

In this week, titled "**Ethics, Safety, Alignment and Responsible AI**," you will delve into the foundational concepts and practices guiding the ethical development of artificial intelligence, emphasizing the importance of creating AI systems that perform effectively while aligning with societal values. You will explore the **core principles of AI ethics**, which serve as a framework for responsible technology development that prioritizes human welfare, and learn about **measuring intangible aspects** such as confidence and trust through appropriate metrics. Key topics include identifying **agentic risk factors** associated with AI systems, understanding potential consequences of unethical practices, and analyzing **real-world examples of ethics and safety failures** in the field. Additionally, you will gain insights into various **techniques for building safer AI**, equipping you with the knowledge to implement ethical guidelines and safety measures in your projects. By the end of the module, you will have a comprehensive understanding of ensuring alignment between AI systems and human values, paving the way for the responsible and ethical deployment of AI in the future.

Learning Objectives

After completing this week, you will be able to:

- 1. Recognize the foundational concepts that underpin ethical considerations in artificial intelligence development.
- 2. Describe key principles that guide the responsible design and implementation of AI systems.
- 3. Evaluate methods for measuring essential intangible factors, such as confidence and trust, in AI applications.
- 4. Identify potential risks associated with agentic AI systems and understand their implications for safety and ethics.
- 5. Analyze real-world cases of ethical and safety failures in AI to reinforce lessons learned.
- 6. Implement effective strategies and techniques to develop safer AI systems that align with ethical standards.

Topics Covered

Week #	Week Name	Topics Covered
8	Ethics, Safety, Alignment and Responsible AI	<ul style="list-style-type: none">1. Foundational concepts guiding ethical AI development.2. Principles for responsible and ethical technology design.

		6. A practical hands-on application to showcase the implementation of Responsible AI
--	--	--

Learning Material

Week#	Week Name	#Videos	Duration (in hrs)	# Test Your Understanding (TYU) Quizzes	Practice Quiz	Graded Quiz
8	Ethics, Safety, Alignment and Responsible AI	10	~1.40	6	1	1

Happy Learning !!

