Azure Solution Document — Red Teaming Agents.

Solution Overview

An Azure-native adversarial assurance platform that runs a coordinated **fleet of Red Teaming agents** to simulate realistic attack campaigns against cloud infra, data pipelines, and AI models. The platform supports repeated, deterministic exercises (sandboxed by default) and controlled live testing with human-in-the-loop (HITL) approvals. Outputs include attack traces, MITRE ATT&CK / ATLAS mappings, SOC playbook updates, and compliance evidence packages.

Key runtime agents referenced here: **ReconSentry, ExploitScout, TTPComposer, PhishCrafter, PayloadMutator, EvasionSim, AlertSpoofer**, plus governance agents **PolicySentinel** and **OpsOrchestrator** for HITL and policy enforcement. The platform is built to integrate with Azure Sentinel, Defender, and standard SOAR pipelines so SOC teams can ingest and act on agent telemetry.

Why this matters: Phase-1 recon found exposed credentials, frontend API keys, and plaintext PII — the platform targets exactly these weak points with automated, repeatable tests to reduce risk and improve detection.

Problem Statement

Organisations face three correlated gaps:

1. **Operational hygiene gaps** — default credentials, secrets in frontends, and exposed orchestration endpoints create easy first steps for attackers. Recon work demonstrates these exact issues.
2. **Model & pipeline blind spots** — current LLM testing often focuses on bias/toxicity while missing tool-misuse, chain-of-tool attacks and data-poisoning paths (residual risk called out by the GPT-OSS report).
3. **Lack of replayable, auditor-grade evidence** — SOCs need deterministic traces they can replay to validate detections and to produce compliance artifacts.

Agents that directly target these problems: **ReconSentry** (attack surface mapping), **ExploitScout** (CVE & exploit path discovery), **PhishCrafter** (social engineering), **TTPComposer** (multi-stage attack orchestration), and **PolicySentinel** (governance).

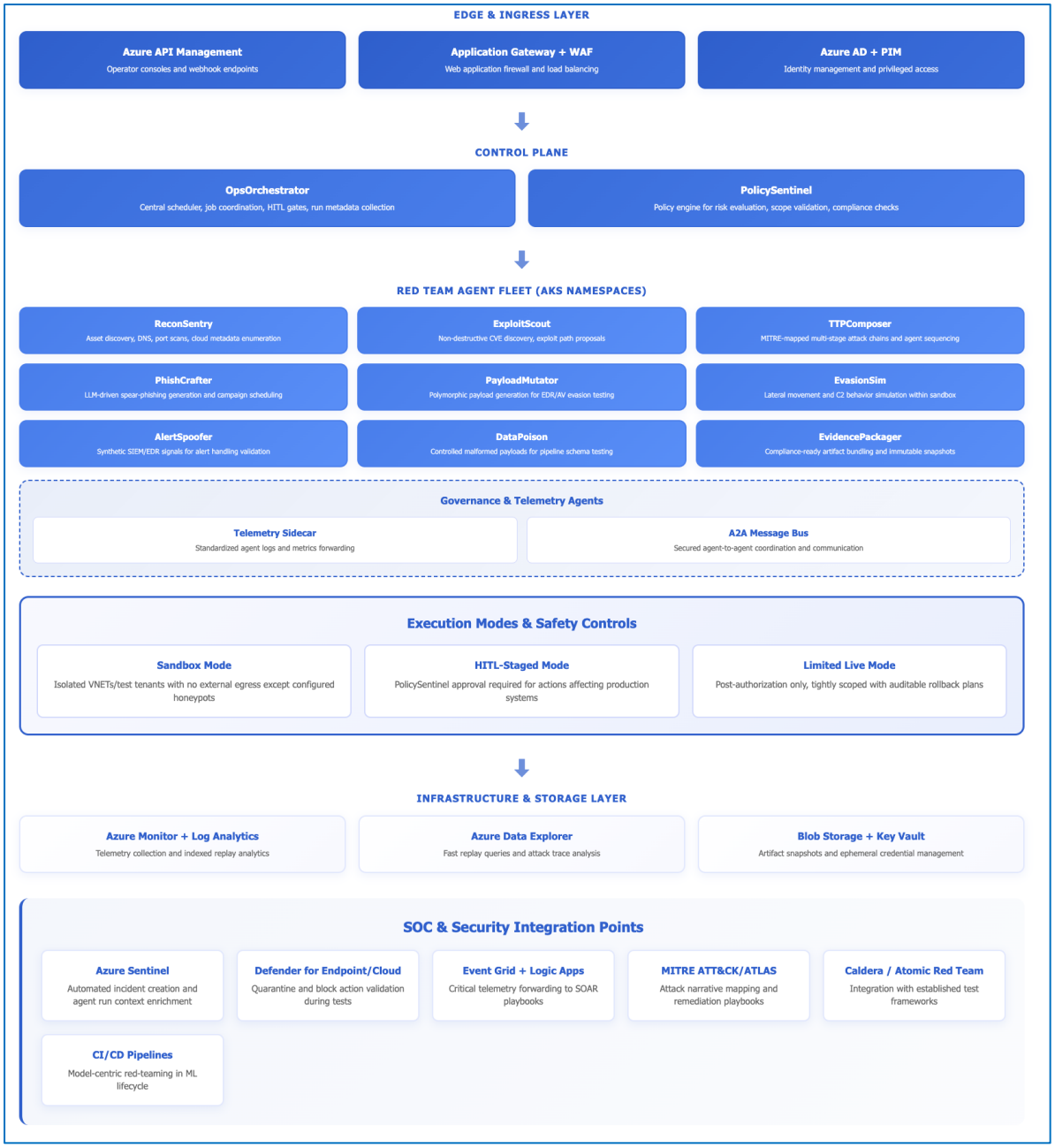
Solution Detail

This section explains how the platform works end-to-end.

1. **Agent Architecture**
    a. Each Red Team agent is implemented as a containerized microservice in AKS. Agents have clearly defined capabilities and role separation (e.g., ReconSentry only collects open-source intelligence; ExploitScout performs non-destructive CVE scanning against scoped endpoints).
    b. Agents coordinate through a secured A2A (agent-to-agent) bus managed by the **OpsOrchestrator**. **TTPComposer** composes steps and instructs agents to run in sequence or parallel.
2. **Execution Modes & Safety**
    a. **Sandbox Mode:** All agents run inside isolated VNETs/test tenants; no external network egress except to configured honeypots.
    b. **HITL-Staged Mode:** OpsOrchestrator requests a PolicySentinel approval when actions could affect production (e.g., limited live exploit simulation).
    c. **Limited Live Mode:** Only permitted after authorization — tightly scoped, auditable, with rollback plans.
3. **Testing Types**
    a. **Recon & Surface Mapping** (ReconSentry)
    b. **Phishing & Social Engineering** (PhishCrafter)
    c. **Payload Evasion** (PayloadMutator) and **Detection Bypass** (AlertSpoofer)
    d. **Vulnerability Chaining** (ExploitScout + TTPComposer)
    e. **Data Poisoning & Pipeline Attacks** (specialized DataPoison agent / pipeline fuzzers)
    f. **A2A Attack Chains** (TTPComposer coordinating agent sequences)
4. **Outputs & Artifacts**
    a. Deterministic replay logs for every run (used by SOC and auditors)
    b. MITRE ATT&CK/ATLAS mapped attack narratives and remediation playbooks
    c. SOC actionable items: rule changes, SOAR playbook updates, endpoint policy hardening

Technical Architecture

A recommended Azure reference architecture:



- **Edge / Ingress**: Azure API Management + WAF for operator consoles and web hooks
- **Control Plane**: OpsOrchestrator service running in AKS, exposes job scheduling, HITL gates, and run metadata.
- **Agent Plane**: AKS namespaces host agents; each agent runs with a sidecar that forwards telemetry to Log Analytics. Agents include ReconSentry, ExploitScout, TTPComposer, PhishCrafter, PayloadMutator, EvasionSim, AlertSpoofer, plus governance agents PolicySentinel and EvidencePackager.
- **Telemetry & Replay**: Azure Monitor → Log Analytics → Azure Data Explorer for indexed replay and analytics; snapshots stored in Blob Storage for reproducibility.

- **Secrets & CI/CD**: Azure Key Vault stores ephemeral credentials used during tests; pipelines in Azure DevOps or GitHub Actions inject short-lived service principals into sandboxes only.
- **SOC Integration**: Event Grid + Logic Apps forward critical telemetry to Azure Sentinel and SOAR playbooks; Sentinel notebooks and workbooks visualize MTTD/MTTR and detection rates.
- **Network Isolation**: Sandboxes use separate subscriptions and VNETs with explicit peering rules; outbound egress is controlled and audited.

(Include diagram in template: operator/UI → APIM/WAF → AKS (Orchestrator + Agents) → Key Vault & Blob → Log Analytics → Sentinel/SOAR → HITL dashboard)

Key Components

- **OpsOrchestrator** — central scheduler; enforces run policies, collects run metadata, coordinates HITL requests.
- **PolicySentinel** — policy engine that evaluates risk level, scope, and compliance posture before allowing live actions.
- **Agent Suite:**
    - **ReconSentry** — asset discovery, DNS, port scans, cloud metadata enumeration.
    - **ExploitScout** — non-destructive CVE discovery, contextual exploit path proposals.
    - **TTPComposer** — assembles MITRE-mapped multi-stage chains and sequences agents.
    - **PhishCrafter** — LLM-driven spear-phish generation and campaign scheduling.
    - **PayloadMutator** — polymorphic payload generation for EDR/AV evasion tests.
    - **EvasionSim** — simulates lateral movement and C2 behavior within sandbox.
    - **AlertSpoofer** — injects synthetic SIEM/EDR signals to test alert handling and false-positive management.
    - **DataPoison** (optional) — injects controlled malformed payloads into ingestion pipelines for schema/poisoning tests.
- **Telemetry Sidecar** — standardizes agent logs and metrics for replay.
- **EvidencePackager** — bundles run artifacts into compliance-ready packs (immutable snapshots).

Integration Points

- **Azure Sentinel** — automated incident creation and enrichment with agent run context.
- **Defender for Endpoint & Cloud** — validate whether quarantine and block actions occur as expected during tests.

- **Azure Key Vault** — store ephemeral credentials and rotated tokens used in exercises.
- **Caldera / Atomic Red Team connectors** — map agent behaviors to established test frameworks.
- **CI/CD** — run model-centric red-teaming (prompt injection, schema fuzzing) pre-merge or post-deploy.
- **ITSM** — auto-create tickets for remediation items discovered during runs.

Use Cases

- **Continuous Model Red-Teaming** — run automated injection campaigns after every model update to detect tool misuse and data leakage paths (uses TTPComposer, PhishCrafter, DataPoison).
- **SOC Maturity Exercises** — AlertSpoofer and EvasionSim create noise and sophisticated attack signals to measure SOC triage and fatigue.
- **Phishing + Endpoint Chain** — PhishCrafter crafts emails, ExploitScout identifies likely hosting CVEs, EvasionSim simulates lateral pivots to measure containment.
- **Post-Breach Tabletop Automation** — TTPComposer produces a replayable scenario to run purple-team exercises for dwell-time reduction.

Customer Pain Points Addressed

- **Inability to test multi-agent attack chains** — solved by TTPComposer coordinating agent sequences.
- **Secrets and credential hygiene failures** — platform exercises include ReconSentry discovery tests and automated remediation recommendations for Key Vault adoption.
- **Lack of SOC replay artifacts** — telemetry sidecars + EvidencePackager produce audit-grade bundles for compliance.
- **Narrow LLM test scope** — integrate model-centric agents to extend beyond bias tests into exfiltration and tool misuse.

Industry-Specific Applications

- **Financial**: simulate fraud pipelines that exfiltrate transaction records (DataPoison, EvasionSim, ExploitScout).
- **Healthcare**: test PHI masking and pipeline integrity (ReconSentry to find exposed endpoints; DataPoison to validate pipeline robustness).
- **Manufacturing / OT**: use EvasionSim to test IT→OT pivot scenarios.
- **Retail**: run large-scale phishing simulations (PhishCrafter) and endpoint lockdown validation (Defender integration).

Sample Customer Journey

1. **Onboard & Legal Clearance** — define scope, SLA, blast radius, and obtain risk approvals.
2. **Baseline Recon** — ReconSentry runs passive and active discovery; ops receives a baseline risk map.
3. **Sandbox Campaigns** — TTPComposer orchestrates sequences involving ExploitScout, PayloadMutator and AlertSpoofer; results streamed to Log Analytics.
4. **HITL Review** — PolicySentinel triggers an OpsOrchestrator approval workflow for any scenario that reaches live systems.
5. **Remediation & Verification** — SOC updates rules and playbooks; re-run targeted campaigns to verify fixes.
6. **Compliance Pack** — EvidencePackager produces a timestamped run bundle for auditors.

Technical Requirements

- Azure resources: AKS, API Management, Application Gateway + WAF, Key Vault, Blob, Log Analytics, Data Explorer, Sentinel, Event Grid, Logic Apps.
- Networking: isolated test subscriptions, VNET peering, NSGs and outbound egress controls.
- Identities: Azure AD with PIM for approvals and short-lived run principals.
- Data: sacrificial datasets and anonymized PII for realistic scenarios.
- Personnel: Red Team operators, SOC analysts, ML engineers, Legal/Risk approvers.

Security Architecture

- **Namespace isolation:** each agent group runs in its own AKS namespace with least-privilege RBAC.
- **Key control:** no secrets in images; Key Vault with RBAC and Managed Identities only.
- **Network posture:** default-deny NSGs; explicit egress rules for sandbox to prevent unintended exfiltration.
- **Telemetry immutability:** append-only logs in Log Analytics and snapshots to Blob with retention policy.
- **LLM guardrails:** models used by PhishCrafter or TTPComposer include content filters and are restricted to test templates unless explicit Permission Flag is set.

Performance Considerations

- Plan node pools by agent type: CPU-heavy (ReconSentry, ExploitScout), bursty (PhishCrafter), long-running (EvasionSim). Use autoscaling and spot instances for cost efficiency.
- Telemetry pipeline: batch from sidecars with backpressure handling, store in Data Explorer for fast replay queries.
- Artifact retention policies: archive older run artifacts to cool storage while keeping last 90 days hot for replay.

Tools & Azure Services Used

- **Compute & Orchestration:** AKS, node pools, Horizontal Pod Autoscaler
- **Security & Gateway:** API Management, Application Gateway + WAF, Azure AD, Key Vault
- **Observability:** Azure Monitor, Log Analytics, Azure Data Explorer
- **Automation & Integration:** Event Grid, Logic Apps, Azure Functions, Azure DevOps / GitHub Actions
- **SOC:** Azure Sentinel, Defender for Endpoint / Cloud

Users of Agents

- **Red Team Engineers** — design and operate agent campaigns.
- **SOC Analysts** — consume telemetry, tune detections, and validate playbooks.
- **ML Engineers** — use model-centric tests to validate agent-facing models.
- **Compliance / Audit** — receive evidence packs for regulation needs.
- **Security Leadership** — review dashboards and high-level risk reduction metrics.

Dependencies

- Pre-approved sandboxes and datasets
- SOC connectors and logging enabled (Sentinel/Defender)
- CI/CD pipelines to run model tests and to store test harnesses
- Legal and risk signoff for staged live testing

Key Benefits & Differentiators

- **Full-chain adversarial coverage:** from reconnaissance to exfiltration, including model-specific attack vectors.
- **Governed automation:** HITL gates and PolicySentinel reduce operational danger of automated tests.

- **Replayability at scale:** deterministic traces enable SOC tuning and auditor validation.
- **CI/CD integration:** prevents regressions by embedding model-centric tests into the ML lifecycle.

Value Proposition

- Lower total risk through continuous, automated adversarial testing.
- Faster SOC improvement cycles: measurable MTTD/MTTR reduction via replayable training scenarios.
- Audit readiness: compliance evidence produced automatically after each campaign.
- Solve hygiene problems surfaced during initial reconnaissance and reduce future blast radius.

Conclusion

A Red Teaming Agents platform on Azure provides a practical, governed, and repeatable approach to adversarial assurance for both infrastructure and AI systems. By orchestrating a fleet of specialized agents (ReconSentry, ExploitScout, TTPComposer, PhishCrafter, PayloadMutator, EvasionSim, AlertSpoofer, PolicySentinel, OpsOrchestrator), organizations can surface and fix real-world vulnerabilities before they are exploited and produce audit-grade evidence for regulators.