TOTALLY PRIVATE AI

Evidence Pack

Your Data. Your AI. Totally Private.

Version	V3.0 EU | For security & compliance review
Company	Totally Private AI (TPAI)
Use Cases	Legal, Healthcare, Financial
Contact	rcastillo@totallyprivate.ai

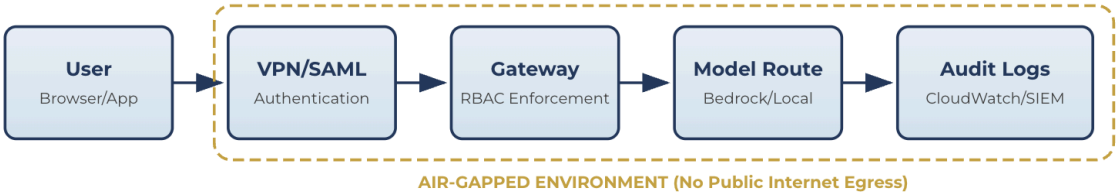
Key Definitions

Air-gapped	No public internet egress. Application plane has no direct internet egress (no IGW/NAT for application subnets) and relies on private connectivity (VPC endpoints) for required cloud services. Air-gapped deployments apply to POC and Enterprise tiers. Pro and Teams are secure multi-tenant offerings with controlled access.
Customer Data	Prompts, file uploads, chat outputs, and any content provided by users. Stored encrypted with envelope encryption (DEKs protected by KMS).
Retention	Customer data: configurable (30-day minimum for backup/operational requirements). Audit logs: metadata only, configurable retention.
Pseudonymous IDs	User identification via HMAC-SHA256 of email address with secret key. Not reversible without key access and privileged system abuse. Re-identification requires separate access to billing systems and is controlled via role-based access.

Security Posture at a Glance

- **Air-gapped:** No public internet egress; entire environment isolated
- **Encrypted:** Envelope encryption with KMS protection
- **Auditable:** Per-request logging with SIEM export
- **Framework aligned:** Mapped to ISO 27001 controls, GDPR Article 32, DORA requirements (not certified yet; certification roadmap available)

Data Flow Overview



1. **Authentication:** Client certificates (Pro) or SAML integration (Enterprise)
2. **Authorization Gateway:** RBAC enforcement before any LLM call
3. **Model Routing:** Bedrock via VPC endpoints
4. **Audit Trail:** Metadata logged to CloudWatch and S3 (no content)
5. **Data Storage:** Customer data encrypted with envelope encryption

Deployment Models & Shared Responsibility

Responsibility	TPAI Pro/POC	TPAI Enterprise
Infrastructure	TPAI AWS account	Customer AWS account
Key Management	TPAI-managed CMKs with strict tenant isolation	Customer-managed CMKs in customer account
GDPR Role	TPAI is Data Processor	With Maintenance: Processor. Without Maintenance: No ongoing processor relationship
Production Access	TPAI (for operations)	Customer only (TPAI requires permission)

What We Guarantee Today (Beta)

- **Air-gapped application plane.** No 0.0.0.0/0 egress from app subnets; Security Groups default-deny; service access via VPC Endpoints (Interface for AWS services such as Bedrock/KMS/Logs) and AWS Gateway (for S3/DynamoDB).
- **Policy-first authorization (RBAC).** The LLM never grants privileges. The platform enforces role-based access at the gateway before any tool/LLM call. (Roadmap: optional policy-as-code engine for attribute/context rules.)
- **Least privilege, explicit tool/API allow-lists.** Each role/tenant has a defined set of callable tools/APIs; platform issues scoped, short-lived STS credentials and blocks undeclared endpoints. No long-lived model tokens on clients.
- **Data governance by default.** Per-user envelope encryption (DEKs protected by KMS CMKs) with kms:EncryptionContext binding; configurable retention (30-day minimum); optional Object Lock for legal hold.
- **Audit & export (beta).** The orchestrator emits a per-request audit event (timestamp, tenant/principal/role, model ID, route ID, token usage where available, policy decision flags, input hash). No direct identifiers in application logs. Events stream to CloudWatch (EMF) and S3 for SIEM.
- **Provider continuity.** Bedrock via Private VPC endpoints; open-weights in VPC available as controlled fallback; every inference logs vendor/model "route" for traceability.
- **No training on customer data.** Explicit no-training guarantee. AWS Bedrock models do not train on customer inputs per AWS service terms.

AWS Bedrock Implementation Details

What Bedrock receives: Inference requests containing prompts and context via VPC private endpoints. May include pseudonymous user IDs for request routing. No customer metadata beyond the inference request itself.

Retention policy: AWS Bedrock does not store customer prompts or outputs per AWS service terms.

Traffic routing: All Bedrock communication via VPC Interface Endpoints; no internet transit.

GDPR & EU Regulatory Alignment

- **Data Minimization (Article 5.1.c):** Only necessary data processed; configurable retention periods; automatic deletion
- **Security of Processing (Article 32):** Encryption in transit and at rest; access controls; regular security assessments
- **Data Subject Rights (Articles 15-22):** Customer controls all personal data. For Pro/POC: TPAI provides data export tools and deletion capabilities. For

Enterprise: Customer has direct database access. Data portability via standard formats (JSON/CSV). Deletion propagates to backups within retention period.

- **DORA ICT Risk Management:** Comprehensive logging, incident response capabilities, operational resilience controls

EU Regulatory Context

TPAI provides ICT software/services that may be used by regulated entities to support internal knowledge work while applying strict security controls (access control, encryption, audit logging, and egress restriction). Under DORA, customers assess whether TPAI is an ICT third-party service provider for their outsourcing/third-party risk processes; TPAI supports these assessments with contractual and technical documentation.

Under the EU AI Act, risk classification depends primarily on the customer's intended use. TPAI does not develop or place a general-purpose AI model on the market; it integrates and controls access to third-party models. Customers remain responsible for use-case classification and any "high-risk" obligations tied to their deployment context.

Data Localization

- EU customer data processed and stored within EEA regions only
- Personnel access restricted to EEA-based support staff for EU customers
- No cross-border transfers for processing
- Customer controls region selection and can enforce via AWS Service Control Policies

EU Contracting

We understand EU clients may have concerns about US jurisdiction. TPAI is establishing an EU entity so European contracts will be governed by EU law, with local data processing agreements and support.

TPAI vs. Alternatives

Alternative	Characteristics	Timeline/Cost
Enterprise AI SaaS	Data processed on vendor infrastructure, subject to their policies	Immediate
Cloud AI Private Endpoints	Network isolation but processing still occurs on cloud provider's managed service	Days to weeks
Build Your Own	Full control but requires ML ops expertise, infrastructure management, ongoing maintenance burden	6-12 months, $500K+
TPAI	**Complete deployment within customer's AWS account, customer-controlled encryption, zero external dependencies for AI processing**	**Weeks not months**

Data Portability & Exit

- Customer owns all data and encryption keys
- Standard export formats (JSON, CSV, markdown)
- No proprietary data formats
- For Enterprise deployments, customer retains full environment after contract termination

Evidence Index

Artifacts available on request:

Claim	How We Prove It	Artifact ID
No public egress from environment	describe-route-tables export showing no 0.0.0.0/0; NAT gateways = 0	EP-01
VPC Endpoint scoping	describe-vpc-endpoints (Bedrock/STS/Secrets/Logs), S3/DDB Gateway endpoints	EP-01
RBAC at gateway (LLM never authorizes)	Role→permission map; sample policy (OPA/Cedar roadmap noted)	EP-07

Envelope encryption	CMK policy with kms:EncryptionContext condition; rotation status	EP-02
Per-request audit event	JSON/CSV sample (principal, model/route IDs, tokens, decisions); SIEM export	EP-03

Legal & Privacy Summary

Data Processing Role

TPAI acts as processor when processing personal data on customer's behalf (Pro/POC deployments; Enterprise with maintenance access). For fully customer-controlled Enterprise deployments without TPAI access, no ongoing processor relationship exists.

Subprocessor Management

- **AWS (Infrastructure):** Compute, storage, networking services
- **Anthropic (via Bedrock):** Claude model inference only
- **OpenAI (via Bedrock):** GPT model inference only
- **Meta (via Bedrock):** Llama model inference only
- **Mistral AI (via Bedrock):** Mistral model inference only

No direct data sharing: Model providers receive inference requests via AWS Bedrock private endpoints. No customer data flows directly to model provider infrastructure.

Data Processing Agreements

DPA templates available on request. Standard GDPR processor terms including data minimization, purpose limitation, security requirements, and data subject rights support.

Breach Notification

TPAI notifies customers without undue delay (target 24 hours). Customers (as controllers) manage supervisory authority notifications; TPAI provides incident details and evidence.

Privileged Access & Break-Glass

TPAI production access is disabled by default and enabled only through an explicit customer-approved workflow (ticket + approval) or break-glass. Break-glass is restricted to a small on-call group, requires MFA, is time-bounded, and generates immutable audit logs (who/what/when). Post-incident review is performed and shared with the customer on request.

Update Delivery

Production updates delivered via private connectivity channel. For customer-hosted Enterprise deployments, updates applied under customer change control.

- **Supply chain transparency:** SBOM, vulnerability scan results, and provenance attestations provided with each release.

- **Offline option:** For strict air-gap requirements, offline update packages available as commercial support option.

Frequently Asked Questions

Q: Exactly what data is stored where?

A: Prompts, outputs, files, and chat history stored in OpenWebUI database (encrypted per-user) or customer S3. Audit logs in CloudWatch/S3 (metadata only, no content).

Q: Can retention be set to zero?

A: Customer data retention configurable with 30-day minimum for backup/operational requirements.

Q: Which AI models are available and how are they governed?

A: Claude Opus/Sonnet, Llama 3.1/3.2, Mistral 7B/8x7B via AWS Bedrock. Customer can restrict available models per user/role. Content filtering via AWS Bedrock Guardrails available. Model routing and decision logging for auditability.

Q: How does deletion work with backups and legal holds?

A: Data ages out based on retention period. Legal holds supported with AWS S3 Object Lock and treated separately from standard retention.

Q: Who can access production and how is it approved/logged?

A: Customer Admin has full access. TPAI accesses only with customer permission or break-glass emergencies (logged and time-bounded).

Q: What does Bedrock receive?

A: Inference requests containing prompts and context. May include pseudonymous user IDs for request routing. No customer metadata beyond the inference request itself.

Q: What's your uptime target?

A: 99.9% for standard deployments, 99.99% available with Enterprise SLA. Incident response included in maintenance plan.

Q: Is penetration testing available?

A: Air-gapped architecture significantly limits attack surface. External penetration testing scheduled for Q2 2026.

Business Continuity & Incident Response

Incident Response

TPAI maintains incident detection, containment, and customer notification processes. For security incidents affecting customer data, we notify customers without undue delay (target: 24 hours) and provide periodic updates. Customers handle regulatory notifications; TPAI provides technical details and evidence.

Business Continuity

Enterprise deployments support multiple resilience patterns based on customer requirements (single-region with backups, pilot light, warm standby, active-active). Target RTO/RPO configured per customer deployment.

Pre-Certification Assurance

While formal certifications are in progress, TPAI implements compensating controls: self-assessment against ISO 27001 Annex A, quarterly security reviews, vulnerability scanning in CI/CD pipeline, and immutable audit logging. Gap analysis and remediation plans available on request.

Certification Status & Roadmap

Current Status	Self-attestation against ISO 27001 controls, GDPR requirements
Q2 2026	SOC 2 Type I assessment initiated
Q3 2026	ISO 27001 certification process (target)
Q4 2026	SOC 2 Type II (target)

Control alignment summaries and gap analysis available on request.

Your Data. Your AI. Totally Private.