

# AI-Powered Web Browsing Experience Reconstruction

Liu YaoMing, Huawei



W3C AC meeting 2026  
Hangzhou, China  
20-22 April 2026





# Lists

## Challenges

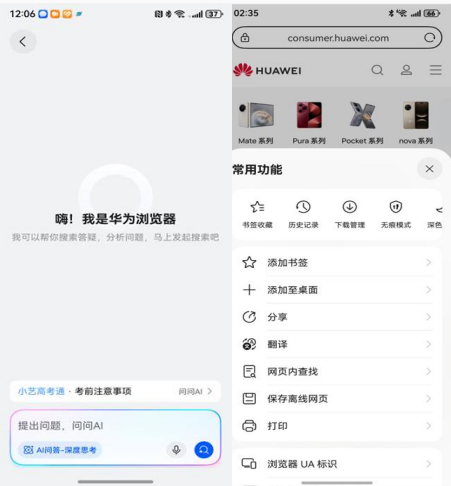
## Implementations & Explorations

- Web Copilot
- Web Agent
- Agentic Web

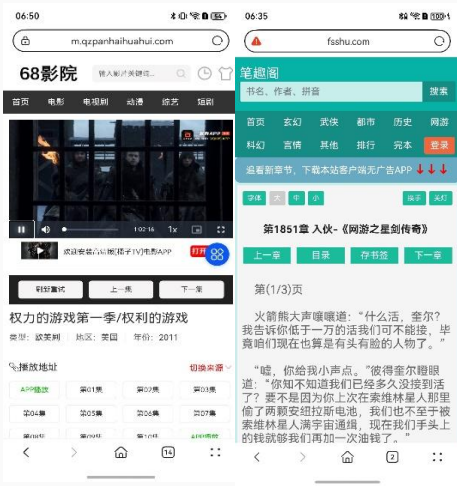


# Challenge – Browser's Top Voice of Customer

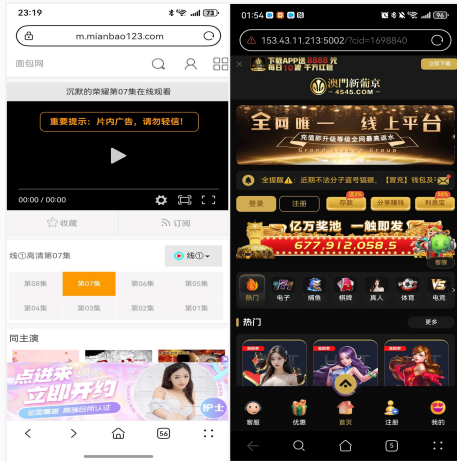
## AI Search, More Tools



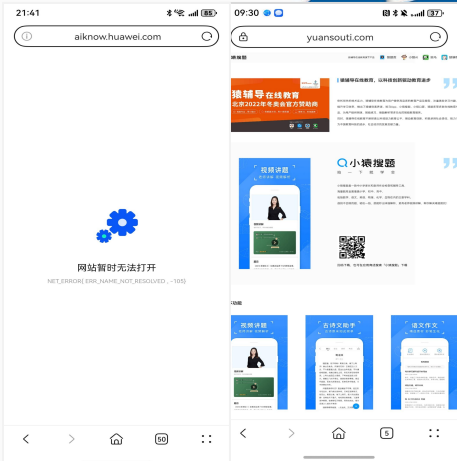
## suboptimal reading & media experience



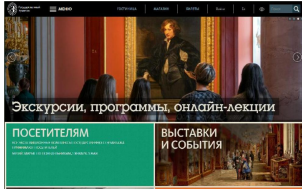
## Too Many Ads



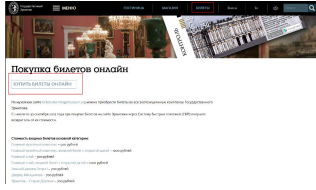
## Load Failures/Layout Issues



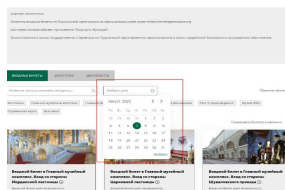
# Challenge – Web Service Experience Pain Points: Service Fragmentation and Process Complexity



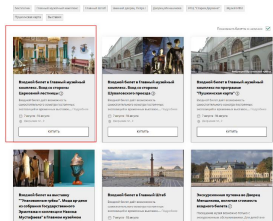
1. the Winter Palace homepage



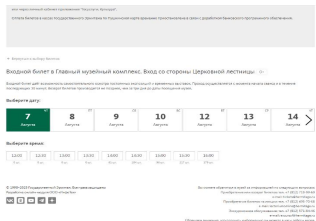
2. ticket purchase page



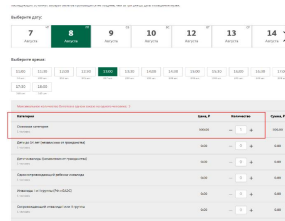
3. Select visit date



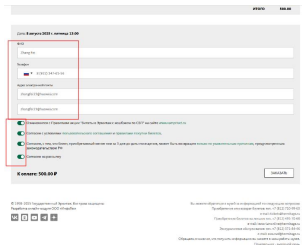
4. Select ticket type and entrance



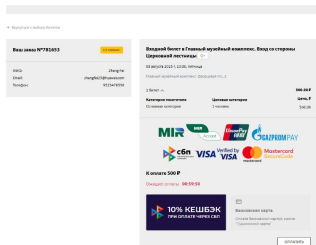
5. Scroll to next option



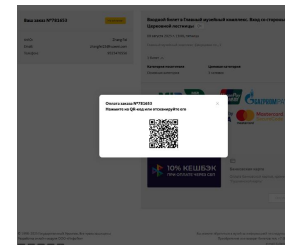
6. Select time slot and visitor category



7. Fill in visitor information



8. Select payment method



9. Scan to pay

## ● Extreme Service Fragmentation

- ✓ **Lack of "Super Apps"**: The Web is the primary interface for practical life tasks: shopping, dining, travel, and medical health. Digital infrastructure is scattered across millions of independent web portals.
- ✓ **The "Web Island" Effect**: Sovereignty and language barriers create thousands of mutually exclusive service silos.

## ● Service Processes complexity

- ✓ **Cross-Island Retrieval**: Users struggle with inefficient information gathering across disconnected domains.
- ✓ **Decision Mazes**: Completing a single task requires navigating massive, complex branching logic within and between websites.

# Web Evolution: "information network" to "action network"



## Web AI

### Web Copilot

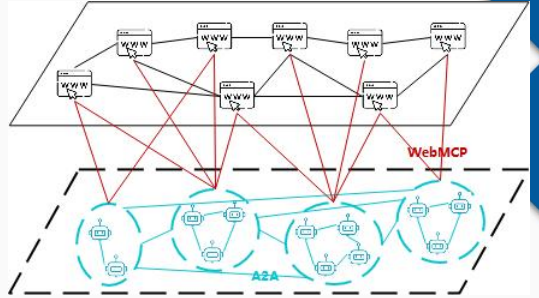
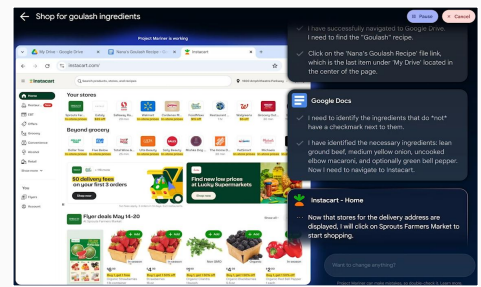
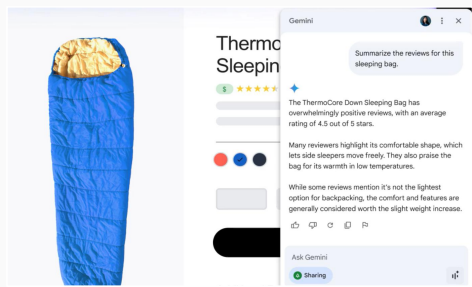
Elevating Browsing Experiences with AI. AI mode Search, Summarize, Rewrite, Translation, Reading mode, etc.

### Web Agent

Intent-driven, autonomously perceives web semantics and executes complex operational sequences to fulfill tasks

### Agentic Web

New Paradigm: web content and architectural design are specifically optimized for "AI Agent operations"





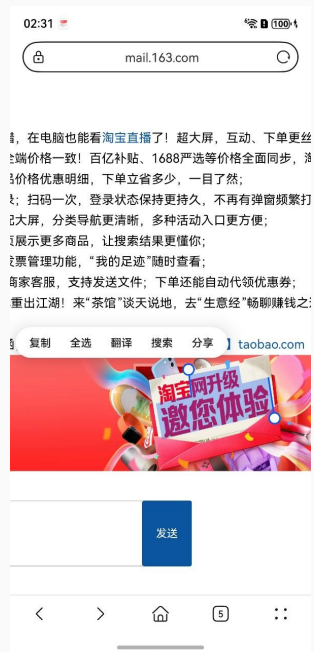
# Web Copilot: AI in web engine makes web browsing more intelligent, more immersive

6

## text selection



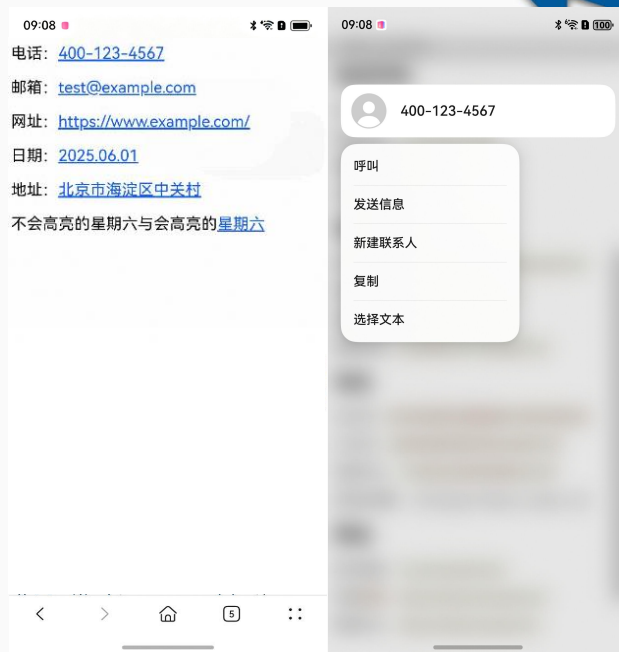
## in-image text recognition



## ReWrite/Summary...



## Special text identification





# Web Copilot: AI in web engine makes web browsing more intelligent, more immersive

## Immersive Translate

Continuous translation.  
In-Image text translation.  
Cross-domain iframe translation.



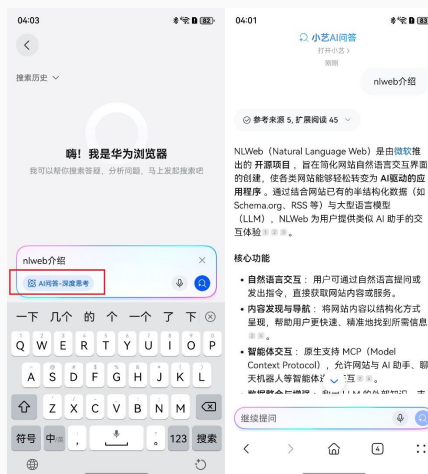
## Reader mode

Intelligent identification of novel-type websites.  
Intelligent recognition of characteristic elements.  
Heuristic analysis of main content.



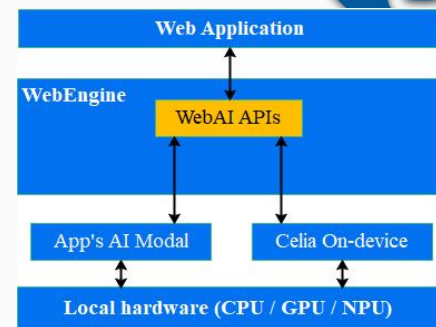
## AI Search

AI mode search based on Celia



## Web AI APIs

Empowering front-end developers with native AI capabilities.  
API: Translation, Write, ReWrite, Language Detection, Summarization, Prompt, Proofreader.



ACC 2024

# Web Agent: Challenges

**Challenge 1: Benchmark**  
**Domain-specific benchmarks** (e.g., for restaurants, hotels, food delivery) capable of evaluating whether performance meets **commercial-grade requirements**.

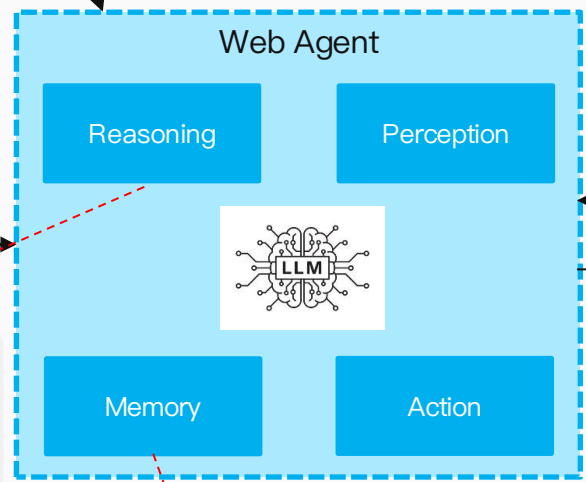


A2UI/LUI

**Challenge 4: Complex Task Planning**  
 Executing a task such as "find and book the cheapest hotel" requires decomposing the objective into 50+ sub-tasks, involving cross-website queries, multi-step price comparisons, and complex decision-making.



Evaluator



**Challenge 5: LLM Context Inflation**  
 Dense web page content.  
 High number of sub-tasks.

**Challenge 2: Perception and Understanding**  
**Cross-modal understanding** of elements such as images, videos, and canvases. Handling **dynamic elements** like pop-ups and lazy-loading components.

web semantics



Web Environment

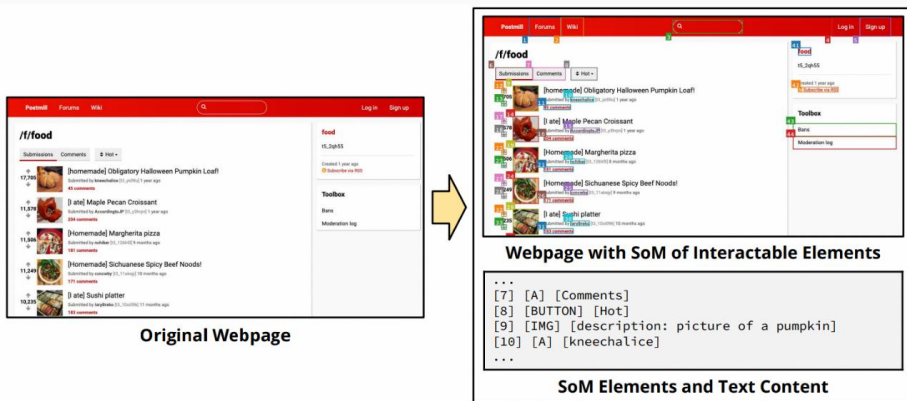
web operation

**Challenge 3: Operation performance**  
 Each "Thought + Action + Observation" cycle requires an LLM call. The overall **execution speed** cannot yet compete with human efficiency. The **success rate** should meet commercial standards

# Web Agent: Accurately perceive web semantics

## Screenshot + Set-Of-Mark

Automated tools pre-identify potential **interactive elements** on web screenshot and overlay them with prominent visual labels. By deeply binding **visual coordinates with semantic metadata**, this technique empowers LLM with the ability to precisely locate and manipulate interface elements.



OpenHarmony ArkWeb:

- get interactive elements
- generate marks for elements

## Annotated Page Content

A hierarchical representation encompassing **text, geometric coordinates, visual styles, interactive attributes, and semantic roles**. This provides the AI with high-fidelity, structured, and action-oriented content input.

**A paradigm shift from 'passive rendering' to 'proactive comprehension'.**

```

[
  {
    "nodeId": 42,
    "elementId": "m-1",
    "role": "button",
    "name": "加入购物车",
    "description": "Add to shopping cart",
    "attributes": {
      "class": "btn-primary",
      "type": "submit"
    },
    "rect": {
      "x": 450,
      "y": 820,
      "width": 120,
      "height": 40
    },
    "isVisible": true,
    "isClickable": true,
    "markLabel": "1"
  },
  {
    "nodeId": 85,

```

OpenHarmony ArkWeb:

- get annotated info for specific elements



# Web Agent: Execute Web Operations Efficiently

## Experience-driven Agent: Learning from Demonstration

### • Demonstration Collection

- ✓ Record User/Agent trajectories using recorder based on Playwright
- ✓ Abstraction & Generalization (semantic encoding)
- ✓ Convert into RPA (Robotic Process Automation) Scripts by Web Agent

### • Online Execute & Evolve

- ✓ When the Agent receives a new task, it performs a pattern match; if a match is found, **it "replays" the RPA Script directly to achieve one-step task completion.**
- ✓ The result of RPA Script execution will be used to update Agent memory and demonstration collection.

## Fallback: "Thought + Action + Observation" cycle

Generate Chrome DevTools Protocol (CDP) commands in real-time to execute webpage operations step by step.

OpenHarmony ArkWeb:

- Execute web operation through CDP commands: navigation, click, scroll, type, send\_keys, select, etc.



# Agentic Web: Building Web MCP Tools for legacy websites through autonomous learning

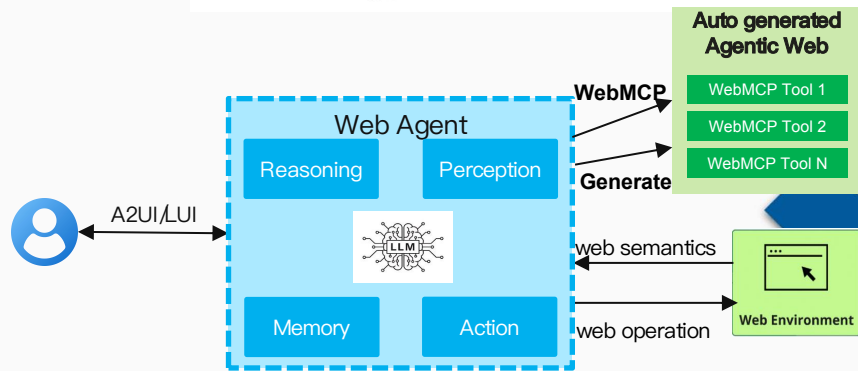
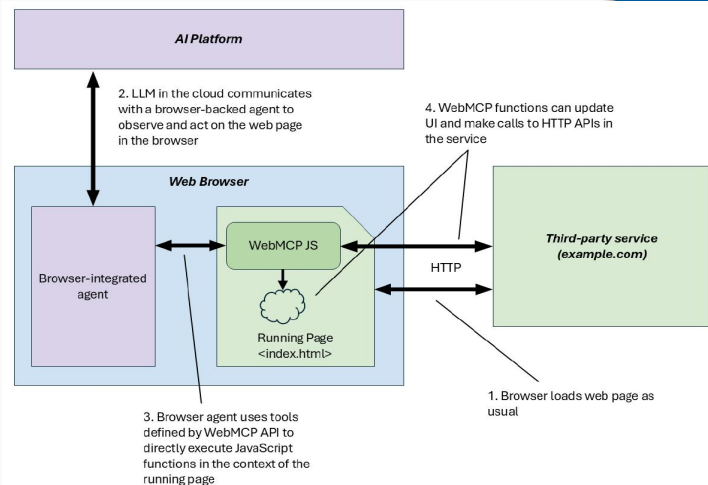
- Agentic Web: **a long-term journey** that requires the entire web ecosystem to adapt through standardized semantic upgrades and protocol integration.
- ✓ **NLWeb**: Legacy websites require refactoring—including data structuring, semantic enhancement, and the deployment of standardized protocol endpoints—to facilitate more seamless access and utilization by AI agents.
- ✓ **WebMCP**: Web developers encapsulate application functionalities and *explicitly register them to browser as JavaScript Tools*, enabling AI agents to directly invoke website features with a level of speed, precision, and reliability that far exceeds human perception

## ● Short-term Solution

- ✓ Autonomously building Web MCP Tools based on interaction behaviors with existing websites, enabling cost-effective and high-efficiency integration of legacy sites into the Agentic Web network.

OpenHarmony ArkWeb:

- Will support WebMCP too





# Thanks



Email: [liuyaoming1@huawei.com](mailto:liuyaoming1@huawei.com)