

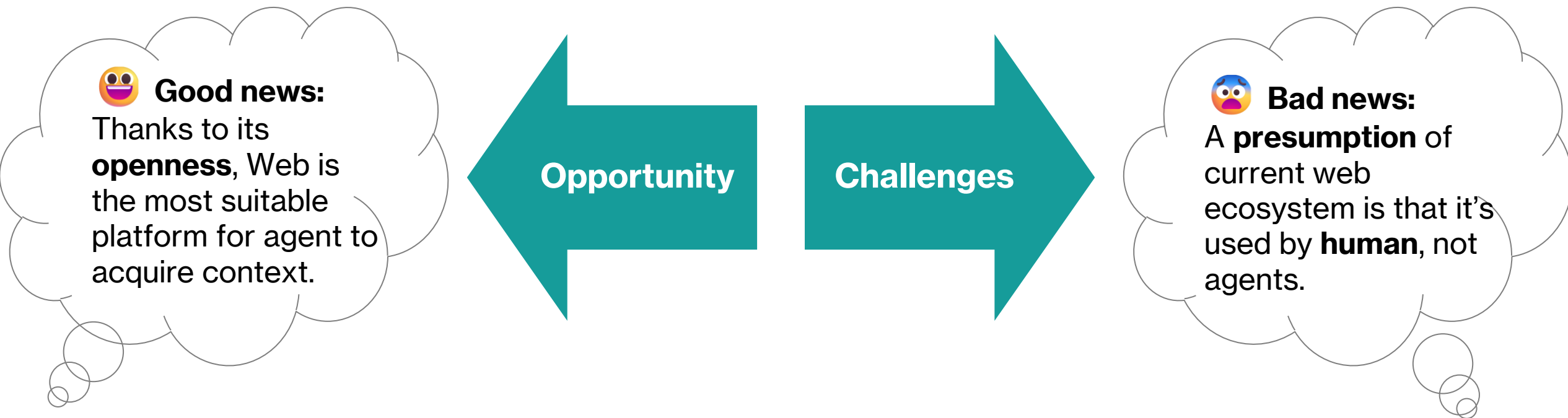
Browser for for Agent to Access Web Context

Lingyan Zhao

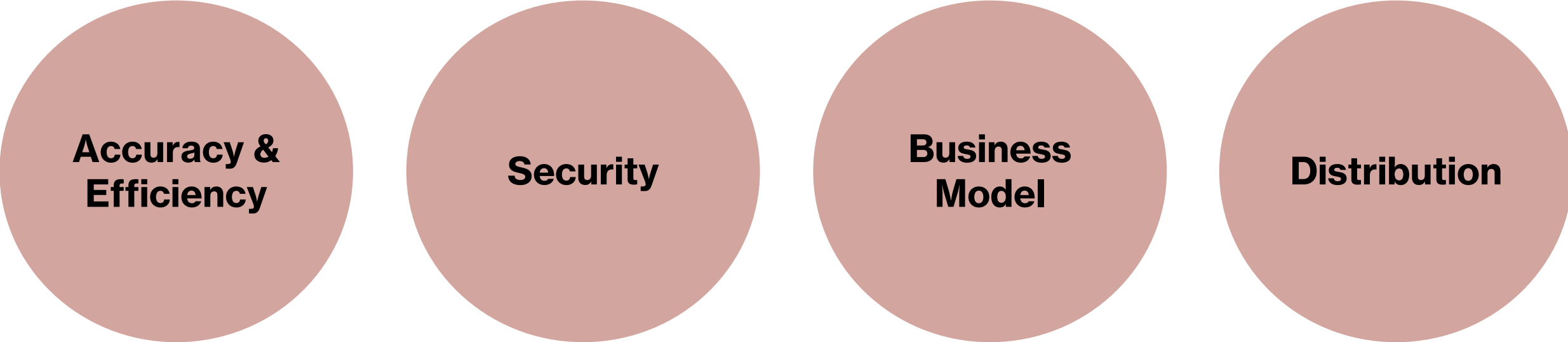
Microsoft Edge Web Platform

2025 September

Web Context, from an AI Agent's POV



Core Challenges

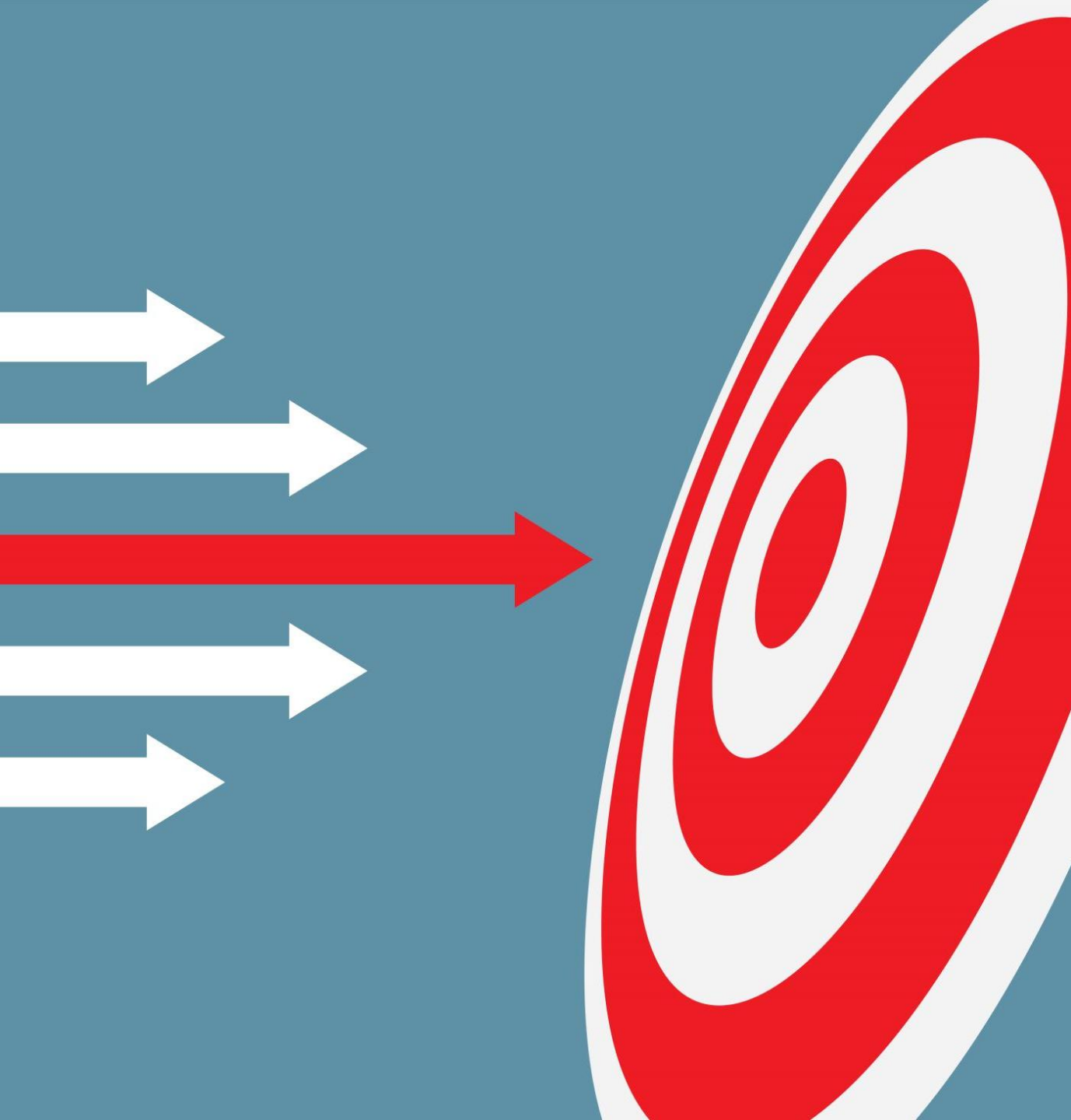


**Accuracy &
Efficiency**

Security

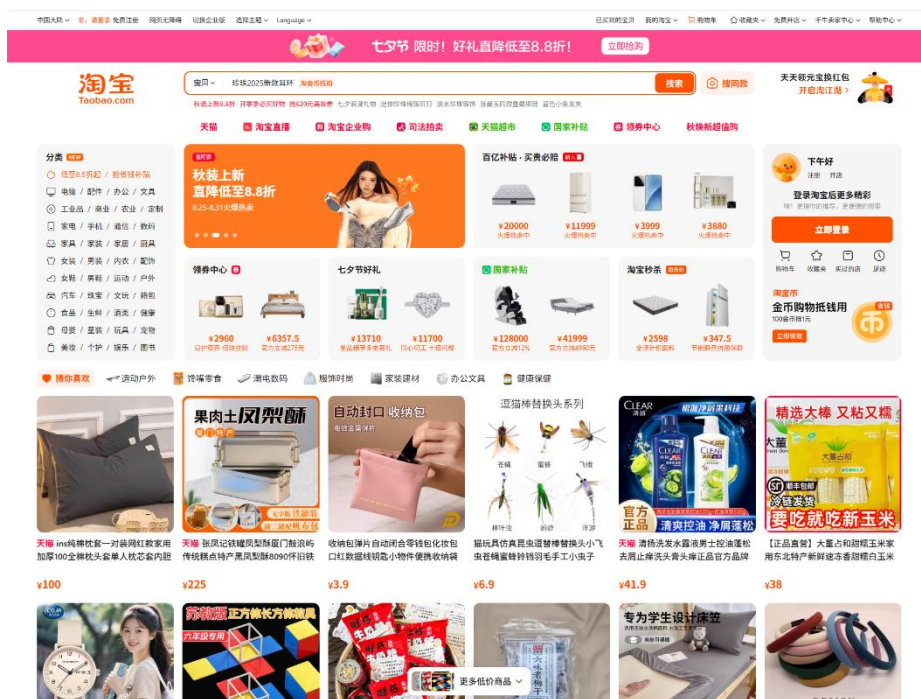
**Business
Model**

Distribution



Problem 1: Accuracy & Efficiency

Limitation of VLM (Vision LLM) or Playwright



Why VLM/Playwright should not be relied as a primary path?

- UI/DOM \neq Semantic
- Low throughput, high token cost \$\$\$
- Non-deterministic flows

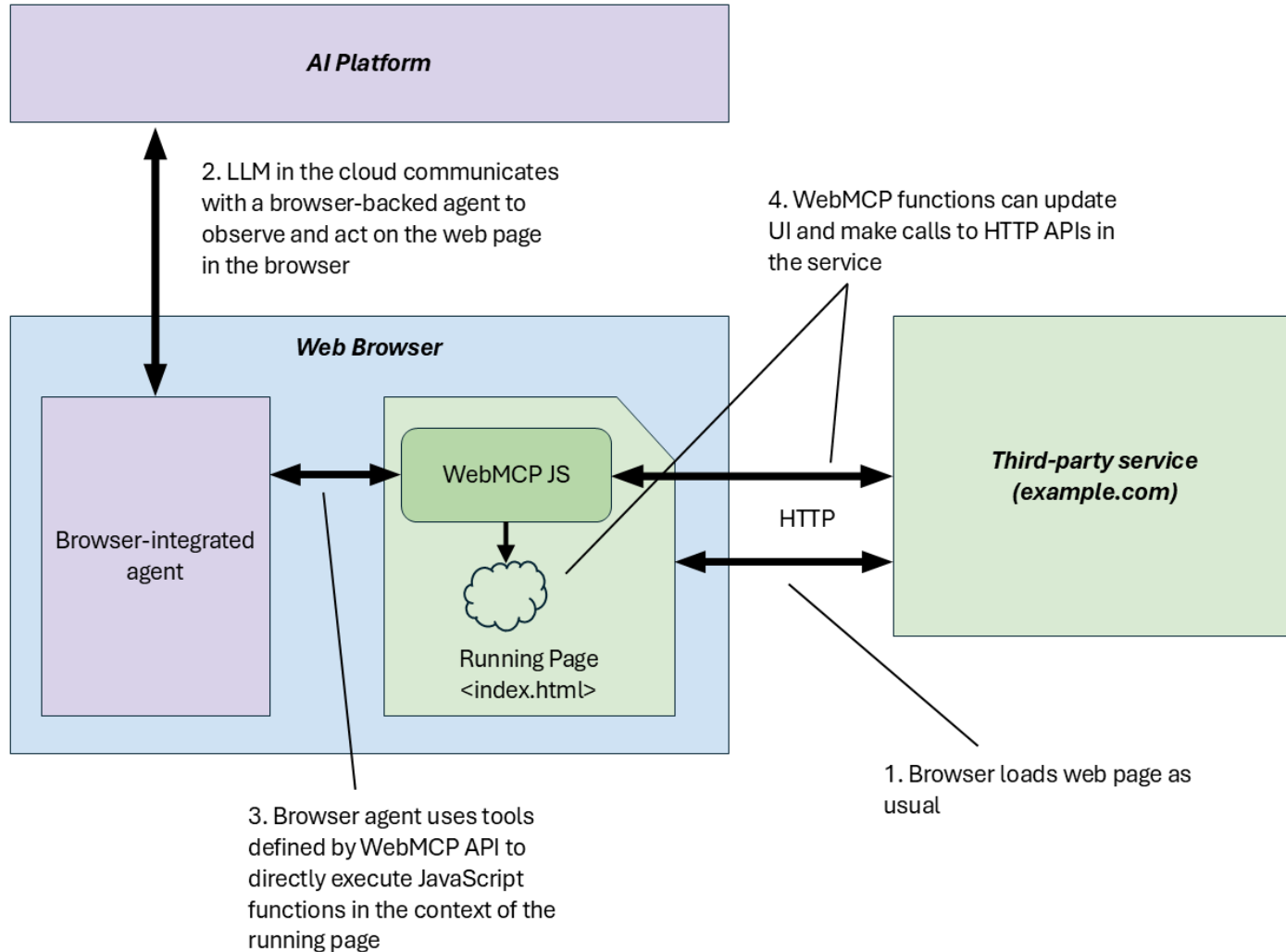
VLM/Playwright approach could serve as a fallback solution, but AI needs AI-native approach in the end.

Solution: atomic functions and context for agent in web ecosystem

 MCP is a fundamental protocol and 'http for agents'. On top of MCP, several solutions are introduced for web ecosystem.

Name	Introduced by	Brief description	Repo/Doc
NLWeb	Microsoft in 2025	'HTML for the agents'	nlweb-ai/NLWeb
WebMCP	Microsoft Edge + Google in 2025	A Web API for web apps to register MCP tools	webmachinelearning/webmcp
MCP and App Action on Windows	App Action GA in 2025, MCP framework demoed in 2025 Build	PWAs could register app actions and Windows will turn the PWA into an MCP server	Enable App Actions on Windows for a PWA

WebMCP



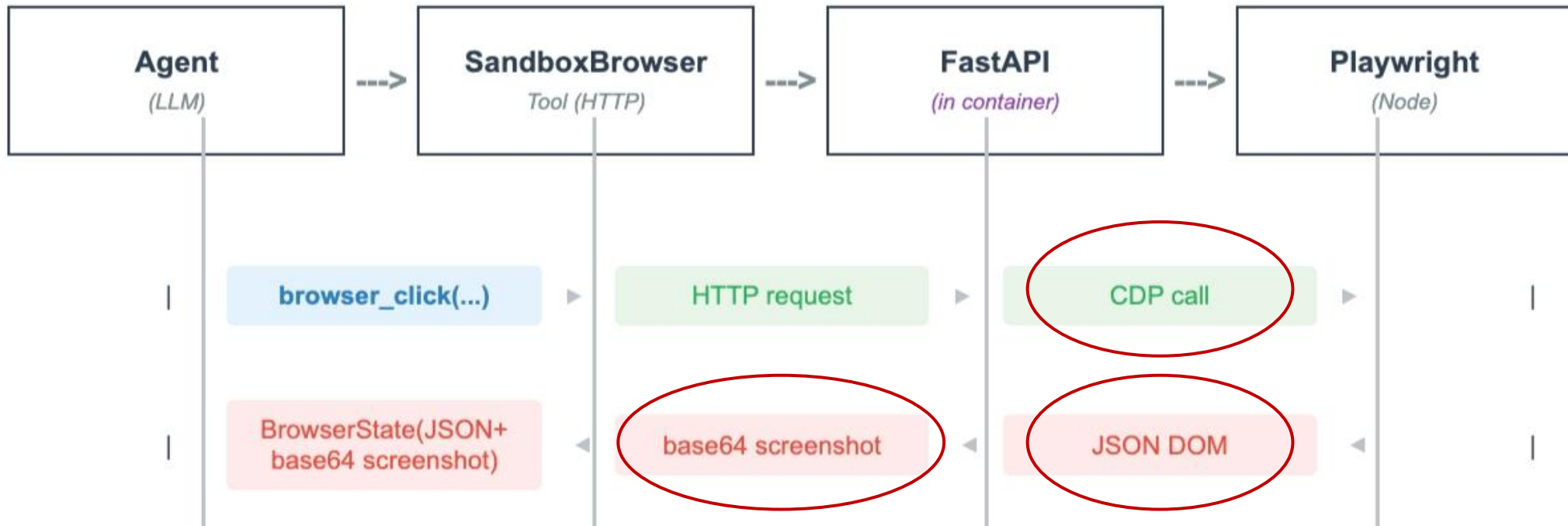
[WebMCP proposal's link](#)

WebMCP enables web pages to provide agent-specific paths in their UI

- Suitable for human-in-the-loop workflows
- WebMCP allows web developer to leverage existing business logic and UI

Fallback solution: native browser support to reduce overhead?

An example flow from one of current open-source agent:



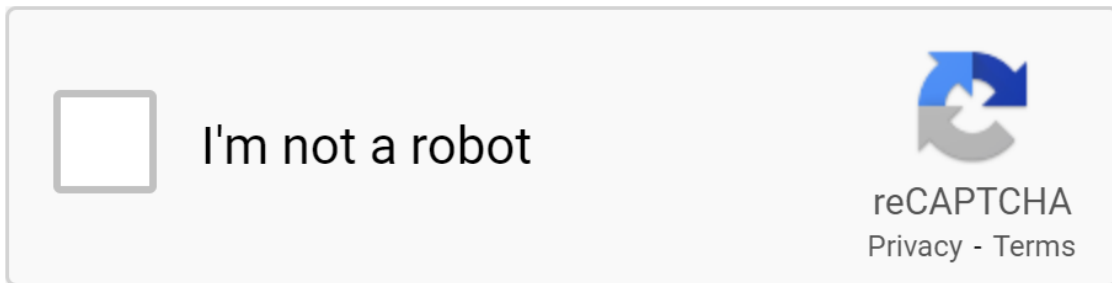
Optimization lies in native browser support (binary buffers, native support for synthetic input, DOM diff) to reduce overhead & improve robustness.



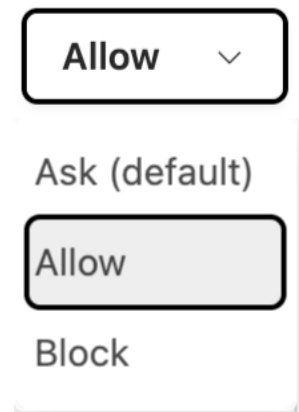
Problem 2: Security

CAPTCHA, Bearer Token, Permissions

- CAPTCHA force “prove human” loops
 - CAPTCHA = Completely Automated Public Turing test to tell **Computers** and **Humans** Apart
- OAuth’s bearer tokens are inherently transferable
 - HTTPS only protects them from interception during transmission
- Weak least-privilege
- Lack of auditing/revocation mechanism



 Microphone



From prove human to prove capability?

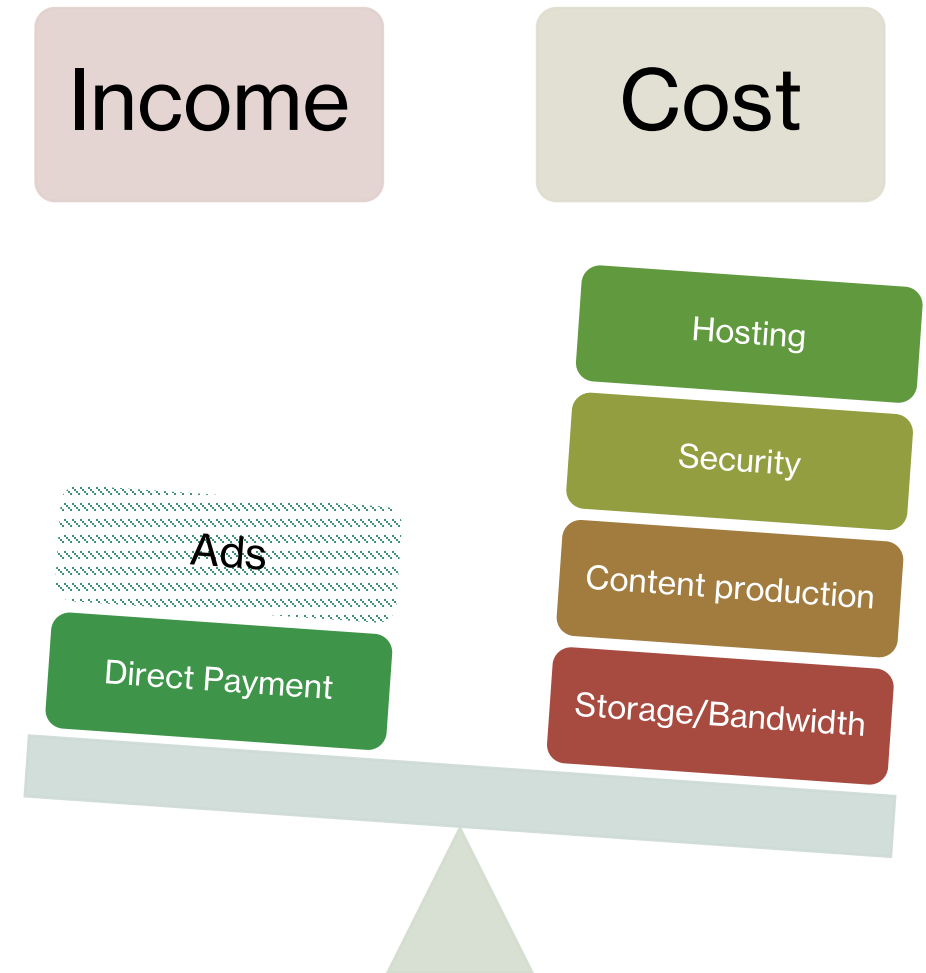
- **Beyond CAPTCHAs**
 - tell legitimate agents and suspicious agents apart
- **Capability based token with task-scoped grants**
 - “This agent can access resource A for 5 mins in order to do task X”
- **Verifiable Delegation**
 - who/what/when/why
- **Audit chain**
 - Actor -> action -> outcome



Problem 3: Business Model

Attention \neq Currency for Agents

“The original web was the human web, and advertising was and is one of the best possible ways to monetize the only **scarce** resource in digital: **human attention**. ”



Pay to access web resources?

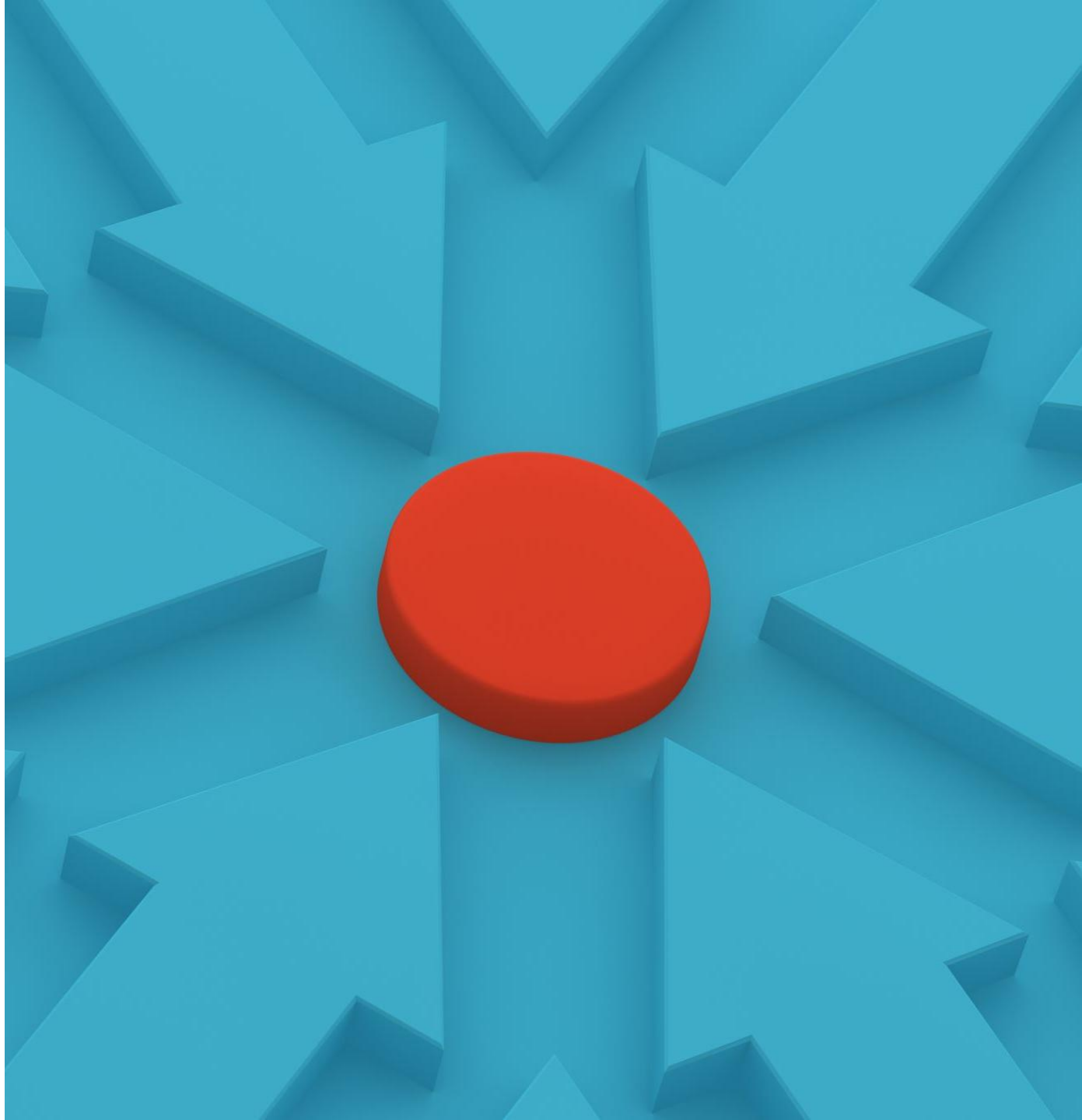
- Cloudflare's [Pay-Per-Crawl](#)
- Value shift from **human views** to metered access or verifiable **outcomes**
- Web browser:
 - Handle HTTP 402 with payment procedure
 - Agent profiles: identity, budget, terms; respect robots/agent headers



Introducing pay per crawl: Enabling content owners to charge AI crawlers for access

2025-07-01

Problem 4: Distribution



How could an agent find the best tool for a task?



Code is cheap -> Tons of overlapping providers



No standardized mechanism to discover tools



Absence of a rating/ranking system

Web browser's new value-add?

- **Discovery UX**
 - .well-known path, web manifests
- **Privacy-preserving usage signals for ranking**
 - Objective signals: task completion count, success rate, latency
 - Subjective signals: user preference
- More ideas...

The background of the slide features a complex network of white lines connecting various circular nodes of different sizes. These nodes are scattered across a dark blue to green gradient. Interspersed among the network are strings of white binary code (0s and 1s), some appearing as small clusters and others as longer sequences, creating a digital, data-driven aesthetic.

Make the Web Agent-Native

Shift from UI to Structured Actions

Replacing traditional UI interactions with structured actions enables precise and efficient agent control on the web.

Scoped Authentication with Key-Bound Capabilities

Improves security and fine-grained access control in new measures

Collaborative Web and Browser Ecosystem

Websites and browsers should expose actions and support scoped authentication while maintaining fallback mechanisms.

Outcome-Based Incentives Over Attention Economy

Replacing attention-based economics with outcome-based incentives fosters a more reliable and efficient environment.



Thanks for listening!