# GeoUI Intent Protocol

A protocol between AI agents and geospatial visualization tools.

**Cheaper per task** and correct where screenshot agents fabricate.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

NASA Grantee

## 01 WHY THIS MATTERS

**An agent that can't tell what region it's looking at can't be trusted to summarize, chart, or recommend on top of it.**

Geospatial visualization applications render through **canvas / WebGL** . The DOM does not expose the map's semantic state (viewport bounds, active layers, time index, selection), so an agent driving the page via screenshots and clicks is reasoning over pixels, not over the map.

**The cost compounds:** every refinement turn re-uploads a screenshot, every guess about region or layer can be wrong, and every click depends on a coordinate the agent inferred from an image.

## 02 BACKGROUND

Two interaction mechanisms are compared on the same downstream application:

A. **Screenshot + DOM baseline.** Agent observes the page through periodic screenshots and accessibility-tree snapshots; acts through browser control (synthetic mouse and keyboard events) .

B. **GeoUI Intent Protocol.** Agent exchanges a typed `GeoIntent`payload with the application through a thin per-app adapter; viewport, layers and time are first-class fields featuring **STAC**-inspired extension capabilities.

**Application:** NASA Worldview, driven by the IESO Worldview agent. Browser control via Playwright in both arms.

**A typed GeoIntent between agent and map is dramatically cheaper and more accurate than driving the same application through screenshots, DOM, and clicks.**

ACROSS 4 WORLDVIEW TASKS

**~15x** **cheaper** per task on every measured axis

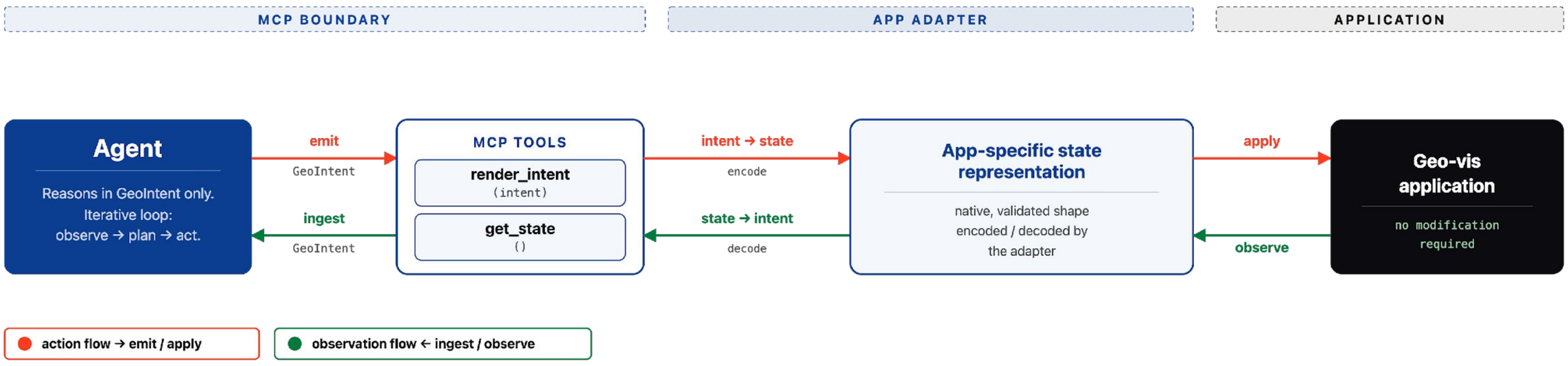
tokens · tool calls · round-trips · execution time

`"what region am I looking at?"`

Screenshot baseline **fabricates**. Protocol reads **viewport.bbox** directly.

PROTOCOL STACK

**Agent ↔ MCP ↔ App adapter ↔ Geo-visualization app**




## 03 APPROACH

A · BASELINE PATH

### Screenshots + DOM + clicks

Map state is invisible — **region, layers, time are guessed from pixels**.
Every turn re-uploads a screenshot; tokens and tool calls scale with refinement count.
Clicks depend on coordinates inferred from a rasterized image.

B · GEOUI INTENT PROTOCOL

### Typed GeoIntent + thin adapter

Viewport, layers, time, selection are **first-class fields** the agent reads directly.
Each task collapses to one or two typed exchanges — no screenshot uploads.
New applications integrate by writing a thin adapter; the agent does not change.

## 04 DATA & RESULTS

**Setup.** Four Worldview queries selected to isolate the interaction mechanism — region recall after a pan, layer state, time-step lookup, and a targeted layer toggle. Same LLM, same target app; only the channel changes.

GEOUI VS VLM · ABSOLUTE COSTS PER TASK · N=4 WORKFLOWS

**15× lower per-task cost — driven by tokens, requests, and tool calls**

| METRIC | GEOUI | VLM | RATIO |
|---|---|---|---|
| Input tokens | 48k | 839k | 17.5× |
| Output tokens | 2k | 18k | 9.0× |
| Total tokens | 51k | 857k | 16.8× |
| Requests | 6 | 36 | 6.0× |
| Tool calls † | 4 | 60 | 15.0× |
| Reasoning tokens † | 764 | 16k | 20.9× |
| Cache-read tokens † | 36k | 848k | 23.6× |

GeoUI Intent Protocol — VLM baseline (screenshot + DOM)

*Bars on log scale, normalized to 857k (Total tokens, VLM). Averaged over n=4.*

| A · BASELINE | B · PROTOCOL |
|---|---|
| After a pan: `"what region am I currently viewing?"` | |
| ✗ Fabricates<br>Names a region inconsistent with the actual viewport — the bounds are not visible in DOM and the screenshot does not carry coordinates. | ✓ Correct<br>Reads `viewport.bbox` directly from the GeoIntent and answers in decimal degrees, no inference. |

## 05 CONCLUSIONS

| | |
|---|---|
| **15×** | **cheaper per task** across tokens, tool calls, round-trips, and wall-clock |
| **1** | **round-trip per task** for typical Worldview queries vs. many refinement turns in the baseline |
| **0** | **pixel uploads** on the protocol path; the agent never sees a screenshot of the map |

The GeoUI Protocol delivers measurable efficiency in agent–application interaction through **fewer tool-use round-trips and lower token usage per task** , while also handling tasks correctly.

### NEXT

N1 **Application-state retrieval in the protocol itself** — remove the Playwright dependency for state read-back.

N2 **More adapters** across geospatial visualization tools beyond Worldview.

N3 **Extensions to GeoIntent** — drawing tools, time animation, layer query semantics.


AUTHORS & CONTACT

**Sanjog Thapa[1], Nishan Pantha[1], Gaurab Pathee[1], Rohit Sahoo[1], Muthukumaran Ramasubramanian[1], Iksha Gurung[1], Rahul Ramachandran[2]**

University of Alabama in Huntsville[1] · NASA Marshall Space Flight Center[2]

sanjog.thapa@uah.edu



ACKNOWLEDGEMENTS

The material contained in this document is based upon work supported by a National Aeronautics and Space Administration (NASA) grant or cooperative agreement. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NASA.

This work is supported by NASA Grant 80MSFC22M004.