

The AI Vendor Due Diligence Cheat Sheet: (for Corporate AI Buyers and Users)

By Rezliant Inc. <https://rezliant.com/>

This cheat sheet equips technology leaders to assess and vet AI vendors for the common AI security weaknesses identified in recent breaches of AI systems.

1. Credential & Access Management

- ☐ Are admin and system credentials strong, rotated, and MFA-enabled?
- ☐ Are test/default accounts removed from production systems?
- ☐ Are role-based access controls (RBAC) implemented for all AI services?

2. Data Storage & Exposure

- ☐ Are AI training data, logs, and user inputs properly secured (encryption at rest/in transit)?
- ☐ Are there automated scans for publicly exposed databases, storage buckets, or docs?
- ☐ Is multi-tenant isolation enforced to prevent cross-customer data leaks?

3. Input Sanitization & Prompt Handling

- ☐ Are all external inputs (emails, calendar events, uploads) validated/sanitized before AI ingestion?
- ☐ Are AI agents sandboxed with strict capability limits?
- ☐ Is taint tracking or provenance used for all prompts to prevent unauthorized actions or exfiltration?

4. Vendor Supply Chain & Extension Security

- ☐ Are third-party AI modules or extensions verified and signed?
- ☐ Are CI/CD pipelines protected against malicious commits or code injection?
- ☐ Are audit logs maintained for deployment of AI modules and extensions?

"... a world where every software company can confidently release secure code knowing they protect not just data or IP, but the livelihoods of all stakeholders"

Rezliant Inc | <https://rezliant.com> & <https://resilientsoftwaresecurity.com>

5. Safety & Guardrails

- ☐ Are adversarial red-team tests performed regularly on AI models?
- ☐ Are layered safety filters applied (model-level + application-level)?
- ☐ Is there continuous monitoring of AI outputs for toxic, unsafe, or sensitive responses?

6. Change Management & Insider Risks

- ☐ Are changes to system prompts or AI configuration audited and peer-reviewed?
- ☐ Are alerts triggered for anomalous configuration or model changes?
- ☐ Is access to critical AI configurations restricted using RBAC?

7. Authentication & Multi-Factor Verification

- ☐ Are high-risk AI-driven actions (fund transfers, data sharing) protected by MFA?
- ☐ Are biometric/voice verifications liveness-tested to prevent AI impersonation?
- ☐ Are human approvals required for critical autonomous AI outputs?

8. Monitoring & Anomaly Detection

- ☐ Are AI inputs, outputs, and system logs continuously monitored for unusual behavior?
- ☐ Are alerts in place for cross-tenant access attempts or large-volume AI data requests?
- ☐ Are suspicious prompt sequences or chain-of-agent behaviors flagged automatically?

9. Regulatory & Privacy Compliance

- ☐ Does the vendor comply with applicable data privacy laws (GDPR, CCPA, SOC 2/ISO 27001)?
- ☐ Are cross-border data flows mapped and controlled?
- ☐ Is there a plan for breach response specific to AI data?

10. Human Oversight & Governance

- ☐ Are humans in the loop for high-risk AI actions?
- ☐ Are AI decisions logged and auditable?
- ☐ Is there a formal AI governance policy including risk assessment, mitigation, and vendor management?

"... a world where every software company can confidently release secure code knowing they protect not just data or IP, but the livelihoods of all stakeholders"

Rezliant Inc | <https://rezliant.com> & <https://resilientsoftwaresecurity.com>