

# MrScraper - AI Engineering Manager - Take Home Test

## AI Chat Web Scraping

---

### Introduction

In today's rapidly evolving technological landscape, AI plays a significant role in enhancing web scraping techniques. From solving complex scraping challenges to handling unstructured data and summarizing extracted content, AI tools such as ChatGPT or DeepSeek can provide innovative solutions.

This test will help us evaluate your web scraping skills, your ability to integrate AI effectively, and your overall problem-solving capabilities.

---

### Task Description

Your challenge is to **create an API that allows chatting with an AI** about scraping a site.

Example URL: <https://medrecruit.medworld.com/jobs/list?location=New+South+Wales&page=1>

Example chat:

- 1. From the URL, crawl the pages to get all job post details.*
- 2. From the URL, get the role, location, and pay of jobs from page 1 until 3.*

The solution design needs to be general enough to accommodate these scenarios as well:

- Link: <https://www.azsoccerassociation.org/member-clubs/>  
Chat: Find emails and phone number of each club
- Link: <https://support.hp.com/us-en/drivers/printers>  
Chat: Find all printer driver details through the pages

Bonus points if your submission already supports these scenarios.

### Requirements:

#### 1. Core Requirements (Mandatory):

- 🟡 Use **Python**.
- 🟡 Use **AI** to get the data from the page.
- 🟡 Result of the API must be returned in **JSON format**.

```
{  
  "text": "...",      // reply text from the AI  
  "results": [...],   // scraping results,  
  <other field such as session ID or others>  
}
```

}

- 🟡 Scrape **data** across **pagination and nested pages**.
  - 🟡 It's an AI chat so it must have sessions. (It must be able to reply to the latest results). For example: after hitting the API with: ***get the role, location, and pay of jobs from page 1 until 3***, you can hit the API with another prompt such as: ***Filter data that have pay higher than \$2,000 per day only***.
  - 🟡 If any field does not have a value, return ' - '.
  - 🟡 Use the Model Context Protocol (MCP) approach to control a browser. Specifically, please use Zendriver (<https://github.com/cdpdriver/zendriver>) as your browser tool for this implementation.
- 

### Submission Guidelines:

1. Push your completed code to a **public GitHub repository**.
  2. Share the repository link by replying-all to this email **within one week** of receiving it.
  3. Ensure the API is well-documented in a README file, including instructions for running the project.
- 

### Interview Details:

- Will be considered after the submission
- 

We look forward to seeing your approach to solving this challenge. If you have any questions or clarifications, feel free to reach out.

Best of luck, and we are excited to discuss your solution further!