# Azure GenAI Development Account – Easy Setup & Cost Guide (Oct 2025)

**Goal:** Provide a simple, easy-to-read checklist and cost guide for setting up an **Azure account for Generative AI development** (no production region selected).

## 1. Core Services Setup

* **Azure Subscription** – one for Dev/Test
* **Identity & Access:** Azure Entra ID, RBAC roles, Key Vault for secrets
* **Networking:** VNet with private endpoints, NAT gateway
* **DevOps Tools:** Azure DevOps (Repos, Boards, Pipelines, Artifacts) or GitHub
* **CI/CD:** Pipelines or GitHub Actions for automated deployments
* **Container Runtime:** Azure Container Apps (easy scaling, low cost)
* **Artifact Storage:** Azure Container Registry (ACR)

## 2. AI & Data Services

* **Azure OpenAI Service** – models (GPT‑4o, embeddings)
* **Vector DB Options:**
  + Azure AI Search (recommended for RAG)
  + Cosmos DB (Mongo vCore) with vector search
  + PostgreSQL Flexible Server with pgvector
* **Storage:** Azure Blob Storage for documents & datasets
* **Content Safety:** Azure AI Content Safety for text/image moderation

## 3. Monitoring & Security

* **Observability:** Azure Monitor + Application Insights
* **Logs:** Log Analytics Workspace (Basic Logs for Dev/Test)
* **Key Management:** Azure Key Vault
* **Security:** Defender for Cloud, Private Endpoints, Managed Identity

# Cost Reference (Simple Estimates, USD)

*Note: Costs vary slightly by region. These are indicative monthly values for planning.*

### Free / Included

* **Azure Container Apps** – free grant: 180k vCPU‑seconds, 360k GiB‑seconds, 2M requests/month
* **Cosmos DB Free Tier** – 1,000 RU/s + 25 GB storage
* **Azure AI Content Safety** – 5,000 text/image checks/month

### Low Usage (POC / Dev)

* Container Apps (light overage): ~$20–$50
* AI Search S1 (1 SU): ~$75
* Blob Storage 250 GB Hot: ~$5
* Key Vault ops: <$1
* DevOps (basic, 1 hosted job): Free
* **Total (excluding model tokens): ~$100–$200/month**

### Medium Usage (Pilot)

* Container Apps (autoscale): ~$100–$300
* AI Search S2 (1–2 SU): ~$200–$400
* Blob Storage 1 TB Hot/Cool mix: ~$20–$25
* Content Safety: 500k text checks: ~$190
* Logs: 30 GB/month: ~$60
* DevOps: 2 hosted jobs + add-ons: ~$100
* **Total (excluding model tokens): ~$600–$1,000/month**

### Higher Usage (Team)

* Container Apps/AKS small cluster: ~$400–$800
* AI Search S3 (2–4 SU): ~$800–$1,500
* Blob Storage 2 TB: ~$40–$50
* Logs: 100 GB/month: ~$200
* Content Safety: 1M text checks: ~$380
* DevOps: multiple jobs/users: ~$200–$300
* **Total (excluding model tokens): ~$2,000–$3,000+/month**

# Quick Notes

* **Model Costs:** Azure OpenAI billed by tokens (input + output). Expect most spend here depending on usage.
* **Save Costs:** Use scale-to-zero in Container Apps, Blob lifecycle (Hot → Cool → Cold), free Cosmos DB tier, and Batch API for embeddings.
* **Best Start:** For Dev/Test – Container Apps, AI Search (S1), Blob Storage, Azure OpenAI (on‑demand), DevOps free plan.

✅ This setup gives you everything needed for **basic generative AI app development** (Dev/Test only, no production).