

AI Security (GenAI firewall)

Use generative AI without handing it your customers' data

See which AI assistants your people use and block the ones you have not approved; put your own AI applications behind a gateway that stops sensitive data and injection attempts.



TotalHeal
PromptShield

Contents

1 Executive summary

2 The challenge

3 Architecture

4 How it works

5 Capabilities in detail

6 Security and data handling

7 Where it fits

8 Deployment and system requirements

9 Technical specifications

10 Getting started

1 Executive summary

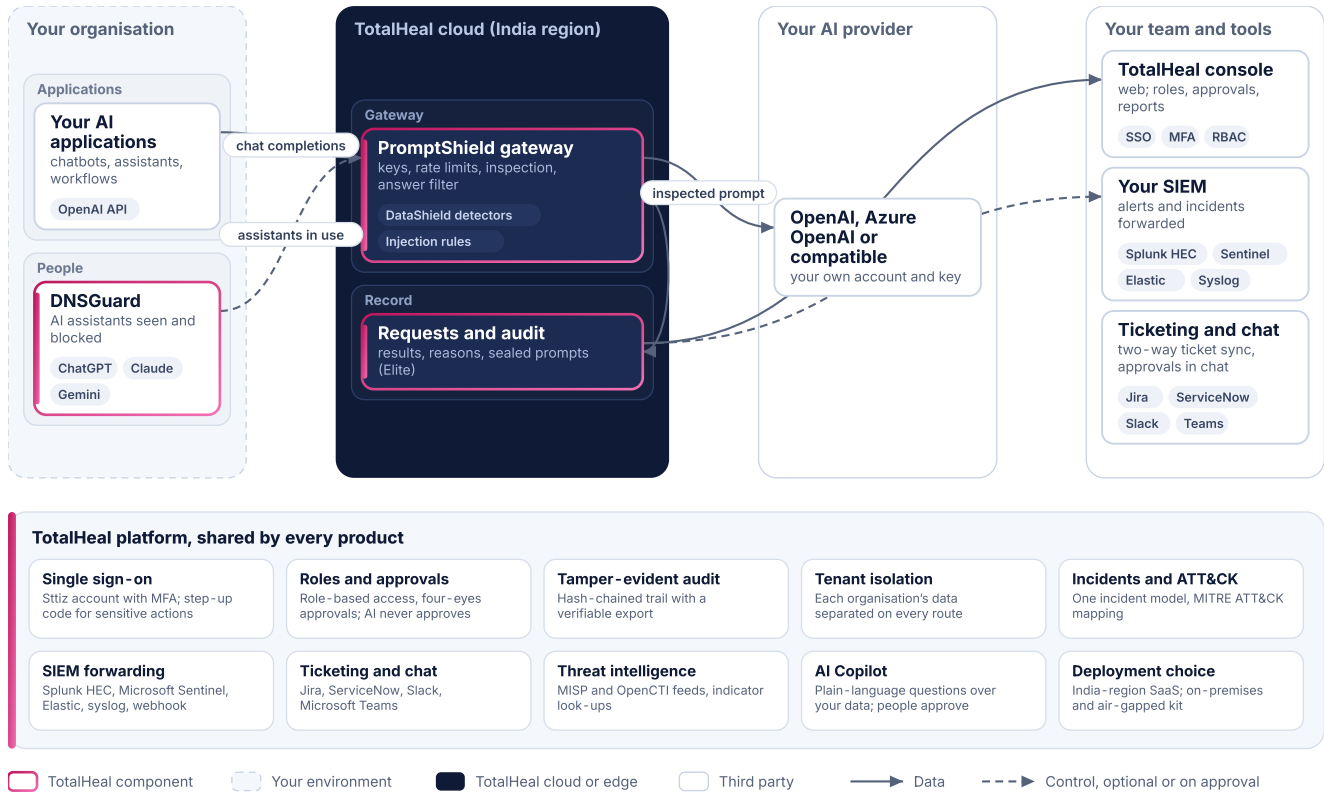
PromptShield shows which AI assistants your people use and blocks the ones you have not approved. Your own AI applications call it like the OpenAI API: every prompt is inspected before it reaches your provider, sensitive data is redacted or blocked, injection attempts are stopped, and every answer is checked on the way back.

2 The challenge

People paste contracts, customer records and source code into whichever AI assistant is open, and the company's own AI applications send whatever their users type straight to a model provider. Nobody can say which assistants are in use, what data has left, or whether someone is trying to talk an internal assistant into ignoring its instructions.

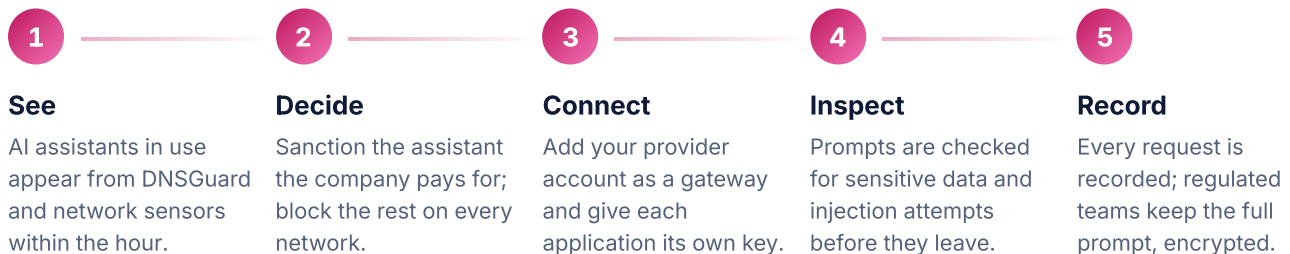
The TotalHeal approach. PromptShield shows which AI assistants are used on your networks and lets you sanction or block each one. For your own applications it is a gateway: they call PromptShield exactly as they would call the OpenAI API, and every prompt is inspected before it reaches your provider — sensitive data redacted or blocked, injection attempts stopped — and every answer is checked on the way back.

3 Architecture



Your applications call PromptShield instead of the provider; each prompt is inspected, redacted or blocked, and only then sent to your own provider account. DNSGuard reports and blocks the AI assistants people use.

4 How it works



5 Capabilities in detail

- AI assistants in use**
 ChatGPT, Claude, Gemini, Microsoft Copilot, DeepSeek, Perplexity and other assistants found from DNSGuard's hourly counts and network-sensor events: how many people, how often, since when. Sanction the one the company pays for; block the rest through DNSGuard on every network.
- A gateway that speaks the OpenAI API**
 Applications change only the base URL and the key. Chat completions and the model list are supported; streaming requests are answered as server-sent events after the answer is inspected.



Your provider, your account

OpenAI, Azure OpenAI or any OpenAI-compatible endpoint — Anthropic's compatibility endpoint, Mistral, Groq, a self-hosted vLLM or Ollama. The provider API key is kept encrypted and never shown again; requests leave through TotalHeal's egress guard.



Answers checked too

High and critical sensitive data in a model's answer is replaced before the application sees it.



Per-model policy

Allow only the models you have approved on each gateway, and cap the tokens of the expensive ones.



Every request on record

Application, end user, model, result, reasons, sensitive data found (masked), latency and tokens. For regulated teams the full prompt and answer are kept encrypted, and opening one is itself recorded in the audit log.



Sensitive data never reaches the model

Every prompt is inspected with the DataShield detectors — Aadhaar, PAN, GSTIN, card numbers, bank accounts and IFSC, passports, cloud keys, tokens and passwords — and allowed, redacted or blocked by severity. The application receives a coaching message to show the person.



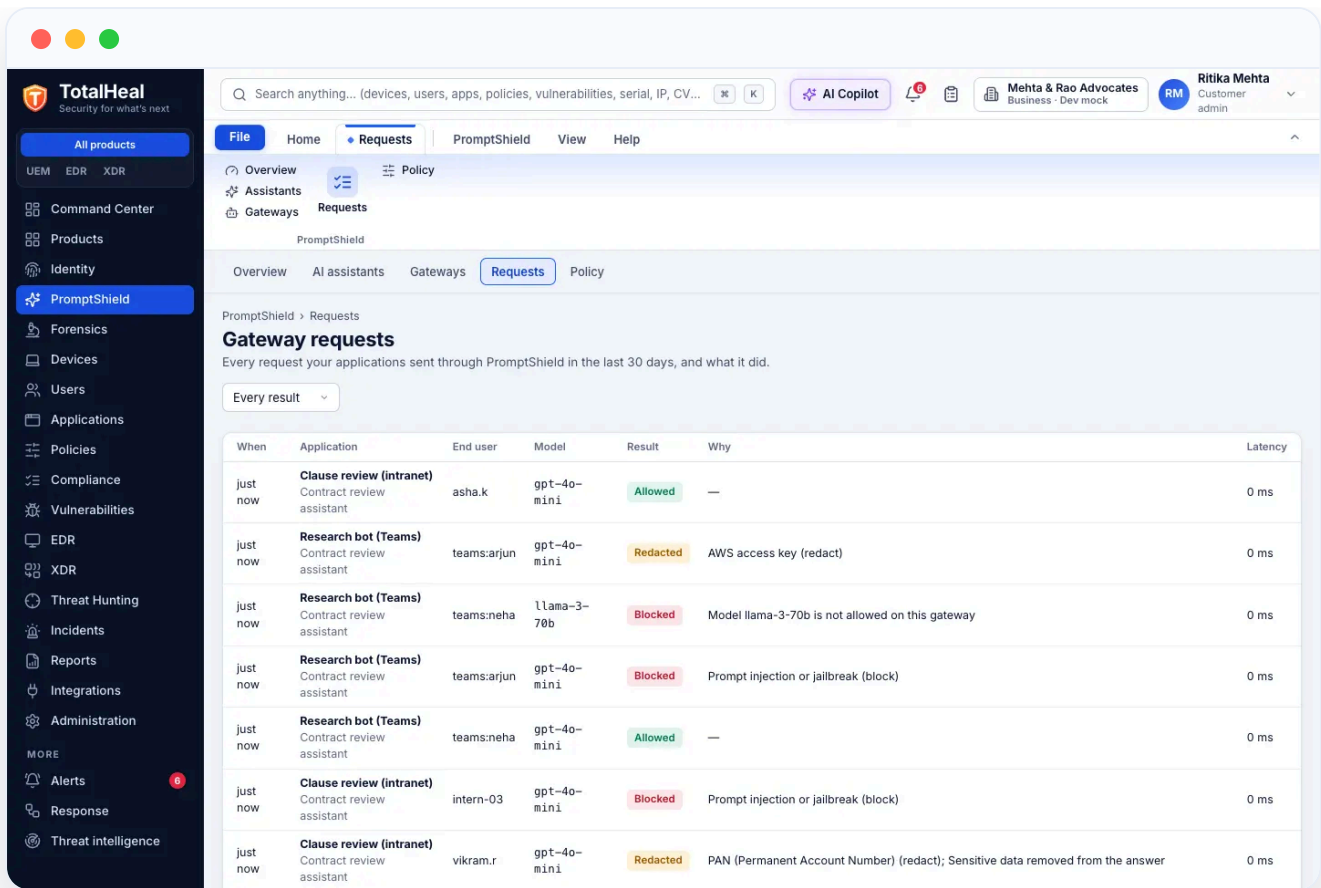
Prompt injection and jailbreaks

A curated rule set of published attack patterns — instructions to ignore earlier instructions, requests for the system prompt, “developer mode” and DAN-style role play, forged chat-template markers, hidden characters, encoded payloads and data-exfiltration links — scored per prompt; above your threshold the prompt is blocked or flagged.



One key per application

Each application gets its own key with its own rate limit; revoke one without touching the others. Only a fingerprint of each key is kept.



Gateway requests in PromptShield for a demo law firm's contract-review assistant. Real screenshot, demo data.

6 Security and data handling

Data checked before it leaves

A prompt reaches your provider only after inspection, with what the policy removes already removed.

Keys you can revoke one by one

Each application has its own key; only its fingerprint is kept.

Your provider key stays sealed

It is kept encrypted, never shown again, and used only to call your provider.

Opening a prompt is audited

Reading a recorded prompt is itself written to the audit log.

7 Where it fits

Client confidentiality

Law firms and banks keep Aadhaar, PAN and card numbers out of model providers' logs.

Internal assistants

Stop employees and outsiders from talking a support bot into revealing its instructions.

Shadow AI

Find the AI assistants nobody approved and block them.

8 Deployment and system requirements



Cloud (SaaS)

Available

The gateway runs in the TotalHeal cloud and calls your provider from there.



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

 OpenAI

 Azure OpenAI

 OpenAI-compatible endpoints (Anthropic, Mistral, Groq, vLLM, Ollama)

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



Network and access

- DNSGuard for discovering and blocking AI assistants
- Your own account with OpenAI, Azure OpenAI or an OpenAI-compatible provider for the gateway

Alerts and tickets to

 Splunk

 Microsoft Sentinel

 Elastic

 Syslog / CEF

 Jira

 ServiceNow

 Slack

 Microsoft Teams

9 Technical specifications

Features	Included
AI assistants in use	✓ Included
A gateway that speaks the OpenAI API	✓ Included
Your provider, your account	✓ Included
Sensitive data never reaches the model	✓ Included
Answers checked too	✓ Included
Prompt injection and jailbreaks	✓ Included

Per-model policy	✔ Included		
One key per application	✔ Included		
Every request on record	✔ Included		
Availability by tier			
	Essential	Professional	Elite
AI assistants in use (ChatGPT, Claude, Gemini, Copilot, DeepSeek and more) from DNSGuard and network sensors; sanction or block each through DNSGuard	✔ Included	✔ Included	✔ Included
AI gateway for your own applications: OpenAI-compatible endpoint, one key per application with its own rate limit, your provider key kept encrypted	—	✔ Included	✔ Included
Prompts inspected with the DataShield detectors: sensitive data allowed, redacted or blocked by severity, with a coaching message	—	✔ Included	✔ Included
Sensitive data removed from answers; every request recorded; alerts for blocked requests	—	✔ Included	✔ Included
Prompt-injection and jailbreak detection, allowed models and token caps per model, full prompt audit kept encrypted	—	—	✔ Included
List price per user per year	US\$18 · ₹1,710	US\$36 · ₹3,430	US\$54 · ₹5,140
Platforms and coverage			
Discovery and blocking	DNSGuard (cloud or on-premises resolver, desktop agents); NetWatch sensors		
Gateway	OpenAI Chat Completions API (POST /v1/chat/completions, GET /v1/models), streaming included		
Providers	OpenAI, Azure OpenAI, any OpenAI-compatible endpoint (Anthropic, Mistral, Groq, vLLM, Ollama)		
Console	TotalHeal web console		
Deployment options			
SaaS (India region)	Available The gateway runs in the TotalHeal cloud and calls your provider from there.		

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	User per year
Counted	People in the organisation covered by discovery and the gateway
Trial	30 days, up to 50 users
Editions	PromptShield is sold on its own, outside the editions; blocking assistants needs DNSGuard.
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	DNSGuard for discovering and blocking AI assistants
Requirement 2	Your own account with OpenAI, Azure OpenAI or an OpenAI-compatible provider for the gateway

10 Getting started

- DNSGuard for discovering and blocking AI assistants
- Your own account with OpenAI, Azure OpenAI or an OpenAI-compatible provider for the gateway

Start a 30 days, up to 50 users 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/promptshield

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.