Web scraping with ChatGPT (Medium level training)

Duration: 24 Hours, 3 Days per week, 2 Hours per day

Module	Major Categories	Duration
Module 01	Introduction to Web Scraping and ChatGPT	6 Hours
Module 02	Advanced Web Scraping Techniques with ChatGPT	6 Hours
Module 03	Selenium for Web Scraping	6 Hours
Module 04	Real-World Projects and Case Studies	4 Hours
Module 05	Final Assessment and Certification	2 Hours

Class No	Module Details	Duration
	Module 1: Introduction to Web Scraping and ChatGPT	
01	Session 1: Getting Started (2 hours) 1.1.1 Introduction to Web Scraping: • Overview of web scraping and its applications. • The importance of data extraction from websites. • Understanding ethical considerations and legal aspects. 1.1.2 Tools and Libraries for Web Scraping: • Introduction to key tools and libraries for web scraping (e.g., Python, BeautifulSoup, Requests). • Setting up your development environment (Python installation and package management). • Basic HTML and CSS overview to understand web page structure. Session 2: Python Fundamentals (2 hours) 1.2.1 Introduction to Python for Web Scraping: • Basics of Python programming language. • Key concepts such as variables, data types, and operators. • Writing your first Python script for web scraping. 1.2.2 Web Requests with Python: • How to make HTTP requests using Python's Requests library. • Understanding HTTP methods (GET, POST, etc.). • Handling HTTP responses and status codes (e.g., 200 OK, 404 Not Found). Session 3: Introduction to ChatGPT for Web Scraping (2 hours)	6 Hours 3 Days

	1.3.1 Introduction to ChatGPT:	
	• Overview of ChatGPT and its capabilities.	
	• Use cases for ChatGPT in web scraping. • Accessing the ChatGPT API and authentication. 1.3.2 Getting Started with ChatGPT for Web Scraping: • Practical demonstration of using ChatGPT for web scraping tasks. • Extracting structured data from web pages using ChatGPT. • Examples of using ChatGPT for text extraction and analysis from websites.	
Module 2: Advanced Web Scraping Techniques with ChatGPT		
02	Session 1: ChatGPT for Text Extraction and Analysis (2 hours)	6 Hours 3 Days
	2.1.1 Text Extraction with ChatGPT:	
	• Expanding on text extraction techniques using ChatGPT.	
	• Strategies for extracting text content from various web page elements.	
	• Examples of using ChatGPT for extracting and cleaning text data.	
	2.1.2 Text Analysis with ChatGPT:	
	• Introduction to text analysis tasks with ChatGPT.	
	• Summarizing text content using ChatGPT.	
	• Building a simple keyword extractor with ChatGPT.	
	Session 2: Handling Complex Data (2 hours)	
	2.2.1 Extracting Data from Tables and Lists:	
	• Techniques for extracting structured data from tables and lists on web pages.	
	• Overview of HTML parsing and traversal.	
	• Practical exercises on extracting tabular data with ChatGPT.	
	2.2.2 Combining ChatGPT and BeautifulSoup :	
	• How to combine the capabilities of ChatGPT and BeautifulSoup for advanced data extraction.	
	• Extracting data from complex web page structures.	

	• Case study: Extracting data from a dynamic website using ChatGPT and BeautifulSoup.	
	Session 3: Best Practices and Ethical Considerations (2 hours)	
	2.3.1 Legal and Ethical Aspects of Web Scraping:	
	• In-depth discussion on the legal and ethical considerations in web scraping.	
	• Intellectual property rights, terms of service, and scraping etiquette.	
	• Strategies for avoiding IP bans and rate limiting when scraping.	
	2.3.2 Error Handling and Optimization:	
	• Techniques for handling errors and exceptions in web scraping scripts.	
	• Debugging and troubleshooting common issues.	
	• Scalability and optimization of web scraping scripts for efficiency.	
Module 3: Selenium for Web Scraping	
03	**Session 1:** Introduction to Selenium (2 hours)	6 Hours 3 days
	3.1.1 Basic concept of Selenium:	
	• Overview of Selenium as a web automation tool.	
	• Use cases and advantages of using Selenium for web scraping.	
	• Installing Selenium and configuring WebDriver for different browsers.	
	3.1.2 Navigating Websites with Selenium:	
	• Practical demonstration of navigating websites using Selenium.	
	• Opening and closing web browser windows.	
	• Navigating through web pages, including back and forward navigation.	
	Session 2: Advanced Selenium Techniques (2 hours)	
	3.2.1 Handling Dynamic Web Pages:	
	• Understanding dynamic web pages and their challenges in web scraping.	
	• Techniques for handling JavaScript-rendered content with Selenium.	

	• Case study: Scraping data from a dynamically loaded web page.	
	3.2.2 Synchronizing Web Element Interactions:	
	• Exploring synchronization strategies to ensure web elements are ready for interaction.	
	• Waiting for page elements to load using explicit and implicit waits.	
	• Handling AJAX requests and delays.	
	Session 3: Data Extraction and Automation (2 hours)	
	3.3.1 Extracting Data from Web Elements:	
	• Techniques for extracting data from various web elements (e.g., text, images, links).	
	• Interacting with forms, input fields, and buttons using Selenium.	
	• Extracting and manipulating data from HTML forms.	
	3.3.2 Automating Web Scraping Tasks with Selenium:	
	• Designing and implementing automation scripts for web scraping tasks.	
	• Combining Selenium with ChatGPT for enhanced data extraction and interaction.	
	• Building practical web scraping automation examples.	

Module 4: Real-World Projects and Case Studies

	Session 1: Building Practical Projects (2 hours)	
	4.1.1 Define a Web Scraping Project:	
	• Discuss various web scraping project ideas and their real-world applications.	
	• Guidelines for defining the scope, objectives, and requirements of a web scraping project.	
04	• Assistance in selecting a project or refining project ideas.	4 Hours 2 days
	4.1.2 Designing the Project Structure and Data Storage:	
	• Planning the project's structure and organization.	
	• Choosing the appropriate data storage methods (e.g., CSV, JSON, databases).	
	• Designing the database schema if applicable.	
	Session 2: Project Work and Troubleshooting (2 hours)	

	4.2.1 Guidance and Assistance on Project Work:	
	• Dedicated time for students to work on their chosen web scraping projects.	
	• Instructor support and guidance for overcoming project-related challenges.	
	• Reviewing and providing feedback on project progress.	
	4.2.2 Addressing Common Issues and Challenges:	
	• Troubleshooting common issues faced during web scraping projects.	
	• Strategies for handling unexpected situations and errors.	
	• Best practices for debugging and refining web scraping code.	
Module 5: Final Assessment and Certification		
05	• Final Assessment Exam	2 Hours 1 day