



SMART SQUAD AI IN YOUR AZURE TENANT

Faction Four Systems Inc.

Applies to Smart Squad 2.9.4 and later · Revised October 2026

Connecting Smart Squad's AI features to an Azure OpenAI resource in your agency's own Azure subscription, so report narratives, ticket images and policy documents are processed by a service you own, in the region you choose.

CONTENTS

Overview	2
What this covers	2
What your team sets up in Azure	3
What to provide Faction Four	4
What Faction Four does	5
What stays on Faction Four services	3
Go-live checklist	5
Frequently asked questions	6
Questions	6



OVERVIEW

Smart Squad's AI features — Smart Draft report writing, Smart Review, AI Chat and more — run on Microsoft's Azure OpenAI service. By default that service is hosted and managed by Faction Four.

For agencies that want direct control over where AI processing happens, Smart Squad can connect to an Azure OpenAI resource in your own tenant instead. **There is no software to install and no change for your officers.** It is a server-side configuration change made by Faction Four, and it can be reversed just as easily.

What you gain

- **Data residency you control.** Pick the Azure region — Canada Central, for example — and keep AI processing inside it.
- **Your governance, your logs.** The resource lives under your tenant's security policies, auditing and retention rules.
- **Your billing.** Token usage lands on your Azure bill with full visibility in the Azure portal. Smart Squad continues to track per-call usage on its side as well.
- **Your capacity.** You set the throughput quota, so another organisation's workload never competes with yours.

Microsoft does not use Azure OpenAI customer data to train its models under either arrangement. This option is about control and residency, not a change in how Microsoft handles data.

WHAT THIS COVERS

All of Smart Squad's language-model features route through your resource once connected.

FEATURE	WHAT IT DOES	CAPABILITY REQUIRED
Smart Draft	Generates a draft report narrative	Text
Smart Review	Reviews a report before submission	Text
AI Chat	Q&A over your policies and charge codes	Text + embeddings
Suggest Edit	Inline writing suggestions in notes	Text
Change Summary	Summarises what changed in a report	Text
Media Summary	Summarises transcripts and media	Text
Form Prefill	Fills form fields from a report	Text
Smart Pulse	Wellness signal from report tone	Text
Paper Ticket Scan	Reads handwritten tickets from photos	Vision

A small number of supporting services use different Azure products and are not part of this change — see what stays on Faction Four services.

WHAT YOUR TEAM SETS UP IN AZURE

Four steps in your subscription, for your IT staff or Azure partner. Most agencies finish in under a day.

1 Create an Azure OpenAI resource

In your subscription, in your preferred region. Two things from it go to Faction Four later: the endpoint address (for example `https://youragency.openai.azure.com`) and an API key.

2

Deploy the models

Inside the resource, create model deployments for the features you use:

PURPOSE	SERVES	EXAMPLE MODEL
Chat	Smart Draft, Smart Review, AI Chat and all other text features	GPT-4.1
Vision	Paper Ticket Scan	GPT-4o
Embeddings	AI Chat document search	text-embedding-3-small

One chat deployment can serve every text feature, or you can deploy different models for different features — both are supported.

3

Open the network path

The Smart Squad servers must reach your endpoint over HTTPS. If the resource uses the standard public Azure endpoint, nothing further is needed. If you lock it down to a private endpoint, coordinate routing and DNS with Faction Four before go-live.

4

Set your capacity

Assign enough tokens-per-minute quota to each deployment for your agency's size. Smart Draft and Paper Ticket Scan are the heaviest users. Quota is adjustable at any time in the Azure portal, and Faction Four can recommend starting values based on your report volume.

Name each deployment after the model it runs — a deployment of GPT-4.1 named `gpt-4.1`. Smart Squad reads the deployment name to work out how to talk to the model, so a generic name like `chat` causes request errors for certain model families. Faction Four will confirm your deployment names with you before go-live.

Embeddings. If you use AI Chat, deploying `text-embedding-3-small` — the model Smart Squad uses today — makes the switch seamless. A different embedding model is supported, but requires Faction Four to re-index your policy documents and charge codes as part of the cutover.

WHAT TO PROVIDE FACTION FOUR

When your Azure setup is ready, send your Faction Four contact:

1. **Endpoint address** — the resource URL, for example `https://youragency.openai.azure.com`
2. **API key** — through a secure channel, not plain email. We will agree the method with you.
3. **Deployment names** — for chat, vision and embeddings as applicable
4. **Embedding model choice** — confirm `text-embedding-3-small` or tell us what you deployed instead
5. **Network details** — only if you are using a private endpoint

WHAT FACTION FOUR DOES

- Applies the configuration for your agency on the Smart Squad servers. The change is per-agency and takes effect the same day.
- Schedules a short validation window with your team: we run Smart Draft, AI Chat and Paper Ticket Scan together and confirm in your Azure portal that the calls are landing in your tenant.
- Remains your single point of support for Smart Squad, exactly as today. You operate the Azure resource; we support everything that talks to it.

If anything needs to be unwound, the configuration reverts to Faction Four's hosted service in minutes — no data migration, and no downtime for officers.

WHAT STAYS ON FACTION FOUR SERVICES

Three supporting capabilities use different Azure products — Speech and Language services rather than Azure OpenAI — and continue to run on Faction Four's cloud in a standard deployment:

CAPABILITY	SERVICE USED
Audio transcription	Azure Speech-to-Text
Voice output (Smart Glasses)	Azure Speech Neural Voice
PII redaction pre-processing	Azure Language

If your agency requires these in your tenant as well, that is available on a dedicated Smart Squad deployment — talk to your Faction Four representative. Alternatively, any of them can be disabled for your agency.

Face recognition already runs locally inside Smart Squad and never calls an external AI service.

GO-LIVE CHECKLIST

- ☐ Azure OpenAI resource created in your chosen region
- ☐ Chat, vision and embedding deployments created, each named after its model

- ☐ Quota assigned to each deployment
- ☐ Network path confirmed from the Smart Squad servers
- ☐ Endpoint, key and deployment names delivered securely to Faction Four
- ☐ Joint validation session scheduled
- ☐ Smart Draft, AI Chat and Paper Ticket Scan verified end to end
- ☐ Token usage confirmed in your Azure portal

FREQUENTLY ASKED QUESTIONS

Who pays for the AI usage?

You do, on your Azure bill, at Microsoft's standard Azure OpenAI rates. Faction Four hosting fees for AI no longer apply to the covered features.

Can we change models later?

Yes. Deploy the new model in your resource, tell Faction Four the deployment name, and we update your configuration. Upgrading models is one of the main reasons agencies choose this option.

What happens if we run out of quota?

Requests are briefly throttled by Azure and Smart Squad reports the error to the officer. Raising quota in the Azure portal resolves it immediately — no Faction Four change needed.

Is our data used to train AI models?

No. Microsoft's Azure OpenAI terms exclude customer data from model training, in your tenant or ours.

Does this require a Smart Squad update?

It requires Smart Squad 2.9.4 or later on your environment. Your Faction Four contact will confirm your version as part of scoping.

QUESTIONS

Talk to your Faction Four representative, or raise it through Faction Four support.