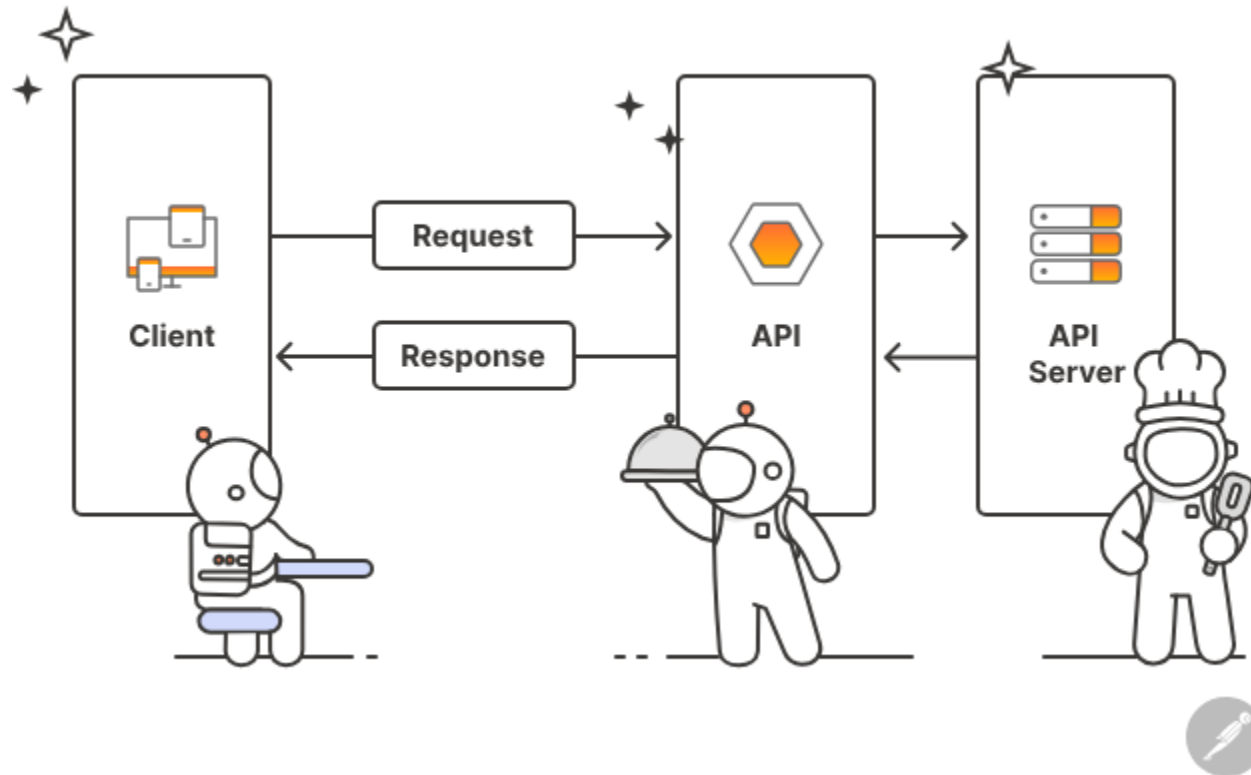# Web Scraping with API

Kai Yang

Assistant Professor

Department of Government and International Affairs

Lingnan University

Feb 25, 2026

What is API?

- Definition:
- An API or application programming interface, is used to pass data back and forth between software apps in a formalised way.

- Metaphor: Library access without ATM service

- You want to search books on "Hong Kong Politics"
- Assisted by librarians

- Written notes from librarians:
- An Intro to HK Politics, 3rd Floor, Michael Cheung, 2012, available; HK Political Reforms, 2nd Floor, 2013, John Smith, available; HK Armed Forces, James Lang, 1998, unavailable
- ...

- But for API scraping, standardized format for each item:

- {
-   "title": "An Intro to HK Politics",
-   "location": "3rd Floor",
-   "author": "Michael Cheung",
-   "year": 2012,
-   "is_available": true,
- }

```
[
  {
    "title": "An Intro to HK Politics",
    "location": "3rd Floor",
    "author": "Michael Cheung",
    "year": 2012,
    "is_available": true,
  },
  {
    "title": "HK Political Reforms",
    "location": "2nd Floor",
    "author": "John Smith",
    "year": 2013,
    "is_available": true,
  },
  {
    "title": "HK Armed Forces",
    "location": null,
    "author": "James Lang",
    "year": 1998,
    "is_available": false,
  }
]
```

Hand-written vs. standardized orders

- JSON (JavaScript Object Notation) is a lightweight, text-based, language-independent data-interchange format. It is widely used in web applications to transmit data between a server and client due to its simple, human-readable structure.