Anar Labs

Anar Guard

Your rules, on every AI prompt.

Guard checks text against your policies, masks GCC personal data and writes each decision to an audit trail, before the text reaches a model.

For
CISOs, data protection officers and AI governance leads

Date
November 2026

The problem

Staff paste personal data into prompts every day.

A prompt a staff member writes

Please update the file for Ahmed. Emirates ID 784-1985-1234567-1, mobile +971 50 123 4567, IBAN AE07 0331 2345 6789 0123 456.

→

Without a control point

The identity number, the phone and the bank account go to the model provider. Nobody has a record that it happened.

Policies on paper

Your AI policy is a document. Nothing checks a prompt against it.

Personal data leaves

National IDs, phones and IBANs travel with the question.

No evidence

The auditor asks what was checked. Nobody can show it.

How it works

One call before the model answers.

01

Your application

Sends the prompt or the answer to Guard.

→

02

Anar Guard

Applies your policies, the content filter and the personal-data scan. Writes an audit row.

→

03

The decision

Allow, warn or block, with the masked text and the reason.

```
POST /api/v1/governance/evaluate  →
{ "allowed": false, "action": "block", "violations": [ ... ],
  "pii_detected": [ ... ], "redacted_content": "..." }
```

Field names from the Guard API. Anar supplies no models. Guard checks the text that your application sends to it.

Personal data

Guard finds GCC personal data in Arabic and English, and masks it.

Input

Please update the file for Ahmed. Emirates ID 784-1985-1234567-1, mobile +971 50 123 4567, IBAN AE07 0331 2345 6789 0123 456.

→

Guard output

Please update the file for Ahmed. Emirates ID `[REDACTED_NATIONAL_ID]`, mobile `[REDACTED_PHONE]`, IBAN `[REDACTED_IBAN]`.

رقم الهوية 784198512345671 والهاتف +966 55 123 4567، أرسل التقرير اليوم.

→

رقم الهوية `[REDACTED_NATIONAL_ID]` والهاتف `[REDACTED_PHONE]`، أرسل التقرير اليوم.

Shipped Output of Guard's detector on these two texts, run 18 November 2026. Names are not masked yet.

Policies

Write a rule once. Guard applies it in both languages.

Policy

Classified project names

Type: keyword. Action: block.

Keywords: Project Falcon, مشروع الصقر

Rules are keyword lists or patterns. Each rule logs, warns or blocks. The strictest result wins.

Shipped

Prompt

Summarise the Project Falcon budget for the minister.

```
"action": "block",
"reason": "Matched keywords: Project Falcon"
```

لخص ميزانية مشروع الصقر للوزير.

```
"action": "block",
"reason": "Matched keywords: مشروع الصقر"
```

Output of Guard's policy check on these prompts, run 18 November 2026.

Alerts

Your security team hears about it in their own tools.

01

Alert rule

A metric, a threshold and a time window. For example, more than 20 blocks in 15 minutes.

→

02

Alert

Guard raises it once, then waits for the cooldown. Your team acknowledges it.

→

03

Signed webhook

Your SIEM or chat tool gets the alert, signed with HMAC-SHA256, with retries.

```
POST https://siem.example.gov.ae/hooks/ai
X-Guard-Signature: sha256 HMAC of the body, with your shared secret
```

Shipped

Audit trail

Every decision is on record, ready for your auditor.

Time	Event	Severity	User	Result
09:14:02	policy_violation	critical	finance.analyst	block: Classified project names
09:14:40	pii_detection	warning	hr.officer	3 entities masked
09:15:11	evaluation	info	citizen.portal	allowed

Example rows. The columns are the ones Guard exports: id, timestamp, event type, content snippet, result, user, model and severity.

Search and filter

Filter by event type, severity, user and time.

Shipped

Export

Download the trail as CSV or JSON for your auditor.

Shipped

Tamper evidence

A chained record that shows any later edit.

Roadmap

Framework reports UAE AI Ethics · SDAIA · Qatar AI Charter · Oman · EU AI Act — Partial

The platform

Other Anar products report to Guard.

Anar Chat

An event for each conversation turn.

Anar Voice

An event for each transcription.

Anar Agents

An event for each run.

```
{
  "event_type": "llm_call",
  "product": "chat",
  "model": "your-model",
  "user_id": "hr.officer",
  "latency_ms": 840,
  "total_tokens": 1210,
  "severity": "info"
}
```

The fields each product sends to Guard. The values are an example. Set the Guard address in each product to turn it on.

Status today

What works now, and what comes next.

Capability	Status	Note
Policies, personal-data masking, audit export	Shipped	Keyword and pattern rules. GCC IDs, phones, IBANs, cards, passports.
Alerts and signed webhooks	Shipped	Thresholds, cooldowns and retries.
Safety classifier	Partial	Uses a model you configure. It marks text safe when that model does not answer.
Inline proxy to Anar Gateway	Partial	Built. No Anar product routes through it yet.
Framework reports	Partial	Five frameworks. The score counts activity. It does not test controls.
Names in Arabic text, tenants, tamper evidence	Roadmap	One customer per deployment today.

Roadmap

Next, in the order a security review asks for it.

Now

- Keyword and pattern policies
- GCC personal-data masking
- Audit trail with export
- Alerts and signed webhooks

Next

Roadmap

- Safety classifier inside your boundary
- Arabizi and Gulf dialect detection
- Names and places in Arabic text

Later

Roadmap

- Human review queue
- Policy versions and replay
- Evidence mapped to each control

Put your rules in front of the model.

A one-month pilot: we deploy Guard in your environment and connect it to one team's AI use.

anarlabs.ai/contact