ASK ARTHUR

Security Overview

Platform architecture, controls, and compliance posture

GENERAL USE

April 2026

Version 1.0

Contents

1. Overview

Ask Arthur is an Australian scam detection platform operated by Young Milton Pty Ltd. The service combines threat intellig assisted classification, and a sub-processor stack hosted primarily in Sydney to deliver real-time verdicts on suspicious con web app, browser extension, mobile app, B2B API, and chat-bot integrations.

This document summarises the security, privacy, and operational controls that protect customer data. It is provided for gen diligence and does not constitute a SOC 2 or ISO 27001 attestation. Where formal attestation is required, contact brendan@ compliance roadmap and supplementary documentation under NDA.

2. Platform Architecture

2.1 Deployment Model

Ask Arthur runs as a multi-tenant SaaS application on Vercel, with PostgreSQL hosted on Supabase (AWS Sydney region, a primary data is processed and stored within Australia.

2.2 Infrastructure

Web hosting	Vercel (Sydney edge, syd1)
Database	Supabase PostgreSQL — Sydney (ap-southeast-2)
Object storage	Cloudflare R2 (Oceania)
Rate limiting	Upstash Redis (Singapore)
AI processing	Anthropic Claude API (US — query data only, no storage or training)
Email delivery	Resend (US)
Disaster recovery	Supabase point-in-time recovery, 7-day window

2.3 Network Posture

No databases are exposed to the public internet. All ingress traffic terminates at the Vercel edge, which proxies to Next.js s TLS 1.3. Outbound traffic to sub-processors uses authenticated TLS connections with SDK-issued credentials.

3. Security Controls

3.1 Data Protection

CONTROL	IMPLEMENTATION
Encryption at rest	AES-256 via Supabase-managed KMS
Encryption in transit	TLS 1.3 enforced; HSTS preload (max-age 2 years)
PII scrubbing	12-pattern pipeline strips emails, cards, Medicare, TFN, SSN, AU/intl phones, IPs, BSBs, addresses, and name
Data retention	SAFE/SUSPICIOUS reports archived after 90 days; HIGH_RISK after 180 days; bot queue cleared at 24h
Backups	Supabase daily encrypted backups; point-in-time recovery to any second within 7 days
Secret hygiene	API keys SHA-256 hashed; plaintext never stored; timing-safe comparisons for all secret checks

3.2 Identity & Access Management

SURFACE	METHOD
End-user authentication	Supabase Auth (PKCE) — email/password, magic link; MFA available
Session management	HttpOnly, Secure, SameSite=Strict cookies; server-side JWT validation via `getUser()`
Authorisation	PostgreSQL Row Level Security on every multi-tenant table
B2B API	Bearer tokens; SHA-256 hashed; per-key rate limits; max 5 active keys per user
Browser extension	Per-install ECDSA P-256 signature; non-extractable private key in IndexedDB; nonce replay protection
Bot webhooks	Platform HMAC verification (Telegram secret token, WhatsApp SHA-256, Slack v0 with 5-min replay w
Privileged access	Admin panel uses dual-mode auth (Supabase admin role with HMAC cookie fallback); 24h session exp

3.3 Application Security

CONTROL	IMPLEMENTATION
Input validation	All external input validated with Zod schemas; 10MB payload cap
Prompt-injection defence	Unicode sanitization (11 invisible-char classes), nonce-based XML delimiters, sandwich defence, 14 regex
HTML sanitization	Email scans strip comments, style/script blocks, hidden elements, and data attributes server-side before
Content Security Policy	No `unsafe-eval` ; `frame-ancestors 'none'` ; explicit allowlist for connect/script/style
Rate limiting	Two layers: edge middleware (60/min per IP on API routes) + per-route burst+daily limits keyed to install/I production.
Code review	All changes via GitHub PR; required Vercel preview deploy gating; squash-merge to main
Dependency scanning	`pnpm audit` for Node; `pip audit` for Python pipeline; lockfiles committed
SSRF protection	Image proxy uses domain allowlist, manual redirect handling, content-type validation, 5MB / 10s caps

3.4 Infrastructure Security

CONTROL	IMPLEMENTATION
DDoS protection	Vercel + Cloudflare edge (sub-processors)
Patch management	Dependencies patched within 7 days for high severity; managed runtime auto-patched (Vercel, Supabase)
Database hardening	RLS-enforced; `SECURITY DEFINER` RPCs scoped with `SET search_path = public` ; admin-only RPCs revoke `public/anon/authenticated`
Webhook idempotency	Stripe events deduplicated via `stripe_event_log` insert-with-conflict gate
Secret management	Environment variables via Vercel; never committed; quarterly rotation policy for high-sensitivity keys

4. Operational Security

4.1 Logging & Monitoring

- **Application logs** — structured JSON via `@askarthur/utils/logger`, retained in Vercel Observability for 30 days
- **Database audit** — Supabase log explorer; advisor scans run weekly via the MCP tooling
- **Cost telemetry** — `cost_telemetry` table records every paid-API call with feature/provider tagging; `/admin/costs` dash anomalies
- **Health monitoring** — `/admin/health` shows queue depth, archive counts, Stripe idempotency log, and feed staleness
- **Alerting** — daily Telegram threshold alerts (default $2 USD/day) and weekly week-over-week digest

4.2 Change Management

- Every change shipped via a GitHub pull request from a feature branch off `main`
- Vercel preview deployment must pass before merge; squash-merge preserves a linear file tree for cache hits
- Migrations are idempotent (`CREATE TABLE IF NOT EXISTS`, `DROP POLICY IF EXISTS`) and applied via the Supabase MCP verification
- Destructive or long-running migrations require a runbook and maintenance window
- Detailed commit messages document the technical hypothesis, approach, and outcome — preserved as contemporaneou

4.3 Backup & Recovery

- **Recovery Point Objective (RPO):** < 1 minute (Supabase point-in-time recovery)
- **Recovery Time Objective (RTO):** < 4 hours for primary database restore
- **Application:** stateless; Vercel rolls back to a prior deployment in < 60 seconds
- **Object storage:** Cloudflare R2 with versioning enabled on evidence buckets

5. Privacy & Data Handling

5.1 Privacy Principles

- **Data minimisation** — submitted text and images are processed for analysis and immediately discarded; only PII-scrubbe retained for HIGH_RISK verdicts
- **Purpose limitation** — customer data is used solely for delivering the contracted service; no secondary marketing use
- **Customer ownership** — all customer-linked data remains the customer's property and is portable on request
- **AI training** — Anthropic's commercial terms exclude API customer data from model training; Ask Arthur does not operate pipeline

5.2 Australian Privacy Act 1988 (Cth)

Ask Arthur operates under the 13 Australian Privacy Principles and provides the following self-service rights to authenticat

RIGHT	ENDPOINT
Access (APP 12)	`GET /api/user/export-data` — JSON bundle of all data linked to the caller
Erasure (APP 11/13)	`POST /api/user/delete-account` — cascades to `user_profiles` and owned `family_grou`
Rectification	Account settings UI; bulk corrections via brendan@askarthur.au
Data breach notification (Pt IIIC)	Eligible breaches notified to the OAIC and affected individuals as soon as practicable

5.3 Cross-Border Data Transfers

Where overseas processing is necessary (AI analysis via Anthropic US, email delivery via Resend US, rate limiting via Upsta Arthur ensures equivalent protection through contractual safeguards and discloses each transfer in the public sub-process

6. Compliance Roadmap

FRAMEWORK	STATUS	DETAIL
Australian Privacy Act 1988	Compliant	13 APPs covered; APP 12 + 13 endpoints live
ASD Essential Eight (ML1)	Self-assessed	Application control, patching, MFA, daily backups in place
SOC 2 Type I	In progress	Target Q3 2026
SOC 2 Type II	Planned	Target H1 2027 following Type I attestation
ISO 27001	Planned	Target 2027
IRAP assessment	On request	Available for Australian Government engagements

7. Sub-Processors

The following third parties process customer data on our behalf. Each holds independent security certifications. The curren
at askarthur.au/trust.

PROVIDER	REGION	PURPOSE	CERTIFICATIONS
Supabase, Inc.	USA / Sydney	Database hosting	SOC 2 Type II
Vercel, Inc.	USA / Sydney	Application hosting	SOC 2 + ISO 27001
Cloudflare, Inc.	USA / Oceania	CDN + object storage	SOC 2 + ISO 27001
Anthropic, PBC	USA	AI analysis	Enterprise DPA; SOC 2
Twilio, Inc.	USA	Phone intelligence	SOC 2 Type II
Upstash, Inc.	Singapore	Rate limiting	SOC 2
Resend, Inc.	USA	Email delivery	SOC 2
Stripe, Inc.	USA	Billing	PCI DSS Level 1; SOC

8. Incident Response & Contact

8.1 Incident Response

- **Triage** — within 24 hours of confirmed report
- **Customer notification** — affected customers notified within 72 hours of confirming a breach
- **Regulator notification** — eligible breaches reported to the OAIC under the Notifiable Data Breaches scheme as soon as p
- **Post-incident review** — written within 14 days of remediation; available to enterprise customers under NDA

8.2 Vulnerability Disclosure

Researchers and customers can report suspected vulnerabilities to brendan@askarthur.au. We do not currently operate a pa programme but will acknowledge responsible disclosures publicly with the reporter's consent.

8.3 Contact

Security & incidents	brendan@askarthur.au
Privacy & data requests	brendan@askarthur.au
Vendor due diligence / DPA	brendan@askarthur.au
Legal entity	Young Milton Pty Ltd · ABN 72 695 772 313
Web	askarthur.au

Further detail on our security and compliance programme is available under NDA. This document is reviewed at minimum annuall whenever a material control changes.