Focus20 Agentic AI Playbook

Engineering Autonomous Revenue & Operations

Executive Summary

The era of conversational chatbots is over. Enterprise value is now driven by 'Agentic AI' — autonomous systems capable of reasoning, breaking down complex objectives into multi-step workflows, and securely interacting with internal tools (APIs, Databases, ERPs) to execute tasks end-to-end.

This playbook outlines Focus20's architectural standards for deploying Reasoning & Acting (ReAct) frameworks securely within regulated environments. By replacing human-in-the-loop bottlenecks with autonomous agents, enterprises can achieve 40-70% improvements in operational velocity while maintaining strict compliance guardrails.

1. The ReAct Framework

Agentic AI relies on a cognitive loop consisting of:

Thought: The Large Language Model (LLM) assesses the current state constraints and identifies the next logical step.

Action: The LLM selects a specific tool (e.g., executing a SQL query, calling a Salesforce API, or running a Python script) and generates the required payload.

Observation: The system executes the tool and returns the result to the LLM. The LLM then reasons over the observation to determine if the objective is met.

2. Enterprise Architecture Patterns

Focus20 implements highly scalable agentic architectures leveraging cloud-native primitives, predominantly on AWS:

Vector Stores & Retrieval Augmented Generation (RAG): To eliminate context window hallucinations, we utilize Amazon OpenSearch Serverless or Pinecone. Proprietary data is chunked, embedded via models like Titan or text-embedding-ada-002, and dynamically injected into the agent's prompt based on semantic relevance.

Multi-Agent Orchestration: For complex tasks, relying on a single monolithic LLM prompt fails. We utilize LangGraph and AWS Step Functions to coordinate specialized sub-agents. For example, a 'Coder Agent' drafts Terraform scripts, an 'Auditor Agent' reviews for security compliance, and an 'Execution Agent' deploys via an approval gate.

3. Zero-Trust AI Security

Deploying autonomous systems requires stringent safeguards:

Prompt Injection Protection: Utilizing deterministic guardrails (e.g., NeMo Guardrails) to validate inputs before they reach the reasoning engine.

Least Privilege Execution: API mutations generated by agents are executed within isolated, ephemeral AWS Lambda containers with rigidly defined IAM roles.

Human-in-the-Loop (HITL) Checkpoints: High-risk actions (e.g., financial transactions, production infrastructure changes) mandate programmatic suspension until explicit human semantic approval is captured.

Conclusion

Deploying Agentic AI is an engineering challenge, not just a data science experiment. Focus20 provides the battle-tested blueprints to transition your organization from fragmented automation to intelligent autonomy. Reach out to our engineering team to formalize your Agentic strategy.