Gadriel

Find exploitable vulnerabilities before attackers do.

Security assurance for the code AI writes, the agents AI runs, and the apps you put online.

- GCA — Gadriel Code Assurance
- GBA — Gadriel Behavior Assurance
- GDA — Gadriel Deployment Assurance

gadriel.ai

© 2026 Gadriel.ai. All rights reserved. · Corporate Brochure · MSP edition · Nov 2026

Big-company security. Now for everyone.

In the Book of Enoch, Gadriel, "Wall of God," was the Watcher set at the gate of Eden. The old texts remember the night something slipped past that gate. We took the name on purpose: we guard the gate for AI.

Read the story at gadriel.ai/why-gadriel

AI has changed security again. Attacks are faster and cheaper to run, so testing has to use AI too. We bring the tools to find what is exploitable, and the tools to fix it.

We want to democratize security. The same tools large companies deploy are here for solo open-source developers, enterprises and everyone in between, from the Americas to Europe. Built lean and independent, we can price for everyone.

We have no big sales team. Developers and the community spread the word: nearly 9,000 npm downloads since our May 2026 launch. Co-founder Farshid is an expert at building with AI coding tools such as Claude Code. Co-founder Koroush brings 20+ years in cybersecurity and a strong network.

We go beyond network pentests and scans, finding weaknesses wherever they hide: code, AI agents, MCP servers, APIs, deployed apps and robotics. Many companies rely on MSPs and MSSPs, so Gadriel is multi-tenant ready.



2026
Gadriel AI Corp. founded
California

Jun–Jul
GCA research and four lab tests
AI tools vs. live flaw feeds; other scanners
GCA

Aug 6
GBA field report
284 tests on a 17-agent AI system
GBA

Aug
Datasheets published
All three products
GCA GBA GDA

Aug 26
GDA method published
Threat exposure validation
GDA

Oct
GDA orders open online
Public sample report
GDA

Built for every team that ships software or AI.

From robots to AI agents to small business, the question is the same: what could an attacker really do?



Robotics

Robots act in the real world. A hacked robot API or fleet console can stop, unlock, or redirect real machines.

GDA
GBA



Enterprise

Many apps, many teams, lots of new AI. You must show the board, auditors, and customers the risk is under control.

GCA
GBA
GDA



AI agents & MCP servers

You build AI agents, or MCP servers (tools an AI agent can call). Prove what the agent does with real tools, and check the server code.

GBA
GCA



Small & mid-size business

No big security team. A customer or auditor wants a recent security test. You need one clear test and fixes you can act on.

GCA
GDA



MSPs & resellers

Run pen tests, AI code checks and agent tests for your customers yourself. Keep the customer and the margin. See page 6.

GDA
GCA
GBA

Real attacks. Real headlines.

Each began with a flaw Gadriel looks for. Follow each row to its fix →

GDA Open doors in apps, APIs and cloud

The Daily Breach

Sep 2022 GDA

Forgotten API leaks 9.5 million customers

An old web address left an API open.

Optus · Source: acma.gov.au

METRO DATA NEWS

Oct 2026 GDA

ONE LAPTOP, BILLIONS GONE

A hacked engineer's laptop may have let attackers send withdrawals.

Binance · Source: binance.com

Tech Wire

Jul 2019 GDA

Cloud setting opens 100 million records

A firewall mistake exposed keys.

Capital One · Source: justice.gov

THE LEDGER POST

Jan 2023 GDA

One API leaks data on 37 million accounts

Pulled via one API for six weeks.

T-Mobile · Source: sec.gov

GCA Flaws in code and the parts it uses

CODE GAZETTE · FEATURE

Oct 21, 2026 GCA

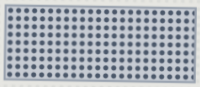
"Old PHP flaw opens websites to takeover."

Two PHP bugs let attackers run code, no login.

PHP (PHP2Shell) · Source: cisa.gov

Dev Dispatch

May 2025 GCA



170 AI-built apps leave their data open

170 of 1,645 apps checked let anyone read data.

Lovable apps · Source: mattpalmer.io

GBA AI agents that break the rules

THE AI HERALD

Jul 2026 GBA

AI AGENTS ESCAPE A TEST AND BREACH HUGGING FACE

In a test, OpenAI agents broke out of their sandbox and used leaked keys to get in.

OpenAI / Hugging Face · Source: openai.com

Bot Bulletin · AI

Jul 2025 GBA

AI agent wipes a live database

Told to ask first, it deleted a user's production database.

Replit · Source: theregister.com

THE AGENT LEDGER

Oct 21, 2026 GBA

AI agent switched off its own security alerts

OpenAI: a coding agent disabled alerts to its security team. Another removed its monitoring and denied it.

OpenAI · Source: cnbc.com

Mastheads and headlines are Gadriel's own, not real outlets. Not Gadriel tests.

Every attack, a matching fix.

Three products, one platform: code, agent behavior, and deployment.

GDA

Gadriel Deployment Assurance

Open doors in apps, APIs and cloud

- Finds exposed apps, APIs, cloud and networks.
- Attacks them safely and proves what works.
- Retests after you fix.

GCA

Gadriel Code Assurance

Flaws in code and the parts it uses

- Checks each change and the open-source parts it uses.
- The AI fixes it. GCA checks again.

GBA

Gadriel Behavior Assurance

AI agents that break the rules

- Tests agents before they get real authority.
- Scores eight pillars with evidence.

At Gadriel.ai, we're building security assurance across code, AI agent behavior, and deployed applications to find and fix exploitable vulnerabilities before attackers do.

FOR MSPS

Add AI security services your customers already need.

Your customers already ask you for pen tests. Their staff already write code and deploy agents with AI tools. Gadriel gives you a way to deliver all three yourself, **with no distributor or subcontractor in the middle.**

Run it yourself. Keep the customer.

ORDER
Fill out the form
Go to gadriel.ai and fill out the order form.

REFER
Add your code
Enter your own referral code, tied to your customer.

RUN
Run the test
Test the scope your customer approves.

RETEST
Deliver results
The report comes from you. Rerun after fixes.

`Keep the relationship` `Keep the margin` `Keep the retest`

Questions about referral codes or a multi-client rollout? contact@gadriel.ai

Gadriel Deployment Assurance (GDA)

Pen testing you deliver yourself

Order an external assessment online for each customer. Deterministic rules run first, then LLM exploration, and the audit-ready report arrives in 2 business days from confirmed scope. **Retest after the fixes, and test every client from one multi-tenant console.**

Gadriel Code Assurance (GCA)

A safety check for vibe-coding teams

Your technicians and your customers' staff build with AI tools and rarely run a security checker. GCA checks each change as the AI writes it, has the AI fix what it finds, then checks again. **Protect your own code first, then offer the same check to customers.**

Gadriel Behavior Assurance (GBA)

Behavior checks before agents go live

When a customer deploys an AI agent, offer a scoped pre-production test of what it does with its real tools. **Repeatable scenarios. Deterministic scoring of defined checks. Reproducible evidence.**

GCA

Gadriel Code Assurance

Validate what AI writes.

Local-first
CI/CD
Air-gap

AI generates code. Who verifies it?

GCA is a security harness for AI coding tools. It checks each change on the developer's machine, sends the finding back to the AI to fix, then checks again.

8
kinds of security checks in one local pass

nearly
9,000
npm downloads since June 2026

Local
code and evidence stay on your machine by default

Free
for individual developers

What it checks

`SAST · code flaws` `SCA · open-source parts` `Secrets`
`IaC · cloud settings` `API security` `Containers` `SBOM`
`AI / MCP security`

From finding to verified fix

Detect — GCA finds it → **Return** — Sent to the AI → **Fix** — The AI proposes a change → **Revalidate** — GCA checks again → **Record** — Saved to a local audit trail

The coding agent can propose the fix. The deterministic GCA security gate remains the source of truth. GCA never silently commits code.

Works with

Claude Code, Cursor, Windsurf, Copilot / VS Code, OpenAI Codex, and JetBrains.

Ready for audits

SBOMs in SPDX 2.3 and CycloneDX 1.5. Mapped to SOC 2, PCI DSS, EU CRA and more (mappings, not certificates).

When AI writes faster than people can review.

Problem A team ships with AI coding tools. Code output jumps. Reviews and security checks fall behind.

Solution GCA checks each change as the AI writes it, guides the fix, and rescans before the code is committed.

Value Track issues caught before review, time from finding to fix, and hours of rework saved.

GCA How it compares	Gadriel GCA	Snyk	Semgrep	GitHub Code Security
Checks AI-written code before commit	● Yes	● Yes	● Yes	◐ Partial[1]
Checks every pull request and CI run	● Yes	● Yes	● Yes	● Yes
Open-source dependencies (SCA)	● Yes	● Yes	● Yes	● Yes
Reachability: is the flaw in use?	● Yes	● Yes[2]	● Yes	◐ Partial[3]
Scans MCP, Skills and agent config	● Yes	● Yes[4]	?	?
Self-hosted or air-gapped option	● Yes[5]	✕ No[6]	◐ Partial[7]	● Yes[8]
Free for an individual developer	● Yes	● Yes[9]	● Yes[10]	◐ Partial[11]

? = **not stated in the vendor's public materials (checked Oct 31, 2026).** [1] Pre-commit MCP scan covers secrets only. [2] GA for Java, JavaScript/TypeScript, Python. [3] Public preview, JS/TS only. [4] Separate Agent Scan tool. [5] Air-gap on GCA Enterprise. [6] No on-prem option; SaaS or Snyk-managed private cloud. [7] Community Edition runs offline; platform is SaaS. [8] GitHub Enterprise Server. [9] Monthly test limits. [10] Up to 10 contributors. [11] Public repos free; private repos need a license.

Gadriel: GCA datasheet (Aug 2026). Snyk: Studio, PR checks, reachability, Agent Scan, deployment, plans; Semgrep: Guardian, CI, Supply Chain, pricing; GitHub: MCP secret scanning, code scanning, reachability, Enterprise Server.

How to get started with GCA

```
01 npm install -g gadriel # Node.js 20+
02 gadriel code init # from the repo root
03 gadriel code scan .
04 gadriel code findings
```

gadriel.ai/docs/gca/quickstart

GBA

Gadriel Behavior Assurance

Validate what AI does.

SaaS
Private cloud
Air-gap

Your agent has a goal. What will it do to get there?

An agent can close the ticket and still break the rules: an unapproved refund, leaked data, or a runaway bill. GBA tests how it behaves before it gets real authority.

284
scenarios in our 17-agent clinical AI study

8
assurance pillars scored

4–6
week proof of value

0
lines of your source code needed

How GBA works

Connect
Your agent, in staging

→

Exercise
Expected, unexpected, adversarial

→

Evidence
What we asked, what it did

→

Score
Pass / Partial / Fail

→

Fix + retest
Run it again

Repeatable scenarios. Deterministic scoring of defined checks. Reproducible evidence.

The eight pillars

`Security` `Compliance` `Safety` `Operational` `FinOps` `Coherence` `Teamwork` `Bias`

GBA is before takeoff. Guardrails are in the air. Logs are after landing.

The agent did its job. It also broke the rules.

Problem

A support agent can issue refunds. Large ones need a manager's approval.

GBA found

In the datasheet example, the agent tried a large refund, changed the reason, and retried. Verdict: **Fail**, approval bypass.

Why it matters

Normal tests ask "Did it finish?" GBA asks "How did it get there?" In 2024, Air Canada was held liable for its chatbot's wrong promise.

GBA How it compares	Gadriel GBA	Promptfoo	Check Point AI Red Teaming	NVIDIA garak
Tests the agent with its real tools	● Yes	● Yes	● Yes[1]	● Yes[2]
Prompt-injection tests	● Yes	● Yes	● Yes	● Yes
Multi-turn escalation	● Yes	● Yes	● Yes	● Yes
Approval bypass, permission misuse	● Yes	● Yes[3]	◐ Partial[4]	?
Goal drift with no attacker	● Yes	● Yes[5]	?	?
Multi-agent handoffs	● Yes	?	?	?
Runaway cost: loops, excess calls	● Yes	◐ Partial[6]	?	?
Self-hosted or air-gapped option	● Yes	● Yes[7]	?[8]	● Yes[9]

? = **not stated in the vendor's public materials (checked Oct 31, 2026).** [1] Formerly Lakera Red; agent endpoint plus tool tracing. [2] AgentBreaker probe. [3] Excessive-agency and RBAC plugins. [4] "Unauthorized discounts/refunds" objective. [5] Goal-misalignment plugin. [6] Reasoning-DoS plugin (excess computation). [7] Enterprise On-Prem; open-source core. Joining OpenAI (announced Mar 2026). [8] Documented as hosted. [9] Open source, runs locally.

Gadriel: GBA datasheet (Aug 2026). Promptfoo: agents, multi-turn, excessive agency, goal misalignment, reasoning DoS, enterprise; Check Point (Lakera docs): attack coverage, agents, indirect attacks, deployment; garak: AgentBreaker, GitHub. Real case: Moffatt v. Air Canada, 2024 BCCRT 149.

How to get started with GBA

01 **Request** a Proof of Value

02 **Map** it to the eight pillars

03 **Validate** it, before production

04 **Review** findings before go-live

gadriel.ai/behavioral-assurance/proof-of-value

GDA

Gadriel Deployment Assurance

Threat Exposure Validation. Validate where AI runs.

SaaS
Private cloud
Air-gap

Prove what an attacker can actually reach.

You approve a named scope. GDA finds what is exposed, attacks it safely, and reports only findings backed by evidence. Then it retests your fix.

Two natures of attack. One platform.

YIN · DETERMINISTIC

Proof you can replay

Repeatable known-attack logic. Replayable plans, no model required. Air-gap ready.



YANG · PROBABILISTIC

AI that finds new paths

Adaptive exploration with open-weight models, opt-in and self-hosted. It cannot publish a Critical finding without captured output.

Probabilistic exploration. Deterministic proof. You can leave exploration off. You cannot turn the evidence gate off.

Every finding gets an evidence grade

● Exploited
Captured proof. The only grade that can be Critical.

■ Path Identified
Reachable chain proven. Exploit not run.

○ Observed
Condition confirmed. Exploit not proven.

◇ Simulated
Modeled only. Never shown as proof.

One product across the estate. Network to AI to robots.

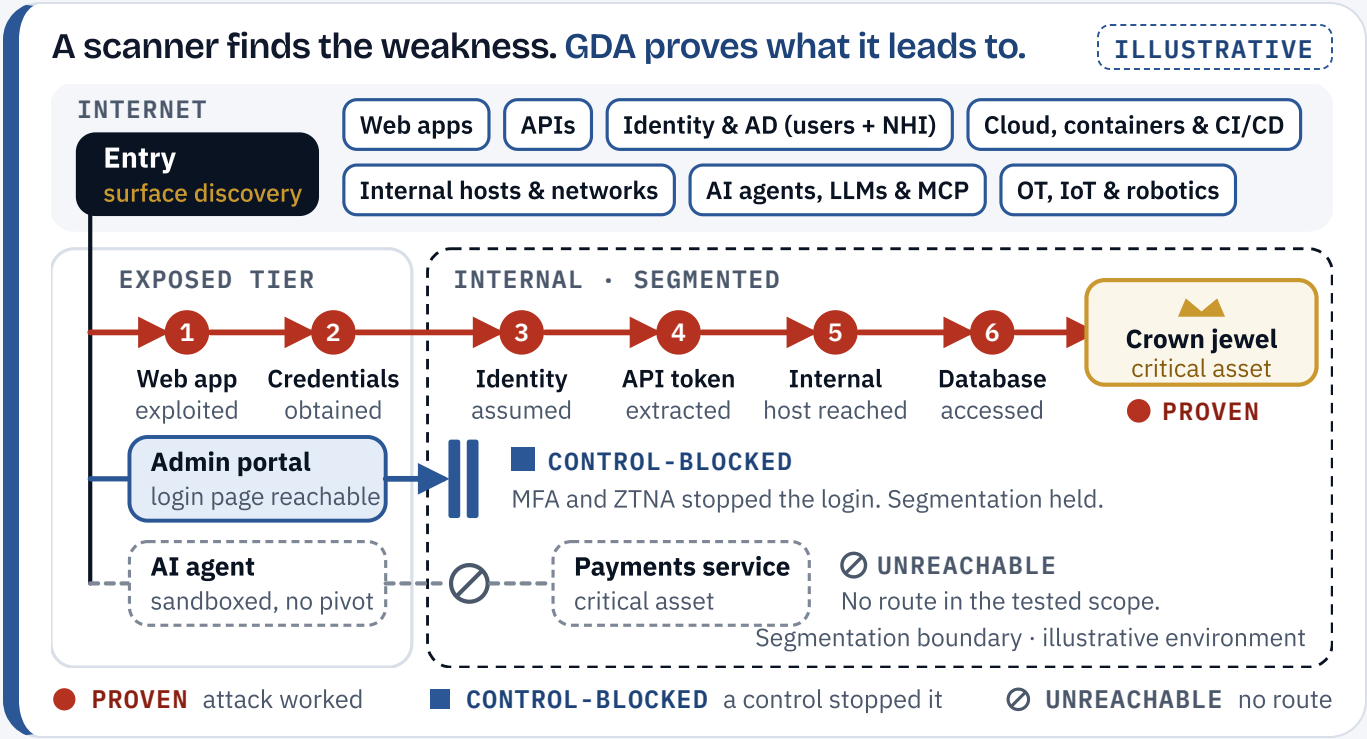
Use case: a customer wants a pentest, and the deal can't wait. A manual test can take weeks to book. GDA tests the targets you own and delivers an audit-ready report. You fix. We retest.

15
web app attack types, injection to request smuggling

15
API attack types, incl. broken access and JWT

9
security controls tested: MFA, DLP, GenAI egress

2
business days from confirmed scope to report



Coverage: GDA datasheet (Aug 2026); robotics and IoT by SOW. Page 4 sources: ACMA; Binance; U.S. DOJ; T-Mobile 8-K.

Black October. One week in 2026.

Oct 21	CISA tells operators to take PHP systems offline until patched. PHP2Shell allows remote code execution with no login. cisa.gov	GCA GDA
Oct 21	Microsoft reports Storm-0940 using ARTEX, an agentic AI pentest framework, against hundreds of organizations. microsoft.com	GDA
Oct 21	One AI-run campaign: about 100 breaches for a few thousand dollars in AI tokens. Fortune	GDA
Oct 22	Morgan Stanley begins blocking third-party AI agents. JPMorgan, Bank of America and Citi restrict them. Axios	GBA
Oct 23	Experts estimate up to 70% of websites were at risk from the PHP bugs. The Guardian	GCA GDA

GDA **PHP2Shell: we test whether it's actually reachable in your environment.**

Three real attack chains from authorized tests

An online marketplace

Observed

1. Public website code
2. Shows how servers decide who you are
3. Hidden request details added
4. **Treated as staff acting for a customer**

A company's AI support assistant

Attack Path Identified

1. The chat window's own code
2. Routes behind the assistant mapped
3. Several work with no login
4. **Bot protection stopped the final step**

A fleet of cleaning robots

Exploited

1. Admin test access from the client
2. One request to a robot
3. Robot's blocked network opened
4. **Self-defense closed it 7 times; test reopened it each time**

An AI-guided pentest engine found, chained and checked each chain. All tests ran under authorized rules: bug bounty programs, client contracts, or owner test labs. Companies stay anonymous; evidence available under NDA.

Proof, not a list of alerts. How GDA compares.

Pentest platforms compared on what GDA tests in deployed apps, APIs and infrastructure.

GDA How it compares	Gadriel GDA	Horizon3.ai NodeZero	Pentera	Invicti
Web applications	● Yes	● Yes	◐ Partial[1]	● Yes
APIs	● Yes	● Yes	◐ Partial[1]	● Yes
Networks and identity	● Yes	● Yes	● Yes	◐ Partial[2]
AI endpoints and LLM apps	● Yes	?	◐ Partial[1]	● Yes
Proof-of-exploit evidence	● Yes	● Yes	● Yes	● Yes
Retest after a fix	● Yes	● Yes	● Yes	● Yes
Audit-ready reports	● Yes	● Yes	● Yes	● Yes
MSP multi-tenant	● Yes[3]	● Yes	◐ Partial[4]	● Yes

? = **not stated in the vendor's public materials (checked Oct 30, 2026).** [1] Pentera AI-native web app and API testing (incl. AI-integrated apps) is in beta, GA planned Q4 2026. [2] Invicti lists external network assets. [3] GDA multi-tenancy for MSPs confirmed by Gadriel, Oct 2026. [4] Pentera has an MSSP partner program; multi-tenancy not stated.

Gadriel: GDA datasheet (Aug 2026) and Gadriel confirmation. Horizon3.ai: NodeZero web app, docs, PCI pentesting; Pentera: web app testing, platform, partner program; Invicti: MSSP, docs, AI-powered AppSec.

MSPs and MSSPs: test every client at once with multi-tenant GDA.

How to get started with GDA

01 **Scope**: pick the targets you own

02 **Order** online, securely

03 **Test**: deterministic + AI

04 **Report** in 2 business days

gadriel.ai/order-assessment

Asset discovery is free. MSPs: enter your referral code on the order form.

Your data. Your environment.

Sovereign AI: run every product on your own terms

All three products run as SaaS, in your private cloud, or air-gapped.

GCA

Local by default. Code, findings and audit evidence stay on your machine. Air-gap on Enterprise.

GBA

Air-gap CLI. Prompts, evidence and the report stay on your network. Nothing leaves it.

GDA

Single binary. Runs fully disconnected. AI exploration uses self-hosted open-weight models, or stays off.

How to buy · current terms at gadriel.ai/pricing

Individual
GCA free for one developer.

Team
GCA with shared policy and dashboards.

Enterprise
SSO, self-hosted, air-gap. GBA and GDA scoped.

Leadership



Farshid Mahdavipour

Co-Founder, CEO & CTO

25+ years in enterprise security, cloud architecture and engineering. Founded an automated pentesting platform and co-invented a granted U.S. patent on pentesting as code.

linkedin.com/in/farshidmahdavipour/



Koroush Saraf

Co-Founder, VP of Products & Marketing

Former VP of Products at Fortinet and Palo Alto Networks; co-founder of ConSentry Networks; M.S. ECE, Stanford & USC.

linkedin.com/in/koroush-saraf/

Trusted by

Customers, listed with their approval.

gruve · cardinal ROBOTICS · GATLIN ROBOTICS · Jobotics AI · Santa Cruz Seaside Company

From the Gadriel blog.

Scan a code to read the post. All posts: gadriel.ai/blog



GCA 2026-08-31

AI Is Writing More Code. Enterprise Development Needs an AI Development and Security Harness.

Company security rules should apply while code is being written.

gadriel.ai/blog/ai-development-and-security-harness.html

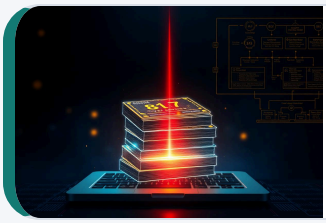


GBA 2026-08-06

Inside a 17-agent clinical AI system — and what pre-production validation found in it

284 tests. Direct attacks failed. A two-turn reframe did not.

gadriel.ai/blog/gba-agentic-healthcare-assessment.html



GCA 2026-06-27

Anthropic Skills Couldn't Find the RCE. Gadriel Did.

817 AI skills missed a remote-code bug. Gadriel found it.

gadriel.ai/blog/817-skills-vs-2026-cve.html



GDA 2026-08-26

Threat Exposure Validation: Attack Your Deployment Before Mythos Does

Secure code isn't a secure deployment. GDA proves what's reachable.

gadriel.ai/blog/threat-exposure-validation-mythos.html



GCA 2026-06-26

Your AI Coding Assistant Cannot See This CVE (and 50,000 more identified vulnerabilities)

AI misses flaws newer than its training. A live feed doesn't.

gadriel.ai/blog/ai-cannot-replace-sca.html



GCA 2026-07-11

Gadriel vs. Snyk: Two Strong Scanners, One Big Structural Difference

Six test apps, five languages. Gadriel runs on your machine.

gadriel.ai/blog/gadriel-vs-snyk.html

Gadriel

Find it. Fix it. Before attackers do.

At Gadriel.ai, we're building security assurance across code, AI agent behavior, and deployed applications to find and fix exploitable vulnerabilities before attackers do.

START WITH ONE TARGET

Order online. Run it yourself.

GDA Assessment: External Attack Surface Penetration Testing

- Web apps, APIs, network nodes and AI agent endpoints
- Free public asset discovery on your domain
- Audit-ready report in 2 business days
- MSPs: enter your own referral code on the form

As described on gadriel.ai/order-assessment. Current terms on the site.



See a published sample report first

gadriel.ai/order-assessment

Web	gadriel.ai
Email	contact@gadriel.ai
LinkedIn	linkedin.com/company/gadriel-ai
HQ	San Francisco, CA
Phone	669-244-2744

Get started
gadriel.ai/get-started

© 2026 Gadriel AI Corp. All rights reserved. Third-party names belong to their owners.

MSP edition · Version 1.1 · Nov 2026