent off without touching the server.



An identity per agent

Every agent has a name, an accountable owner, a token shown once (only its fingerprint is kept) and an allow list of servers and tools. When an agent lists tools it sees only the ones it may call.



A person approves what matters

Tools that change or delete data can require approval. The call waits; a person approves or rejects it in the console with a note, and that exact call — same tool, same arguments — goes through once within 15 minutes. An AI agent can never approve.



Unusual behaviour

After a learning week, an agent that calls a tool it has never used, or makes three times more calls in an hour than its busiest hour so far, raises an alert in the same incident queue as everything else.



Arguments checked for sensitive data

Before a call is forwarded its arguments are inspected with the DataShield detectors — Aadhaar, PAN, bank accounts and IFSC, card numbers, cloud keys and passwords — and stopped at the threshold you set per server, with a message the agent can act on.



Rate limits and a kill switch

Each agent has its own limit of tool calls a minute. Disable one agent, or stop every agent at once, and requests are refused immediately.



A record of every call

Agent, server, tool, decision and reason, argument names and sizes (never their values), sensitive data found (masked), result and latency — for audits and for investigating what an agent did.

The screenshot displays the TotalHeal AgentGate web interface. On the left is a dark sidebar with navigation links for various security products. The main panel shows the 'Tool calls' section under the 'AgentGate' tab. It includes a search bar, a filter dropdown set to 'Every decision', and a table of tool call logs. The table columns are: When, Agent, Server / tool, Arguments, Decision, Why, and Latency. The logs show various actions by 'Support copilot' and 'Reconciliation agent' agents, with some calls being blocked due to sensitive data or tool restrictions.

When	Agent	Server / tool	Arguments	Decision	Why	Latency
1m ago	Support copilot	Core banking tools issue_refund	amount, reason, transaction	Waiting for approval	Waiting for approval AGA-E99EBE7AC41F	20,046 ms
1m ago	Support copilot	Core banking tools issue_refund	amount, reason, transaction	Allowed	—	0 ms
1m ago	Support copilot	Core banking tools issue_refund	amount, reason, transaction	Waiting for approval	Waiting for approval AGA-3BC19E832F48	20,047 ms
1m ago	Support copilot	Core banking tools close_account	customer_id	Blocked	Tool not allowed for this agent	0 ms
1m ago	Support copilot	Core banking tools send_email	body, to	Blocked	Sensitive data in the arguments (high) · Email address c*****@example.in; Indian bank account number *****7812; IFSC code *****1234; PAN (Permanent Account Number) *****4F	0 ms
1m ago	Reconciliation agent	Core banking tools issue_refund	amount, transaction	Blocked	Tool not allowed for this agent	0 ms
1m ago	Reconciliation agent	Core banking tools search_transactions	from, status, to	Allowed	—	0 ms
1m ago	Support copilot	Core banking tools search_transactions	customer_id, days	Allowed	—	0 ms
1m ago	Support copilot	Core banking tools get_customer	customer_id	Allowed	—	0 ms

Tool calls through AgentGate for a demo fintech's support and reconciliation agents. Real screenshot, demo data.

6 Security and data handling

Agents never hold server credentials

AgentGate keeps them encrypted and adds them to each forwarded request.

Argument values are not kept

The record keeps argument names and sizes and masked sensitive findings only.

An AI never approves

Approvals come from people signed in to the console, never from agent or service identities.

Revocation is immediate

A deleted, disabled or stopped agent is refused on its next request.

7 Where it fits

Support agents with refunds

Let an agent look up customers freely while every refund waits for a person.

Data that must not leave

Stop an agent e-mailing a customer's PAN or bank details, whatever its prompt says.

Agent sprawl

Know which agents exist, who owns them and what each actually called.

8 Deployment and system requirements



Cloud (SaaS)

Available

The gateway runs in the TotalHeal cloud and calls your MCP servers over HTTPS.



On-premises

Available

Delivered with our team from the on-premises kit.



Air-gapped

Available

Offline kit, offline licence file and offline update bundles.

Protected platforms



MCP
servers
(Streamable
HTTP)



MCP
clients with
bearer
tokens

On-premises server

Operating system

RHEL, Rocky Linux or AlmaLinux 8 / 9; Ubuntu 20.04, 22.04 or 24.04; Debian 11 / 12

Processor

4 vCPU (tested configuration)

Memory

8 GB recommended (6 GB minimum)

Disk

10 GB free for software, releases and backups, plus data

Sign-in

Active Directory / LDAP, SAML 2.0 or OpenID Connect

Software

Network and access

- MCP servers reachable over HTTPS that speak Streamable HTTP
- Agents whose MCP client can send a bearer token

Alerts and tickets to

splunk> Splunk

Microsoft
Sentinel

Elastic

Syslog /
CEF

Jira

ServiceNow

Slack

Microsoft
Teams

Nothing to pre-install: no
container runtime, database, web
server or Java



9 Technical specifications

Features	Included		
Registry of MCP servers	✓ Included		
Credentials agents never hold	✓ Included		
An identity per agent	✓ Included		
Arguments checked for sensitive data	✓ Included		
A person approves what matters	✓ Included		
Rate limits and a kill switch	✓ Included		
Unusual behaviour	✓ Included		
A record of every call	✓ Included		
Availability by tier	Essential	Professional	Elite
Registry of MCP servers with their tools read automatically; server credentials kept encrypted and added by AgentGate — agents never hold them	✓ Included	✓ Included	✓ Included
An AI agent identity per agent: its own token, an owner, an allow list of servers and tools; the tool list each agent sees is filtered to it	✓ Included	✓ Included	✓ Included
Every tool call recorded: agent, server, tool, decision, argument names and sizes, result, latency	✓ Included	✓ Included	✓ Included
Arguments checked with the DataShield detectors and blocked at the threshold you set per server	—	✓ Included	✓ Included
High-impact tools held for a person's approval (an AI never approves); rate limits per agent	—	✓ Included	✓ Included

Alerts on unusual agent behaviour (a tool never used before, a burst of calls); kill switch for one agent or all	—	—	 Included
List price per AI agent per year	US\$180 · ₹17,130	US\$360 · ₹34,260	US\$600 · ₹57,100
Platforms and coverage			
MCP servers	Model Context Protocol over Streamable HTTP (JSON and event-stream answers); tools/list and tools/call inspected, other messages passed through		
Agents	Any MCP client that sends a bearer token (custom agents, agent frameworks, IDE assistants)		
Console	TotalHeal web console		
Deployment options			
SaaS (India region)	Available The gateway runs in the TotalHeal cloud and calls your MCP servers over HTTPS.		
On-premises control plane	Available Delivered with our team from the on-premises kit.		
Air-gapped	Available Offline kit, offline licence file and offline update bundles.		
Licensing			
Priced per	AI agent per year		
Counted	AI agents registered in AgentGate		
Trial	30 days, up to 5 agents		
Editions	AgentGate is sold on its own, outside the editions.		
Government profile	Government is a profile, not a separate tier or price list: the Elite tier with Government defaults switched on — data residency in India, four-eyes approval of policy changes, a distinct approver for every approval-level action, retention of at least 180 days, and the CERT-In and DPDP reporting clocks.		
Procurement dossier	Compliance matrix, bill of materials, architecture and hardening, SLA and OEM authorisation letter, on request		
On-premises server			
Operating system	RHEL, Rocky Linux or AlmaLinux 8 / 9; Ubuntu 20.04, 22.04 or 24.04; Debian 11 / 12		
Processor	4 vCPU (tested configuration)		
Memory	8 GB recommended (6 GB minimum)		
Disk	10 GB free for software, releases and backups, plus data		
Sign-in	Active Directory / LDAP, SAML 2.0 or OpenID Connect		

Software	Nothing to pre-install: no container runtime, database, web server or Java
Requirements	
Requirement 1	MCP servers reachable over HTTPS that speak Streamable HTTP
Requirement 2	Agents whose MCP client can send a bearer token

10

Getting started

- MCP servers reachable over HTTPS that speak Streamable HTTP
- Agents whose MCP client can send a bearer token

Start a 30 days, up to 5 agents trial in the TotalHeal console with a Sttiz account. For a quote, an on-premises installation or the procurement dossier (compliance matrix, bill of materials, architecture and hardening, SLA, OEM authorisation letter), write to business@totalhealsecurity.com.

About TotalHeal

TotalHeal is a security platform from Freaquer Corporation Pvt Ltd: endpoint, network, application, cloud, security-operations and backup products, each sold on its own and run from one console.

Datasheet

totalhealsecurity.com/datasheet/agentgate

No third-party certification is claimed. Figures are from our own tests where stated, with their dates. Specifications can change; this whitepaper is dated 27 October 2026.