Votal Shield: DESC AI Security Policy & ISR Control Mapping

First draft for validation. Control IDs and clause numbers must be reconciled to the customer's official DESC / ISR control catalog.

Document status	DRAFT v0.1, to be validated with the customer's compliance / auditor team.
Date	17 June 2026
Classification	Confidential
Purpose	Clause-level mapping of Shield controls to DESC AI Security Policy and ISR, with responsibility split, evidence, and status.

Status legend

Met (Shield-enforced)	Shield enforces this control directly at runtime.
Shared (Shield + customer config)	Shield provides the control; the customer configures / operates it.
Partial	Partially addressed by Shield; complete with additional controls.
Customer responsibility	Owned by the customer's environment / process, not Shield.
Out of scope	Not an AI-gateway control (e.g. physical security, model training).

DESC AI Security Policy

Control ID	Domain	Control / Requirement	How Shield Addresses It	Responsibility	Evidence Artifact	Status	Notes
DESC-AI-01	Governance & accountability	AI/agent use is governed by documented, enforceable policy with clear ownership.	Per-tenant policy registry defines agents, role-to-tool permissions, guardrails, and data policies; enforced at runtime.	Shared	Policy registry export; tenant policy version	Shared (Shield + customer config)	Customer owns policy content; Shield enforces it.
DESC-AI-02	Agent identity	Every AI agent has a verifiable, unique identity bound to its context.	Cryptographically signed agent identity tokens bound to user, agent, and build; OIDC/Keycloak integration.	Shield	Token issuance events; identity claims	Met (Shield-enforced)	Issuance gated to registered, active agents.
DESC-AI-03	Least-privilege authorization	Agents may only invoke tools/actions explicitly permitted to their role.	RBAC role-to-tool matrix; per-invocation authorization; tool ownership prevents cross-agent use; deny by default.	Shared	Authorize decision logs; registry export	Met (Shield-enforced)	Customer defines role-to-tool matrix.
DESC-AI-04	Action authorization (capabilities)	Privileged actions require a scoped, time-bound, single-use grant.	Signed single-use capability tokens scoped to one tool, verified at the tool boundary; one-time nonce prevents replay.	Shield	Capability mint/verify audit; nonce burn	Met (Shield-enforced)	High-assurance flows.
DESC-AI-05	Prompt-injection defense	AI inputs are screened for injection/jailbreak attempts.	Adversarial/prompt-injection guardrail runs inline on input; blocks or flags per policy.	Shield	Guardrail decision logs	Met (Shield-enforced)	Tunable thresholds per tenant.
DESC-AI-06	Sensitive data protection (PII)	Personal/sensitive data is detected and protected in inputs and outputs.	PII detection with detect/mask/redact/block actions; output sanitization before egress.	Shared	Sanitization log entries	Met (Shield-enforced)	Customer defines data rules.
DESC-AI-07	Data/prompt leakage prevention	System prompts and sensitive context are protected from exfiltration.	System-prompt-leak guardrail; output sanitization; data-access scopes per role.	Shield	Guardrail decision logs	Met (Shield-enforced)	
DESC-AI-08	Content safety	Harmful, toxic, or off-topic content is controlled.	Toxicity, bias, topic-restriction, keyword/regex guardrails on input and output.	Shared	Guardrail metrics; decision logs	Met (Shield-enforced)	Customer sets allowed topics.
DESC-AI-09	Human oversight	Sensitive actions allow human confirmation; policies can run in observe mode.	Monitor (dry-run) vs enforce modes; sensitive-action confirmation; kill switch for operators.	Shared	Policy mode; confirmation events	Met (Shield-enforced)	
DESC-AI-10	Transparency & explainability	AI decisions are traceable and explainable.	Every decision recorded with guardrails triggered and reason; reconstructable action lineage.	Shield	Immutable audit log; board report	Met (Shield-enforced)	
DESC-AI-11	Model governance / lifecycle	Models in use are inventoried and governed.	Per-model and per-tenant policy; model-agnostic gateway covers in-house and external SaaS models.	Shared	Model/agent registry	Partial	Model build/training governed by customer MLOps.
DESC-AI-12	Third-party / SaaS AI governance	External AI services are governed and bounded.	Shield governs the request/response and tool boundary for external models (e.g. Claude); provider internals governed by contract.	Shared	Gateway config; audit log	Met (Shield-enforced)	Provider training/retention is contractual.
DESC-AI-13	Abuse / rate control	AI usage is bounded to prevent abuse and runaway cost.	Rate limits on token issuance and capability mint; input length limits; tool-call rate limiting.	Shield	Rate-limit events	Met (Shield-enforced)	
DESC-AI-14	Logging & monitoring	All AI decisions are logged and monitored.	Immutable audit log; guardrail metrics; board/compliance report; SIEM export.	Shared	Audit log; SIEM events	Met (Shield-enforced)	Customer connects SIEM.
DESC-AI-15	Incident detection & response	AI incidents are detected and can be contained quickly.	Real-time alerts on blocks/violations; webhook/SOAR triggers; kill switch for instant containment.	Shared	Alerts; containment events	Met (Shield-enforced)	Customer owns SOAR playbooks.
DESC-AI-16	Identity revocation	Compromised agents/sessions can be revoked.	Agent/instance revocation; disabled agents blocked at runtime; shadow-agent detection.	Shield	Revocation events; block logs	Met (Shield-enforced)	

Control ID	Domain	Control / Requirement	How Shield Addresses It	Responsibility	Evidence Artifact	Status	Notes
DESC-AI-17	Secure deployment & residency	AI controls deploy securely and keep data in-region.	Self-hosted / on-premises / air-gapped deployment; multi-tenant isolation; in-region processing & storage.	Shared	Deployment architecture	Met (Shield-enforced)	Customer chooses deployment model.
DESC-AI-18	Testing & assurance	Agent controls are tested, including adversarial cases.	Automated regression tests plus a live verifier asserting allow/deny outcomes against a deployment.	Shared	Test results; verifier output	Met (Shield-enforced)	Demonstrable in PoC.
DESC-AI-19	Data retention & deletion	AI logs/data follow retention and deletion rules.	Configurable TTL on audit/metrics; customer-owned datastore; no training on customer data.	Shared	Retention config	Shared (Shield + customer config)	
DESC-AI-20	No training on customer data	Customer prompts/data are not used to train models.	Inference-time control plane; traffic processed transiently for a decision; not used for training.	Shield	Architecture statement	Met (Shield-enforced)	

ISR Controls

Control ID	Domain	Control / Requirement	How Shield Addresses It	Responsibility	Evidence Artifact	Status	Notes
ISR-GOV-01	Governance	Security governance and documented policies.	Per-tenant policy registry, versioned policies, deny-by-default posture.	Shared	Policy export	Shared (Shield + customer config)	
ISR-RSK-01	Risk management	Risk-based controls applied to systems.	Risk-tiered tool classification; sensitive-action confirmation; monitor-to-enforce rollout.	Shared	Policy mode; risk tags	Partial	Enterprise risk process owned by customer.
ISR-AM-01	Asset management	Inventory of assets/agents in scope.	Agent registry and tool registry; shadow-agent detection for unregistered agents.	Shared	Registry export	Met (Shield-enforced)	
ISR-AC-01	Access control: identity	Unique identity for users/agents.	Signed agent identity tokens; OIDC/Keycloak; per-process identity.	Shield	Token issuance events	Met (Shield-enforced)	
ISR-AC-02	Access control: least privilege	Access limited to what each role needs.	RBAC role-to-tool matrix; tool ownership; capability tokens; deny by default.	Shared	Authorize logs; registry	Met (Shield-enforced)	Customer defines roles.
ISR-AC-03	Access control: revocation	Timely removal of access.	Agent/instance revocation; kill switch; disabled-agent enforcement.	Shield	Revocation/block logs	Met (Shield-enforced)	
ISR-CR-01	Cryptography	Approved cryptography for data and tokens.	Signed tokens/capabilities; TLS in transit; data at rest protected by customer datastore/keys.	Shared	Token signing; TLS config	Shared (Shield + customer config)	Key ownership with customer.
ISR-OPS-01	Operations: logging	Security events are logged and retained.	Immutable audit log; guardrail metrics; configurable retention.	Shared	Audit log	Met (Shield-enforced)	
ISR-OPS-02	Operations: monitoring	Continuous monitoring of events.	Real-time alerts; board/compliance report; SIEM export.	Shared	Alerts; SIEM events	Met (Shield-enforced)	Customer connects SIEM.
ISR-OPS-03	Operations: malware/abuse	Protection against malicious use.	Prompt-injection, content, and rate-limit guardrails; tool-call validation.	Shield	Guardrail logs	Met (Shield-enforced)	
ISR-COM-01	Communications security	Secure data in transit; segmentation.	TLS; per-tenant isolation; air-gapped option with no outbound egress.	Shared	Network/deploy config	Shared (Shield + customer config)	
ISR-DEV-01	Secure development	Security in the SDLC of in-scope systems.	Governed MCP/codegen with enforcement built in; tested controls; regression suite.	Shared	Test results	Partial	Customer SDLC owned by customer.
ISR-DP-01	Data protection / privacy	Protect personal data; classification.	PII detection, mask/redact/block; data-access scopes; no training on customer data.	Shared	Sanitization logs	Met (Shield-enforced)	Customer classifies data.
ISR-IR-01	Incident management	Detect, respond, and contain incidents.	Alerts; webhook/SOAR automation; kill-switch containment.	Shared	Alert/containment events	Met (Shield-enforced)	Customer owns IR process.
ISR-BC-01	Business continuity	Availability and resilience.	Stateless data plane; horizontal scaling; fail-closed authorization.	Shared	Deploy architecture	Partial	DR/BCP owned by customer.
ISR-TP-01	Third-party security	Govern third-party/SaaS services.	Gateway governs external SaaS models; boundary documented; per-model policy.	Shared	Gateway config	Met (Shield-enforced)	Provider contracts owned by customer.
ISR-CP-01	Compliance & audit	Demonstrable, auditable controls.	Replay-proof, tamper-evident lineage; exportable evidence; board/compliance report.	Shield	Audit lineage; reports	Met (Shield-enforced)	
ISR-HR-01	Human resources security	Personnel screening and awareness.	Not an AI-gateway control.	Customer	HR records	Customer responsibility	Owned by customer.
ISR-PE-01	Physical & environmental	Physical protection of facilities.	Not an AI-gateway control; on-prem deploy runs in customer-secured facilities.	Customer	Facility controls	Out of scope	Owned by customer.
ISR-RET-01	Retention & disposal	Retention and secure disposal of data.	Configurable TTL; customer-owned datastore controls deletion.	Shared	Retention config	Shared (Shield + customer config)	

OWASP LLM Top 10

Control ID	Domain	Control / Requirement	How Shield Addresses It	Responsibility	Evidence Artifact	Status	Notes
LLM01	Prompt injection	Manipulating the model via crafted input.	Adversarial/prompt-injection guardrail on input.	Shield	Guardrail logs	Met (Shield-enforced)	
LLM02	Sensitive information disclosure	Leaking sensitive data via outputs.	PII detection, output sanitization, system-prompt-leak guardrail.	Shield	Sanitization logs	Met (Shield-enforced)	
LLM05	Improper output handling	Unsafe handling of model output downstream.	Output guardrails + data-policy sanitization before egress.	Shield	Guardrail logs	Met (Shield-enforced)	
LLM06	Excessive agency	Agent takes actions beyond intent.	RBAC, tool ownership, capability tokens, kill switch, confirmation.	Shared	Authorize/cap logs	Met (Shield-enforced)	Customer sets roles.
LLM07	System prompt leakage	Disclosure of system instructions.	Dedicated system-prompt-leak guardrail.	Shield	Guardrail logs	Met (Shield-enforced)	
LLM09	Misinformation	Unreliable or biased output.	Bias/toxicity guardrails; human confirmation for sensitive actions.	Shared	Guardrail logs	Partial	Human oversight recommended.
LLM10	Unbounded consumption	Resource exhaustion / runaway cost.	Rate limits, length limits, token/cost controls.	Shield	Rate-limit events	Met (Shield-enforced)	
LLM03	Supply chain	Compromised models/components.	Model build and supply chain governed by customer MLOps, not Shield.	Customer	MLOps records	Out of scope	
LLM04	Data & model poisoning	Tampered training data/models.	Shield does not train models; not applicable to the gateway.	Customer	MLOps records	Out of scope	
LLM08	Vector / embedding weaknesses	RAG/embedding attacks.	Input guardrails apply to retrieved content; embedding store governed by customer.	Shared	Guardrail logs	Partial	

OWASP Agentic Threats

Control ID	Domain	Control / Requirement	How Shield Addresses It	Responsibility	Evidence Artifact	Status	Notes
T1	Memory poisoning	Tampering with agent memory/context.	Input guardrails screen retrieved/context content; agent memory store governed by customer.	Shared	Guardrail logs	Partial	
T2	Tool misuse	Agent uses tools improperly or beyond intent.	RBAC, tool allowlist, tool ownership, tool-call validation, capability tokens, kill switch.	Shared	Authorize/cap logs	Met (Shield-enforced)	
T3	Privilege compromise	Privilege escalation by an agent.	Least-privilege RBAC, capability scoping, deny by default, no cross-agent tool use.	Shared	Authorize logs	Met (Shield-enforced)	
T4	Resource overload	Exhaustion of resources or runaway cost.	Rate limits, input length limits, token/cost controls.	Shield	Rate-limit events	Met (Shield-enforced)	
T5	Cascading hallucination	Errors propagate across agents/steps.	Output guardrails, bias/toxicity checks, human confirmation; factuality not verified.	Shared	Guardrail logs	Partial	
T6	Intent breaking & goal manipulation	Hijacking the agent's goal via input.	Prompt-injection/adversarial guardrail, topic restriction, monitor/enforce.	Shield	Guardrail logs	Met (Shield-enforced)	
T7	Misaligned & deceptive behavior	Agent acts against intended policy.	Guardrails, audit lineage, human oversight; alignment is shared.	Shared	Audit log	Partial	
T8	Repudiation & untraceability	Actions cannot be attributed.	Immutable, tamper-evident audit lineage; SIEM export.	Shield	Audit lineage	Met (Shield-enforced)	
T9	Identity spoofing & impersonation	Agent or user identity is forged.	Signed, build-bound agent identity; capability verify; rogue-agent denial.	Shield	Token/verify logs	Met (Shield-enforced)	
T10	Overwhelming human-in-the-loop	Flooding approvers to bypass review.	Sensitive-action confirmation, rate limits, monitor mode.	Shared	Confirmation events	Partial	
T11	Unexpected code execution / RCE	Agent triggers code/command execution.	Tool-call validation, allowlist, input guardrails; tool runtime owned by customer.	Shared	Guardrail logs	Partial	
T12	Agent communication poisoning	Malicious inter-agent messages.	Guardrails and sanitization on tool/inter-agent messages via the MCP proxy.	Shared	Guardrail logs	Partial	
T13	Rogue agents in multi-agent systems	Unregistered/malicious agents act.	Agent registry, shadow-agent detection, rogue and cross-agent denial, kill switch.	Shield	Registry/block logs	Met (Shield-enforced)	
T14	Human attacks on multi-agent systems	Insiders/attackers abuse the system.	RBAC, authentication, audit; some vectors are organizational.	Shared	Audit log	Partial	
T15	Human manipulation	Social engineering of the agent's user.	Output guardrails and audit; social engineering of users is largely out of scope.	Customer	Audit log	Out of scope	