# Lesson 6: Crawling

## Objective:

Understand web scraping (crawling) techniques and how they can be used to gather external knowledge for your AI models.

## Key Concepts:

- **Web Crawling:** The process of automatically navigating web pages and extracting relevant data.
- **Scraping Tools:** Tools such as BeautifulSoup, Scrapy, or Selenium for Python are used to scrape data from the web.
- **Legal and Ethical Considerations:** Respect robots.txt files and ensure your scraping activities comply with website terms of service.

## Application:

1. **Scraping Relevant Data:**

   o  Write crawlers that extract structured data (e.g., news articles, product information) from websites and use it to feed into your AI system.

2. **Data Processing and Clean-up:**

   o  Parse, clean, and store the scraped data for later use.

## Application:

1.

## Building a Web Crawler:

1. Use libraries like **BeautifulSoup** or **Scrapy** to scrape structured data from websites, such as articles, product listings, or news stories.

## Data Cleaning:

After scraping, clean and process the data before feeding it into your AI model.