QUICK RECAP

In the previous week, we explored AI Agents. We revisited Large Language Models (LLMs), understood reasoning in LLMs, and how it extends into agentic AI capabilities. We also studied how AI agents operate, interact, and are enhanced through architectures, communication mechanisms, and memory. Finally, we introduced Model Context Protocol (MCP) servers and their role in enabling more robust agentic AI workflows.

WEEK OVERVIEW

This week, we shift our focus to Responsible Generative AI Solutions. The goal is to understand the risks, challenges, and necessary safeguards when deploying LLM-powered systems in real-world environments.

We will begin by comparing classical software systems and LLMs, then move into building the right mental model for how LLMs function. A major part of this week will focus on the need for guardrails, common security risks (including confidentiality, integrity, and availability), and real-world risk examples.

Finally, we will explore risk mitigation layers and discuss what a secure Generative AI solution architecture looks like, along with key considerations for building solutions responsibly.

The following topics will be covered in this module:

- The Difference between Classical Software and LLMs
- Common Security Risks in Large Language Models
- Common Risk Mitigation Layers in Generative AI Solutions
- Overview of a Secure Generative AI Solution Architecture
- Considerations for Building a Secure Generative AI Solution

LEARNING OBJECTIVES

By the end of this module, you will be able to:

- Differentiate between classical software and LLMs and build a correct mental model of how LLMs work.
- Understand why guardrails are essential for responsible deployment of LLMs.
- Identify common security risks in LLMs, including confidentiality, integrity, and availability challenges.
- Examine real-world examples of security risks in Generative AI.
- Explore different risk mitigation layers and their role in strengthening Generative AI systems.
- Describe the components of a secure Generative AI solution architecture.

LEARNING INSTRUMENTS

Week	Week Name	No. of Videos	Total Duration	No. of Test Your Understanding Quizzes	No. of Graded Quizzes	No. of Practice Assignments
3	Responsible Generative AI Solutions	14	~2 hours	13	1	1

[illegible]