Computer Vision for Browser Use Agents Opportunity Assessment

Browser Use Agents Can Complete Tedious Tasks on Users' Behalf

People collect, collate, and filter information across many websites and web apps to take actions such as deciding on what to buy. Browser Use Agents can complete these tedious tasks. Agents can use existing UIs to complete tasks instead of waiting for agent specific APIs to launch. As long as humans need to interact with software, UIs won't go away.

UI Description Language Accelerates Development of Task Specific Browser Use Agents

Instead of a single "superman" browser use agent that gets "all" web tasks done for everyone, a variety of agents that are optimized for narrow sets of tasks may emerge as the cheapest, fastest and most reliable.

One key challenge that developers face in implementing browser use agents is transforming UIs into meaningful information that agents can act on. Inputting HTML into LLMs is expensive because of HTML's high token count.

A technology that transforms web UIs via computer vision into a structured language, UI Description Language or UDL, can help agent developers overcome their common technical challenge. UDL is lighter than HTML but saves the key qualities of web components such as position on page, size, shape and text that help agents understand web pages, identify targeted information or click paths to acquire targeted information.

This transformer helps developers advance ahead to addressing app specific challenges. Although UIs are made from similar web components and are thus expressible as UDL, UIs of LexisNexis, email clients, spreadsheets, shopping web apps and concert/sports event web apps have different information architecture and usage of the English language.

Existing players include:

Product Category	**Players**	**Description**	**Bets**
"Superman" Browser Use Agent for **Consumer**	1. Open AI's Operator 2. Google's Gemini App Agent Mode/Project Mariner	Prompt driven input that gets everything done for everyone.	Bets that single general model will be the most cost effective, quick, and reliable.
"Superman" Browser Use Agent for **Enterprise**	1. Narada AI	Prompt driven input that gets everything done for everyone.	Bets that single general model will be the most cost effective, quick, and reliable.
Tools for Browser Use Agent	1. Tinyfish AI's AgentQL/DocQL	Language that AI Agents can write/run to extract information from web pages and documents.	Bets that AI Agents developers will train their agents to write AgentQL/DocQL.
Transformer of Web UI into Structured Data, the Opportunity	1. Human-Eye AI	Transforms Web UIs into language that ML boxes and deterministic programs understand via computer vision.	Bets that different AI Agents will be task specific because mastering a particular task requires different resources and training methods.

Human-Eye AI's computer vision technology transforms web UIs into UDL. This technology enables agent developers to avoid the high costs of inputting HTML into LLMs or having to develop their own computer vision solution. They can concentrate on getting data and developing training methods that differentiate their agents.