fellowship.ai

Demo and Status Report: Autonomous Web Agent

- 21.03.2025

Overview

Recent progress

- Progress with the agents
- Benchmarking

Agent

Since the web scraping agent was not possible due to Proxy-lite not supporting OCR but rather uses JS to interact with the webpage; we have decided to change the way the agent also interacts with the webpage using the following:

The agent now uses AI Model querying of proxy-lite-3b by using Blip2Processor and AutoModelForVision2Seq to process user input.

Google search and navigation via Playwright.

Trying to replicate the real time element tracking by injecting JavaScript to monitor dynamic elements (buttons, inputs) and highlight them.

Agent - Continued

Takes screenshots of webpages to show what the model is interacting with.

Records frames of browser actions and saves them as “output.gif”

Logs everything such as browser action, debugging info, and extracted data.

Saves all of that into a json file called “results.json”

Agent - What's left?

What's left underworks is:

1. Needs form filling logic: only chooses first link, does not detect correct input field.
2. Needs to analyze page structure and decides where to click/type.
3. Needs JS interaction for drop-downs.
4. Needs exception handling for better stability and performance (if a page does not load AI does not retry)
5. It does not improve over time. Reinforcement learning from human interaction can be added.

Benchmarking

As for the benchmarking more tasks have been collected and evaluated the proxy-lite on 50 actions in total; the tasks can be broken down into 20 click, 10 type, 10 scroll, 10 return values). The Recall, Precision, and F-1 score have been used as metrics.

(benchmark.pdf)

Goals for next week

- Wrap up everything
- Document the notebooks