Secure AI Adoption and Deployment

This AI cybersecurity training course teaches experienced developers and entry-level AI engineers how to build and operate production AI applications responsibly. Address real-world security gaps and operational risks hands-on in Google Colab, enhancing capabilities and hardening controls at every step. Modules cover prompting, context management, hallucination mitigation, RAG, tool integration, AI agents, security guardrails, observability, and deployment.

CBT Nuggets course material →

STUDY PLAN

WEEK 1

Prepare for Secure AI Development

The Course Arc and Google Colab

Local Notebooks, IDEs, and APIs

Challenge

Explain Software 1.0, 2.0, and 3.0

Introduction

API Keys and the Software Eras

Deterministic, Learned, and Generative Software

Building a Classifier with Teachable Machine

Challenge

Build a Modern AI Application

Introduction

The Model Is Not the Application

Setting Up the API Client

Anatomy of the First Request

Challenge

Apply Prompts, Context, and Model Controls

Introduction

Configuration and the Client Helper

System Prompts and the Request Builder

Sending Requests and Conversation History

Temperature and Model Deprecations

Challenge

Explain Model Knowledge and Limitations

Introduction

The Model Is Not a Database

Why Models Hallucinate

Supplying Runtime Context

Challenge

Ground AI with Retrieval and RAG

Introduction

The RAG Loop

Chunking and Similarity

Building the RAG Pipeline

Challenge

Integrate AI Models with Software Tools

Introduction

The Tool Call Lifecycle

The Tool Schema

Validate, Execute, and Return

Challenge

Build a Simple AI Agent

Introduction

What Is an Agent?

Giving the Agent a Goal and a Tool

Bounded Loops and Stop Conditions

The Agent Loop

Challenge

Choose Between Workflows and Agents

Introduction

The Autonomy Spectrum

Building the Workflow

The Agentic Version and the Trade-Offs

Challenge

Evaluate AI Application Quality

Introduction

The Application and Its Test Cases

Deterministic Checks and Quality Dimensions

Human Review, Model Graders, and the Harness

Pass Rates, Failures, and the Regression Gate

Challenge

Improve AI Reliability and Guardrails

Introduction

Building a Deliberately Bad Assistant

Guardrail Layers, Fallbacks, and Bounded Retries

Grounding, Retries, Fallbacks, and Escalation

Challenge

Secure AI Applications and Data

Introduction

Trust Boundaries and Authentication Before Prompting

Authorization, Protected Data, and Minimization

Redaction, Prompt Injection, and the Secure Request Path

Challenge

Optimize AI Performance and Cost

Introduction

Measuring the Whole Request Path

Tokens, Cost, and Context Management

Model Routing, Output Budgets, Caching, and Streaming

Routing Requests and Scoring Optimizations

Challenge

Monitor AI Applications in Production

Introduction

Following One Request with Structured Events

Spans, Safe Logging, and Key Metrics

Online Evaluation, Alerts, and the Monitoring Record

Challenge

Deploy and Operate AI Applications

Introduction

Manifests, Configuration, and Secrets

Release Bundles and Release Gates

Canary Rollouts, Rollback, and Incident Response

Challenge

Design a Production AI System

Introduction

Responsibilities and the Deterministic Control Shell

Workflow or Agent and the Production Control Plane

Readiness, the Decision Record, and Go or No Go

Challenge

Last updated 9.18.2026