CORUZANT TECHNOLOGIES · CHECKLIST

AI Production Readiness Checklist

Review one workload at a time. For each dependency, collect the evidence, ask the readiness question, and mark whether the answer is only stated, verified, or a gap.

Workload: _____ Review date: _____ Target RTO / RPO: _____

Table 1: AI workload dependency map

Dependency	Readiness question	Evidence to collect	Failure or gap to surface	Status	Notes / owner
Source data and producer	Are sources identified, owned, and permitted for this workload?	Data inventory; producer and consumer owners; access and use constraints	A source exists but is inaccessible, stale, or not approved for the intended use	☐ Stated ☐ Verified ☐ Gap	
Data contract and change handling	Are schema, semantics, quality, freshness, and change rules explicit? Which are actually validated?	Versioned contract; compatibility policy; validation results; documented unvalidated obligations	Schema passes while meaning, freshness, or quality has changed	☐ Stated ☐ Verified ☐ Gap	
Data delivery and storage	Can ingestion, transformation, and storage meet volume, latency, retention, and recovery needs?	Throughput and freshness observations; storage and retention design; backup or replication evidence	The pipeline or storage becomes the bottleneck even when accelerators are available	☐ Stated ☐ Verified ☐ Gap	
Compute and capacity	Does the hardware fit the workload, and can the required capacity be obtained?	Workload sizing assumptions; quota; reservation or procurement status; scaling plan	Quota exists but capacity is unavailable, or accelerators are insufficient for the target workload	☐ Stated ☐ Verified ☐ Gap	
Network and orchestration	Can the workload schedule and communicate across required nodes and services?	Network and topology requirements; scheduler configuration; deployment dependencies	A multi-node workload cannot operate as assumed because of topology, scheduling, or network limits	☐ Stated ☐ Verified ☐ Gap	
Model and serving dependencies	Are model artifacts, runtime, endpoint, and supporting services available at rollout and recovery?	Artifact and version inventory; serving architecture; dependency owners and availability expectations	Recovery restores the application code but not the model, runtime, or serving capacity	☐ Stated ☐ Verified ☐ Gap	
Identity, security, and policy	Can authorized components access required data and services, including at recovery time?	Identity and access design; classification and policy records; recovery-environment permissions	The failover environment lacks permissions, secrets, or approved access paths	☐ Stated ☐ Verified ☐ Gap	
Recovery objectives and mechanisms	Are workload-level RTO and RPO agreed, and is the recovery approach consistent with them?	Business-approved targets; dependency-level recovery assumptions; backup or replication approach	Component targets look adequate but full workload restoration is slower or loses more data	☐ Stated ☐ Verified ☐ Gap	

Table 2: Ownership and sign-off questions

Evidence or decision	Owner to identify	Review question	Named owner	Signed off
Producer-consumer contract, versions, and change rules	Data producer and consuming data or platform owner	Who approves a breaking change, and how are downstream consumers told?		☐
Validation coverage and exceptions	Data quality or data platform owner	Which contract terms are automatically tested, and which are only documented?		☐
Source permissions, classification, and permitted use	Data steward, security, and privacy or compliance owner	Are the intended data and recovery copies usable under the same controls?		☐
Workload sizing assumptions and accelerator needs	ML platform or infrastructure owner, with the workload team	What workload profile supports the estimate, and what assumptions are unverified?		☐
Quota, reservation, supply, and scale-out route	Cloud or infrastructure capacity owner	Is capacity requested, or actually available for rollout and expected growth?		☐
Storage, network, orchestration, service dependencies	Platform owner for each dependency	Are dependencies available at the required scale and in the recovery location?		☐
Business RTO and RPO and service priority	Business service owner, with the resilience or IT owner	What downtime and data loss are acceptable for this workload, not just one component?		☐
Recovery approach, backups or replication, target validation	Resilience or DR owner, with application and data owners	Does the approach plausibly meet the approved targets, and where will testing be recorded?		☐

If a review question has no clear owner, the honest status is not green. From "The AI Infrastructure Readiness Review" on coruzant.com.