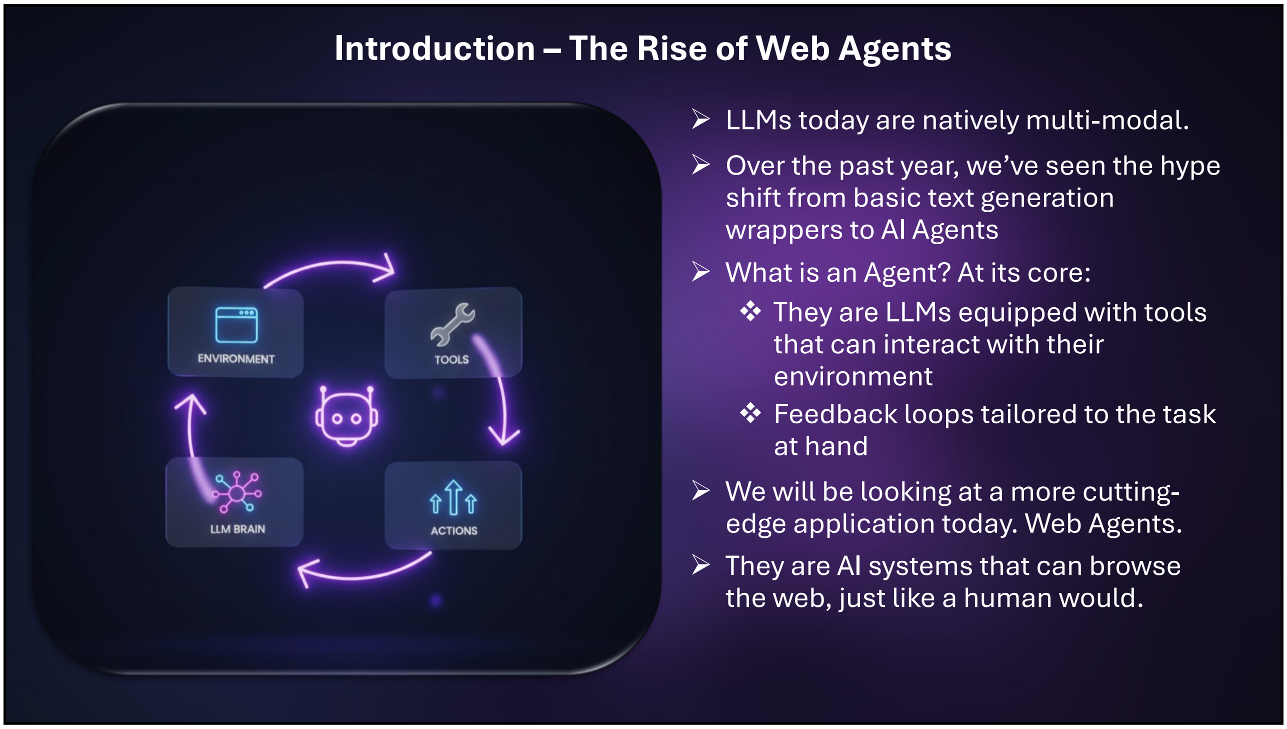
# Modern Web Agents: A Comparative Analysis

- Agnay Srivastava, Edward Kang

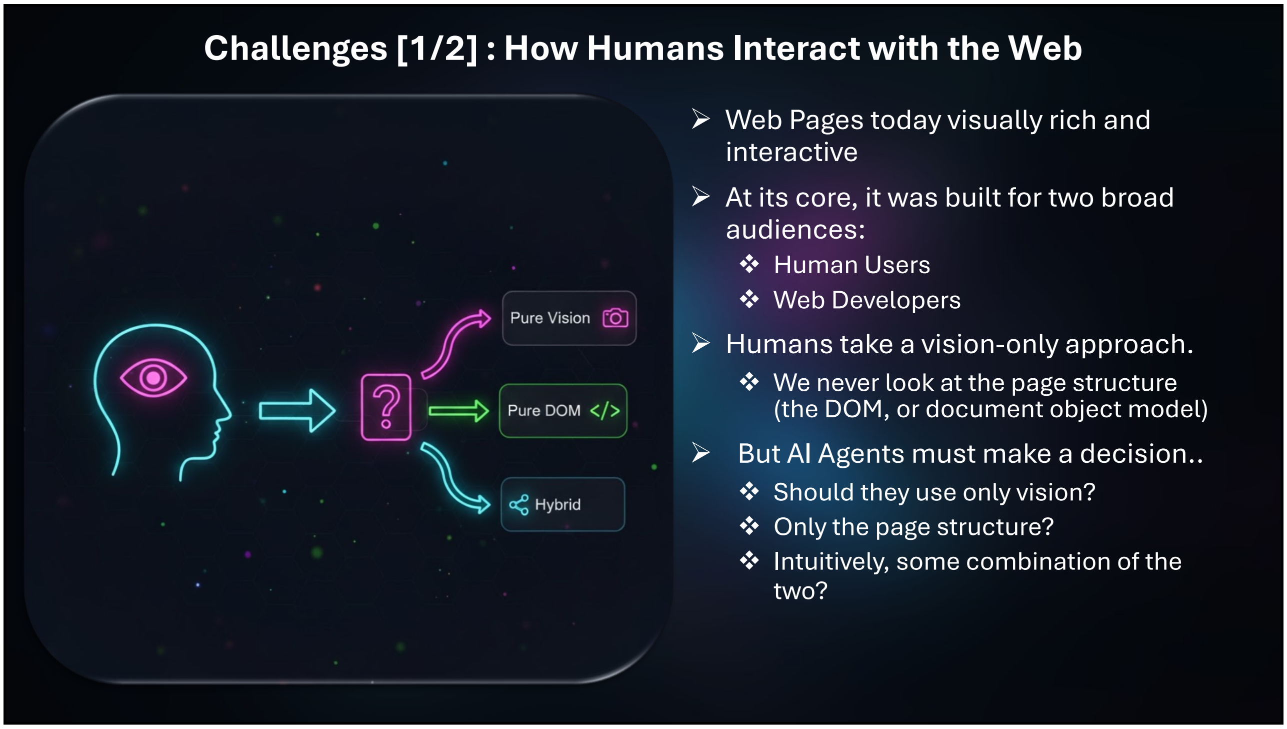
1



## Introduction – The Rise of Web Agents

- LLMs today are natively multi-modal.
- Over the past year, we've seen the hype shift from basic text generation wrappers to AI Agents
- What is an Agent? At its core:
  - They are LLMs equipped with tools that can interact with their environment
  - Feedback loops tailored to the task at hand
- We will be looking at a more cutting-edge application today. Web Agents.
- They are AI systems that can browse the web, just like a human would.

2

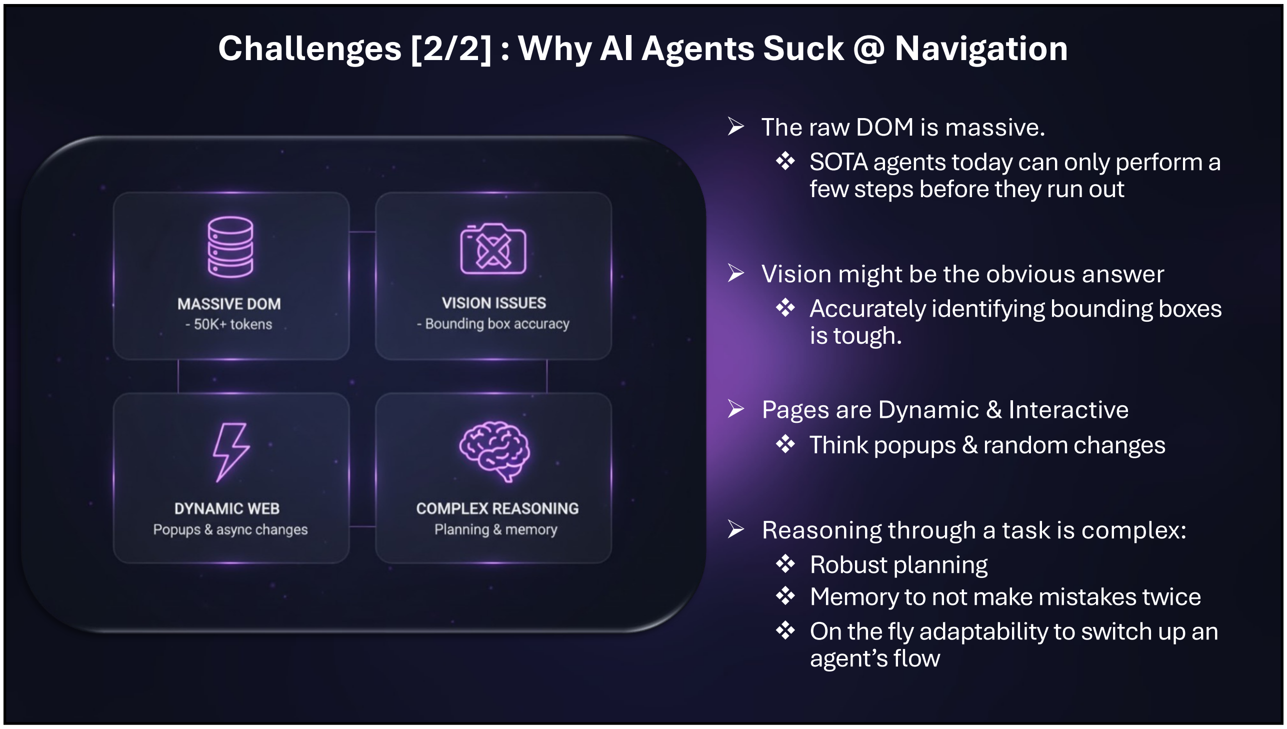


## Challenges [1/2] : How Humans Interact with the Web

Pure Vision
Pure DOM </>
Hybrid

- Web Pages today visually rich and interactive
- At its core, it was built for two broad audiences:
  - Human Users
  - Web Developers
- Humans take a vision-only approach.
  - We never look at the page structure (the DOM, or document object model)
- But AI Agents must make a decision..
  - Should they use only vision?
  - Only the page structure?
  - Intuitively, some combination of the two?

3



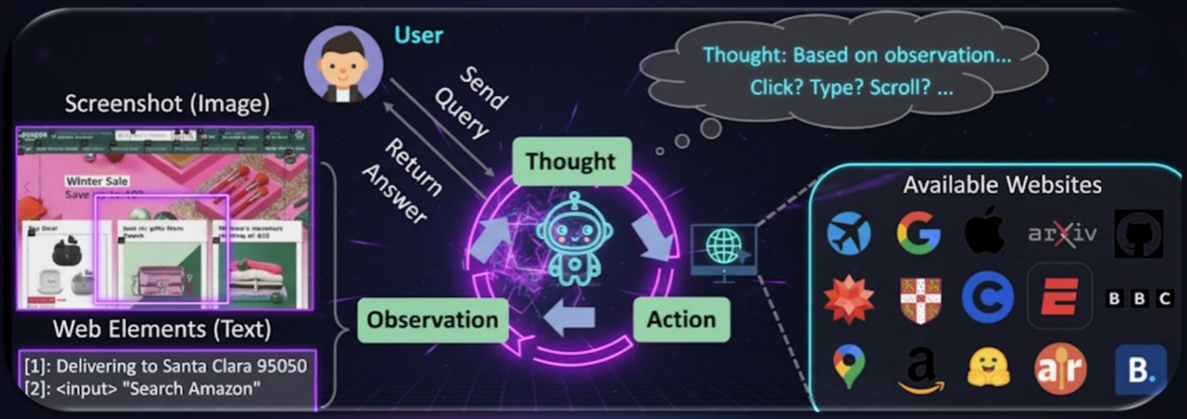
## Challenges [2/2] : Why AI Agents Suck @ Navigation

MASSIVE DOM
- 50K+ tokens
VISION ISSUES
- Bounding box accuracy
DYNAMIC WEB
Popups & async changes
COMPLEX REASONING
Planning & memory

- The raw DOM is massive.
  - SOTA agents today can only perform a few steps before they run out
- Vision might be the obvious answer
  - Accurately identifying bounding boxes is tough.
- Pages are Dynamic & Interactive
  - Think popups & random changes
- Reasoning through a task is complex:
  - Robust planning
  - Memory to not make mistakes twice
  - On the fly adaptability to switch up an agent’s flow

4

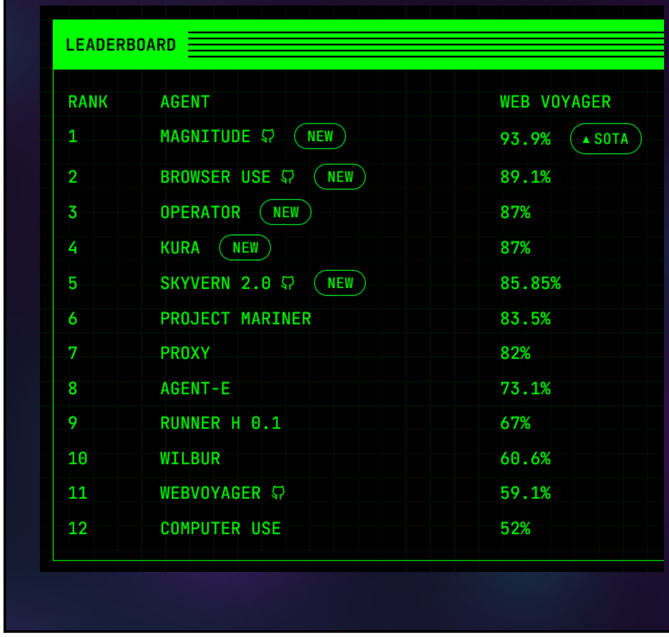
## Web Voyager - The Benchmark Used



- A multimodal evaluation suite (643 tasks across 15 popular sites)
- Automatically evaluates open-ended web task completion using screenshots & page element info.

5

## Our Approach

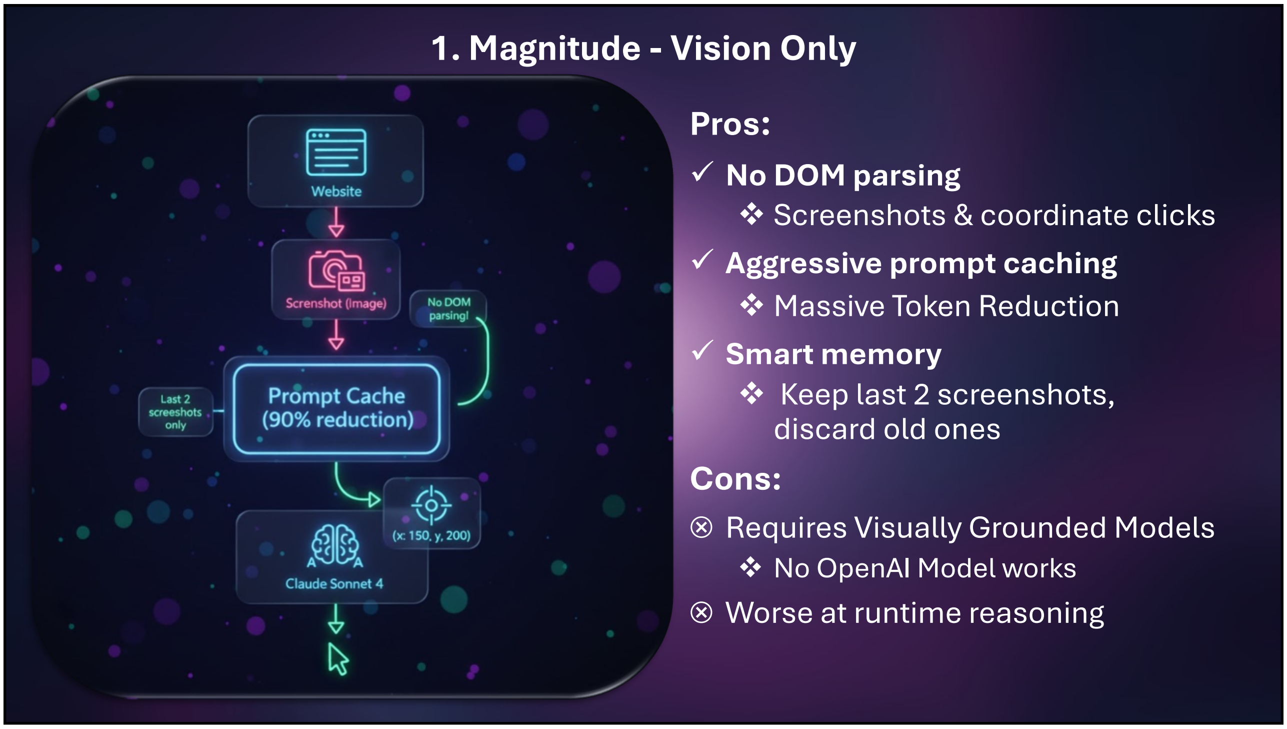


LEADERBOARD

| RANK | AGENT | WEB VOYAGER |
|---|---|---|
| 1 | MAGNITUDE NEW | 93.9% ▲ SOTA |
| 2 | BROWSER USE NEW | 89.1% |
| 3 | OPERATOR NEW | 87% |
| 4 | KURA NEW | 87% |
| 5 | SKYVERN 2.0 NEW | 85.85% |
| 6 | PROJECT MARINER | 83.5% |
| 7 | PROXY | 82% |
| 8 | AGENT-E | 73.1% |
| 9 | RUNNER H 0.1 | 67% |
| 10 | WILBUR | 60.6% |
| 11 | WEBVOYAGER | 59.1% |
| 12 | COMPUTER USE | 52% |

- ➢ Deep Dive into the top open-source web agents.
- ➢ Analyze architectures & winning patterns
- ➢ Compared trade-offs (cost, speed, accuracy)
- ➢ The 2 we are looking at are:
  - ❖ Magnitude:
    - ▪ Vision only
  - ❖ Browser Use:
    - ▪ Hybrid, but more DOM heavy

6



## 1. Magnitude - Vision Only

**Pros:**

✓ **No DOM parsing**

❖ Screenshots & coordinate clicks

✓ **Aggressive prompt caching**

❖ Massive Token Reduction

✓ **Smart memory**

❖ Keep last 2 screenshots, discard old ones

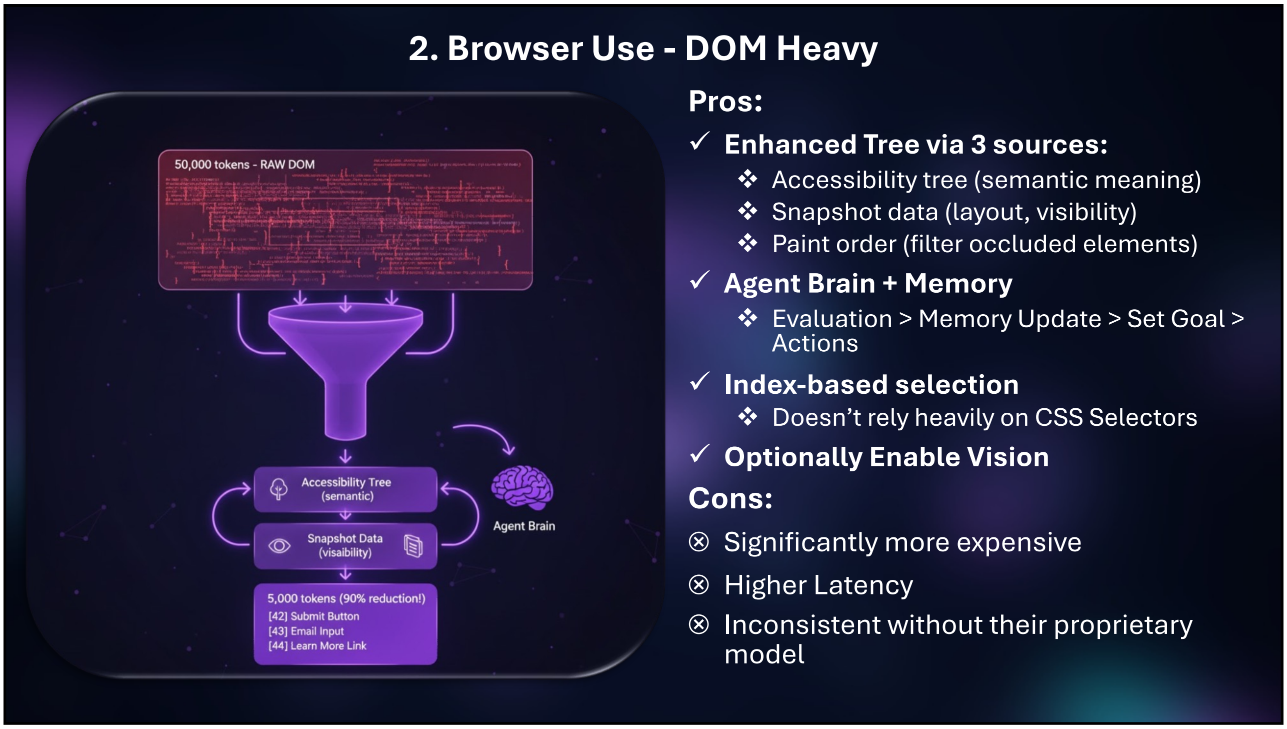
**Cons:**

⊗ Requires Visually Grounded Models

❖ No OpenAI Model works

⊗ Worse at runtime reasoning

7



## 2. Browser Use - DOM Heavy

**Pros:**

✓ **Enhanced Tree via 3 sources:**

❖ Accessibility tree (semantic meaning)

❖ Snapshot data (layout, visibility)

❖ Paint order (filter occluded elements)

✓ **Agent Brain + Memory**

❖ Evaluation > Memory Update > Set Goal > Actions

✓ **Index-based selection**

❖ Doesn’t rely heavily on CSS Selectors

✓ **Optionally Enable Vision**

**Cons:**

⊗ Significantly more expensive

⊗ Higher Latency

⊗ Inconsistent without their proprietary model

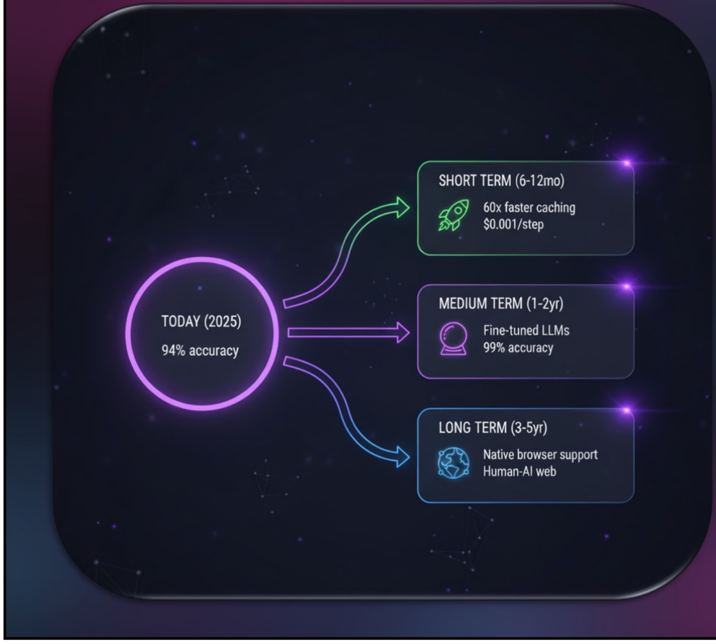
8

## Ethical Considerations

- **Terms of Service Violations**
  - Robots.txt
  - Rate limiting
  - CAPTCHAs exist for a reason
- **Security & Abuse**
  - Credential harvesting
  - Scalping bots
- **AI Agents make everything easier**
  - Any problems that existed with scraping before are now made worse
- **Our Take:**
  - The Technology Isn't Evil
  - The Use Case Determines Morality

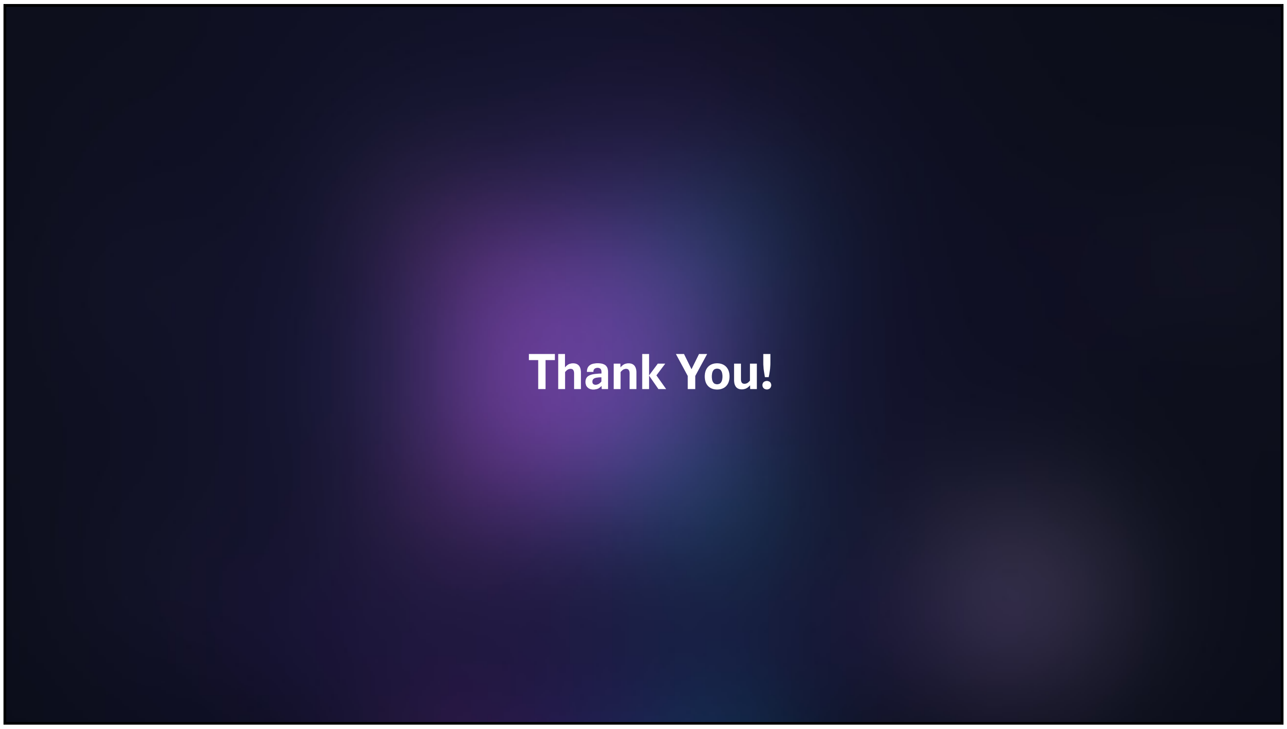
9

## Future Work



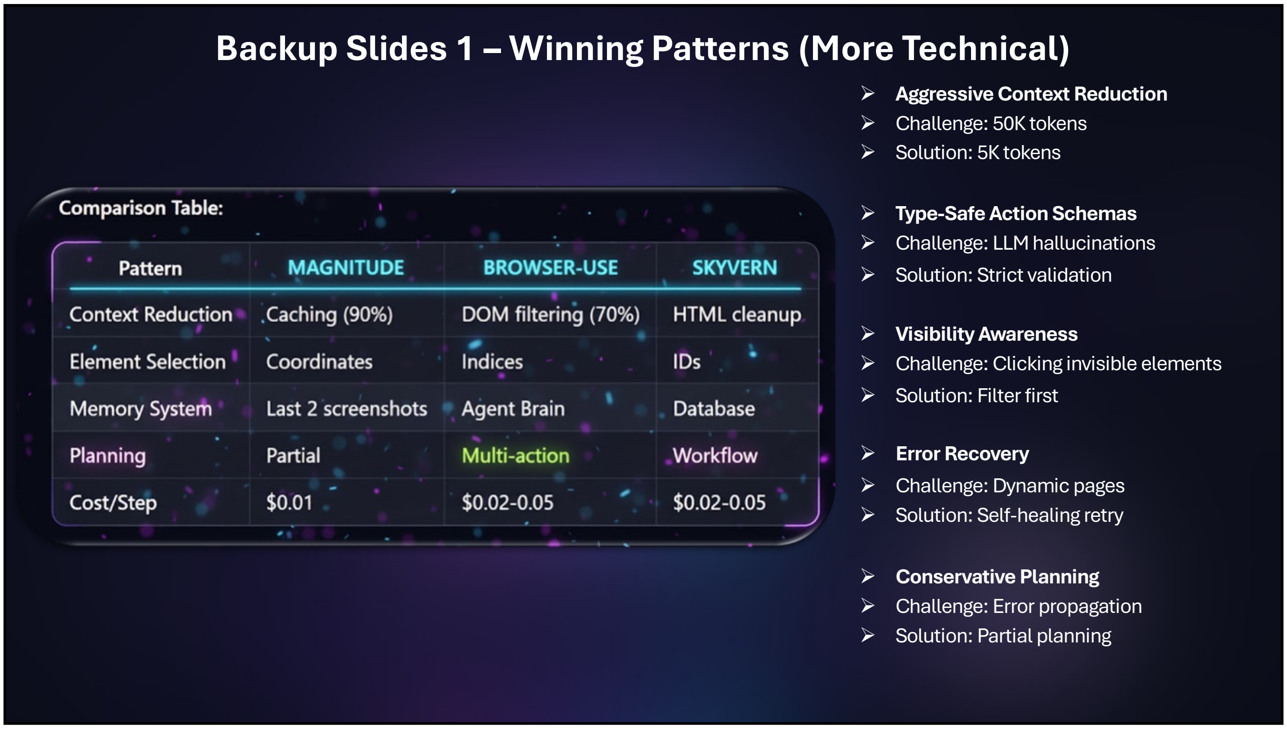
- Models have plateaued
  - Still becoming cheaper & faster
- Short Term:
  - Step Caching
  - Imitation + Transfer Learning
- Medium Term:
  - Agent API Frameworks that get baked into websites for better AI SEO
  - Fine-Tuned LLMs on web task flows
- Long Term:
  - Browser Vendors add native agent support
- Pure vision techniques can also be applied to OS level tasks
  - Claude CUA
  - Gemini CUA

10



# Thank You!

11



## Backup Slides 1 – Winning Patterns (More Technical)

**Comparison Table:**

| Pattern | MAGNITUDE | BROWSER-USE | SKYVERN |
|---|---|---|---|
| Context Reduction | Caching (90%) | DOM filtering (70%) | HTML cleanup |
| Element Selection | Coordinates | Indices | IDs |
| Memory System | Last 2 screenshots | Agent Brain | Database |
| Planning | Partial | Multi-action | Workflow |
| Cost/Step | $0.01 | $0.02-0.05 | $0.02-0.05 |

- **Aggressive Context Reduction**
- Challenge: 50K tokens
- Solution: 5K tokens

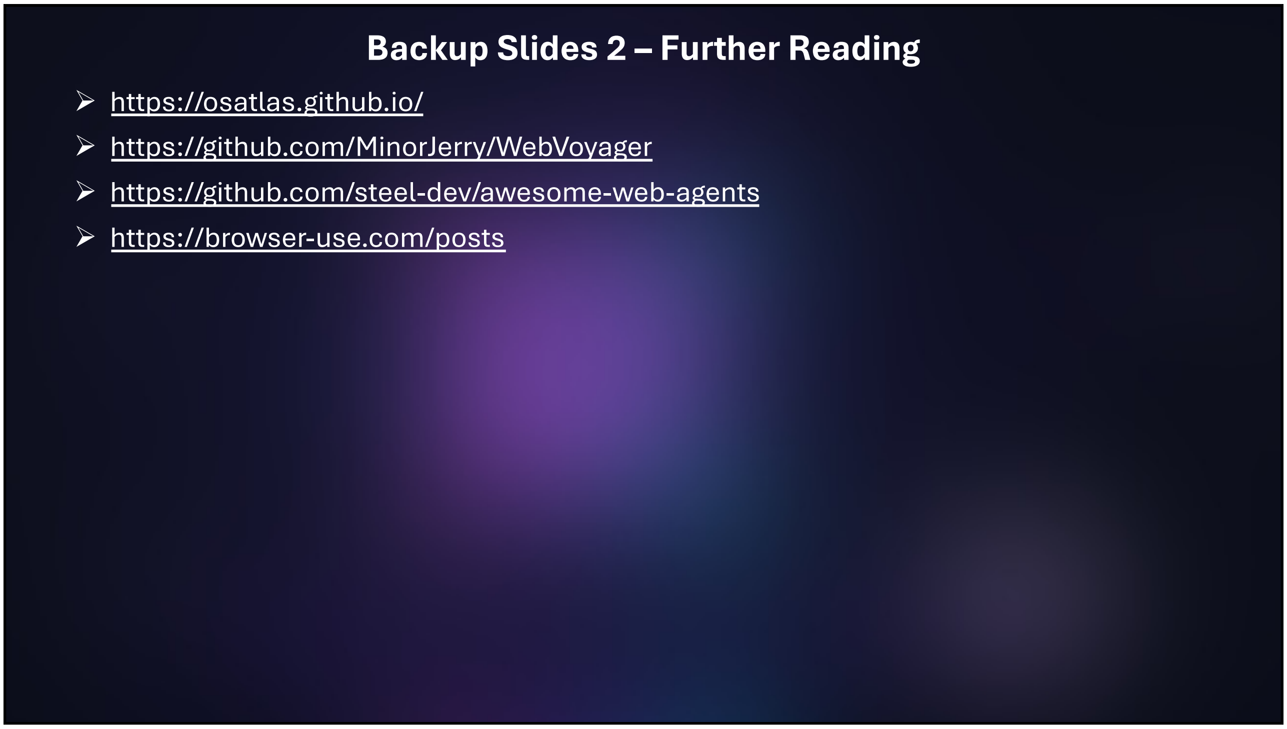
- **Type-Safe Action Schemas**
- Challenge: LLM hallucinations
- Solution: Strict validation

- **Visibility Awareness**
- Challenge: Clicking invisible elements
- Solution: Filter first

- **Error Recovery**
- Challenge: Dynamic pages
- Solution: Self-healing retry

- **Conservative Planning**
- Challenge: Error propagation
- Solution: Partial planning

12



## Backup Slides 2 – Further Reading

- https://osatlas.github.io/
- https://github.com/MinorJerry/WebVoyager
- https://github.com/steel-dev/awesome-web-agents
- https://browser-use.com/posts

13