Responsible Machine Learning

Lecture 8: Algorithm Auditing Tutorial

CS 4973-05

Fall 2023

Instructor: Jeffrey Gleason
gleason.je@northeastern.edu
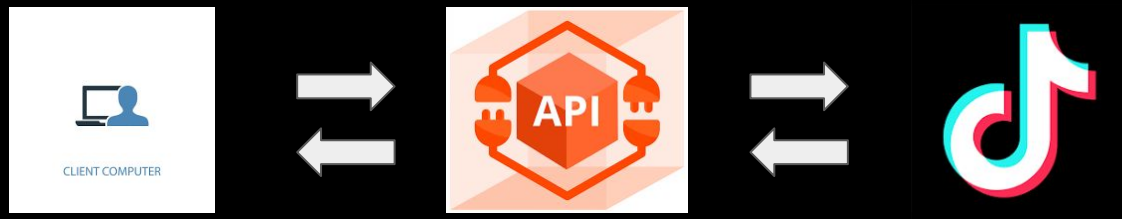**Northeastern University, Boston, MA**

Northeastern
University

Agenda

1. **Web Scraping Background**
2. **Web Scraping Tutorial**
3. **Web Scraping Practice**

What is an API?

- **API = Application Programming Interface**
  - A software intermediary that allows to applications to talk to each other

1. Videos
2. User Info
3. Video Comments

Documented APIs

- Documented APIs = free/paid APIs that organizations want 3rd-parties to use
  - TikTok Research API
  - OpenAI API
  - Jigsaw Toxicity API
  - Google Ads API

Undocumented APIs

- Undocumented APIs = "unsung heroes making sure websites run"
  - Infinite scroll to load more posts → **instaloader**
  - Autocomplete suggestions → **suggests**
  - Obscured labels → **Amazon Brand Detector**

https://inspectelement.org/apis.html#documented-apis

Why Undocumented APIs are Useful

- Example: [Markup's investigation of YouTube's hidden advertiser keyword blocklist](#)
  - Explored keyword search tool for YouTube advertisers and found evidence of a hidden keyword blocklist
  - Found undocumented API that was powering the hidden keyword blocklist
  - Demonstrated pattern of YT blocking racial justice terms, but allowing hate terms

https://themarkup.org/google-the-giant/2021/04/08/google-youtube-hate-videos-ad-keywords-blocklist-failure