

Are You Ready to Deploy GenAI in Production?

This checklist is for CTOs and Founders of early to mid-stage startups experimenting with Generative AI.

Whether you're fine-tuning a model or embedding LLMs into your product, this guide will help you assess whether your stack, processes, and team are ready for a real-world GenAI rollout.

GenAI Production Readiness Checklist

- ☐ YES We have a clearly defined GenAI use case tied to a business outcome (not just R&D).
- ☐ YES Prompt engineering and tuning are version-controlled and testable.
- ☐ PARTIAL Our team can track and observe token usage, latency, and error rates in production.
- ☐ NO We have automated retraining or refresh cycles for evolving data/model performance.
- ☐ YES We're using orchestration tools like Step Functions or agents to manage flow control.
- ☐ YES There's a fallback or retry mechanism for prompt failures or inference timeouts.
- ☐ PARTIAL We've implemented IAM roles and audit logging for GenAI API access.
- ☐ YES Our CI/CD pipelines support model + infra deployments (not just app code).
- ☐ NO We've tested GenAI behavior under load and edge-case prompts.
- ☐ YES We've projected and budgeted for cloud cost at scale (tokens, compute, storage).
- ☐ PARTIAL Prompt inputs/outputs are being logged securely for iteration and RLHF.
- ☐ YES We've defined what success looks like for GenAI adoption in our product roadmap.

AWS Tools for GenAI: What to Use and When

- ☐ NO Our LLM provider or stack allows us to swap models if needed (no vendor lock-in).
- ☐ YES Our team understands and enforces safety guardrails for model outputs.
- ☐ PARTIAL We've assigned shared ownership of GenAI features across product + engineering.

How did you score?

- 12-15 checks** You're production-ready or close. Tidy up the edges and go.
- 8-11 checks** You're getting close. Focus on infra observability, failure modes, and process maturity.
- <8 checks** Start with architectural fundamentals and define your GenAI roadmap.

NEXT STEPS

Want help getting to full readiness?

Book a free readiness audit with our cloud + AI architects. We'll score your stack, highlight gaps, and outline a plan to launch GenAI features that scale without surprises.

VeUP Build is built for exactly this moment.