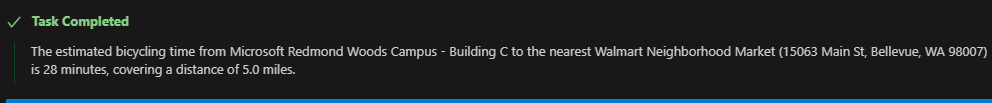
Agent marketplace engineering weekly update

1. Preparing to onboard Google Map MCP for case that Aworld not able to handle

Question: How long does it take from Microsoft Redmond Woods Campus- Building C to the nearest Walmart by Bike? (28 mins). 3 MCP calls could do the work. Currently integrate with Roo Code to test.



While using Aworld, heavily rely on Google Search and Browser Use to get Address info and interact with Google Map Webpage, and result is wrong and slow for too much attempts. Browser based info retrieval often result in wrong answer, because we could have too much source of truth.

1. Validation run comparison

Overall tool selection matched tool usage for Aworld.

Some test failed for lacking mcp\_\_reasoning\_\_mcp\_complex\_problem\_reasoning. But other without it could still pass, but the basic selection for mcp is correct, for example needed to read excel.

A computer screen with white and blue text

AI-generated content may be incorrect.

A black screen with white text

AI-generated content may be incorrect.

A computer screen shot of a computer code

AI-generated content may be incorrect.

A computer screen with text

AI-generated content may be incorrect.

Other are heavily use search tools, then marketplace version will use Tavily most. But still have correct result.

A screenshot of a computer

AI-generated content may be incorrect.

Task break down might be different, for no marketplace version, it seems to search all item in wiki to get correct info for categorizing.

Aworld: GPT 5

"Step 1: Task Analysis\n- Sub-task: Identify vegetables that are not botanical fruits.\n  Goal: Select items that are plant parts other than fruits.\n  Action: Verify plant part classification for candidate items.\n\n- Sub-task: Verify correctness using reliable sources.\n  Goal: Ensure selected items are not botanical fruits.\n  Action: Use Wikipedia summaries to confirm plant part nature.\n\n- Sub-task: Thinking Process Reviewing.\n  Goal: Diagnose potential oversights.\n  Action: Use the reasoning tool to double-check selection criteria.\n\n- Sub-task: Final formatting.\n  Goal: Provide alphabetized, comma-separated list.\n  Action: Sort and output within required tags.\n\nStep 2: Information Gathering\n- Candidate vegetables from the list: fresh basil, broccoli, celery, lettuce, sweet potatoes.\n- We will verify each to ensure none are botanical fruits.\n\nStep 3: Tool Selection\n- Use Wikipedia summaries in parallel to confirm each item’s botanical classification.\n"

Marketplace: GPT 4.1

LLM result for task breakdown and keywords: {

  "keywords": [

    [

      "botanical classification",

      "food categorization",

      "vegetable identification",

      "list sorting",

      "alphabetization"

    ]

  ],

  "subtasks": [

    "Identify which items from the provided grocery list are vegetables using botanical criteria and exclude botanical fruits.",

    "Sort the vegetable list alphabetically and format it as a comma-separated list."

  ],

  "explanation": "1. \*\*Identify botanical vegetables\*\*: This step requires knowledge of botany to accurately distinguish botanical vegetables from fruits based on strict scientific criteria (for example, items with seeds or developing from flowering parts are typically fruits in botanical terms). The necessary action here is \"botanical classification\" and \"food categorization.\" Also, \"vegetable identification\" is the relevant tool function.\n2. \*\*Sort and format the list\*\*: Next, the correctly identified vegetables need to be sorted alphabetically and formatted as a comma-separated list. So the tool's functions here are \"list sorting\" and \"alphabetization,\" focusing on the arrangement and formatting rather than the content itself."

}

A computer screen shot of text

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

All requested tools from tool marketplace are search related tools.

A screenshot of a computer

AI-generated content may be incorrect.

For Marketplace Disabled cases, different types of tools are used. Video analyze, Wikipedia, code runner, YouTube video download, docx

A screenshot of a computer

AI-generated content may be incorrect.

A group of white rectangular shapes

AI-generated content may be incorrect.

1. Vector retrieval with IDB integration done
2. Add setup script and data migration for Tool description loading
3. Improving current tool keyword extraction logic from subtask.

Next steps:

1. Find 10 Gaia test that current Marketplace MCP supported to run to compare result.
2. Gaia MCP tool keywords to update current prompt

A white background with blue text

AI-generated content may be incorrect.