Herzberg Astronomy and Astrophysics Research Centre
Canadian Astronomy Data Centre

StarAI

empowering access to astronomy data handled by Canadian Astronomy Data Centre and its CANFAR platform with artificial intelligence

Adrian Damian, Hossen Teimoorinia, Mrunal Mustapure, Serhii Zautkin

Building a system that can orchestrate use of large language models within CADC infrastructure and data
starai.ca • canfar.testapp.ca/star-ai

Presented at ADASS XXXV by Serhii Zautkin

Why StarAI (and Why Now)

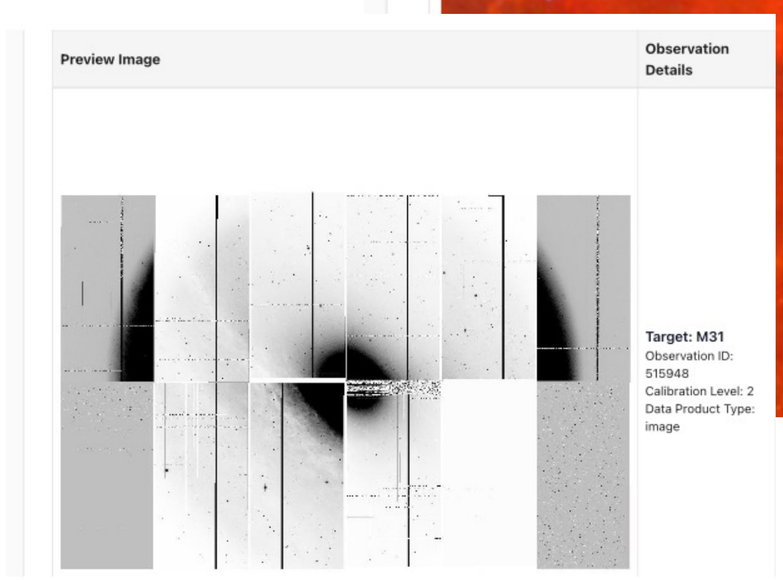
The Canadian Astronomy Data Centre (CADC) hosts one of the world's richest multi-observatory archives and provides a unique research experience developing and supporting the CANFAR science platform.
StarAI is an attempt to find an optimal way to connect Large Language Models (LLM) and artificial intelligence agents (built on top of them) with CADC infrastructure. Also, StarAI is an opportunity to reimagine research activity supported by the CANFAR science platform with an enabled artificial intelligence (AI) agent that can facilitate interactions with the CANFAR platform and speed up user productivity.

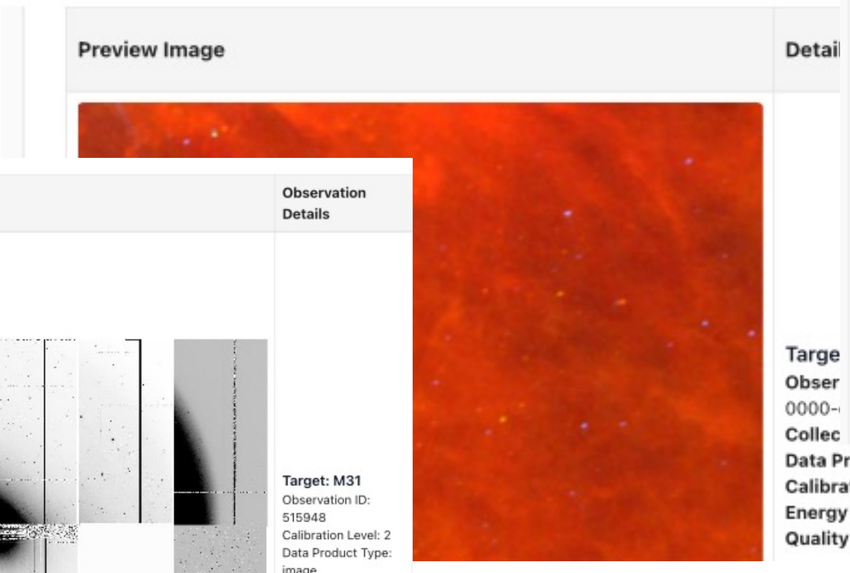
What It Does (Prototype)

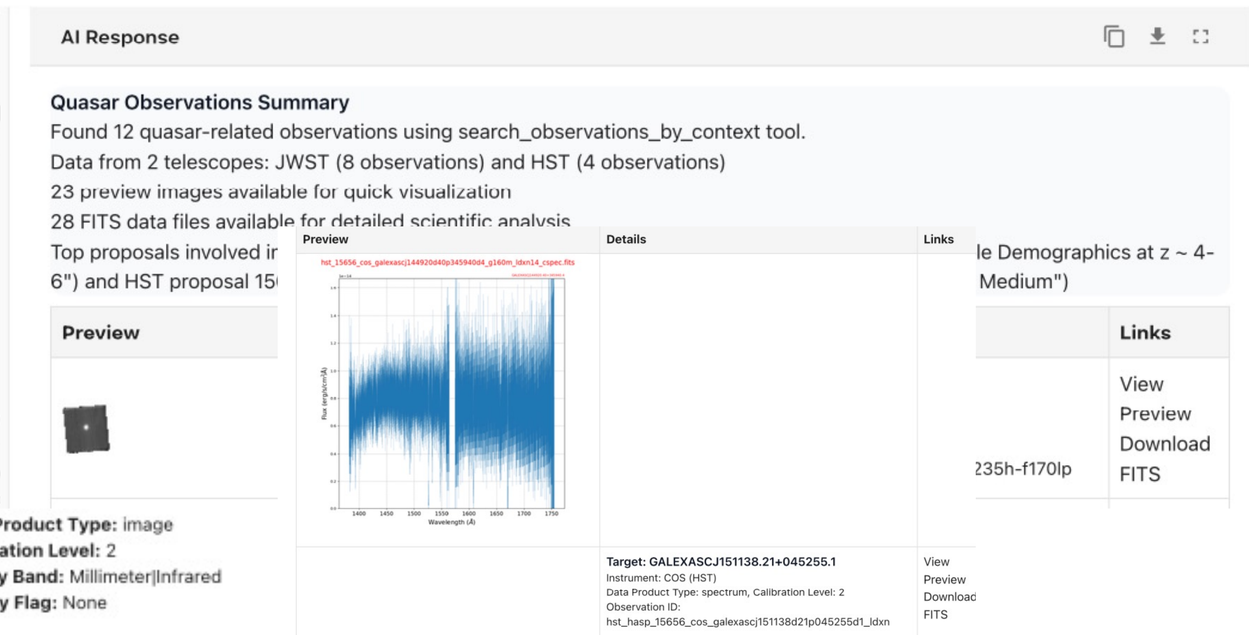
Conversational queries are translated into precise ADQL over CADC's TAP, returning observation records with resolved links to preview images, and FITS files.
Based on the Model Context Protocol (MCP), backend service bridges models to tools including ADQL, metadata, coordinate resolvers, images, and FITS file links.
Retrieval augmented generation (RAG) implemented as a separate service that supplements models with information about available observations based on the related research proposals (JWST and HST).

Sample Queries & Outputs

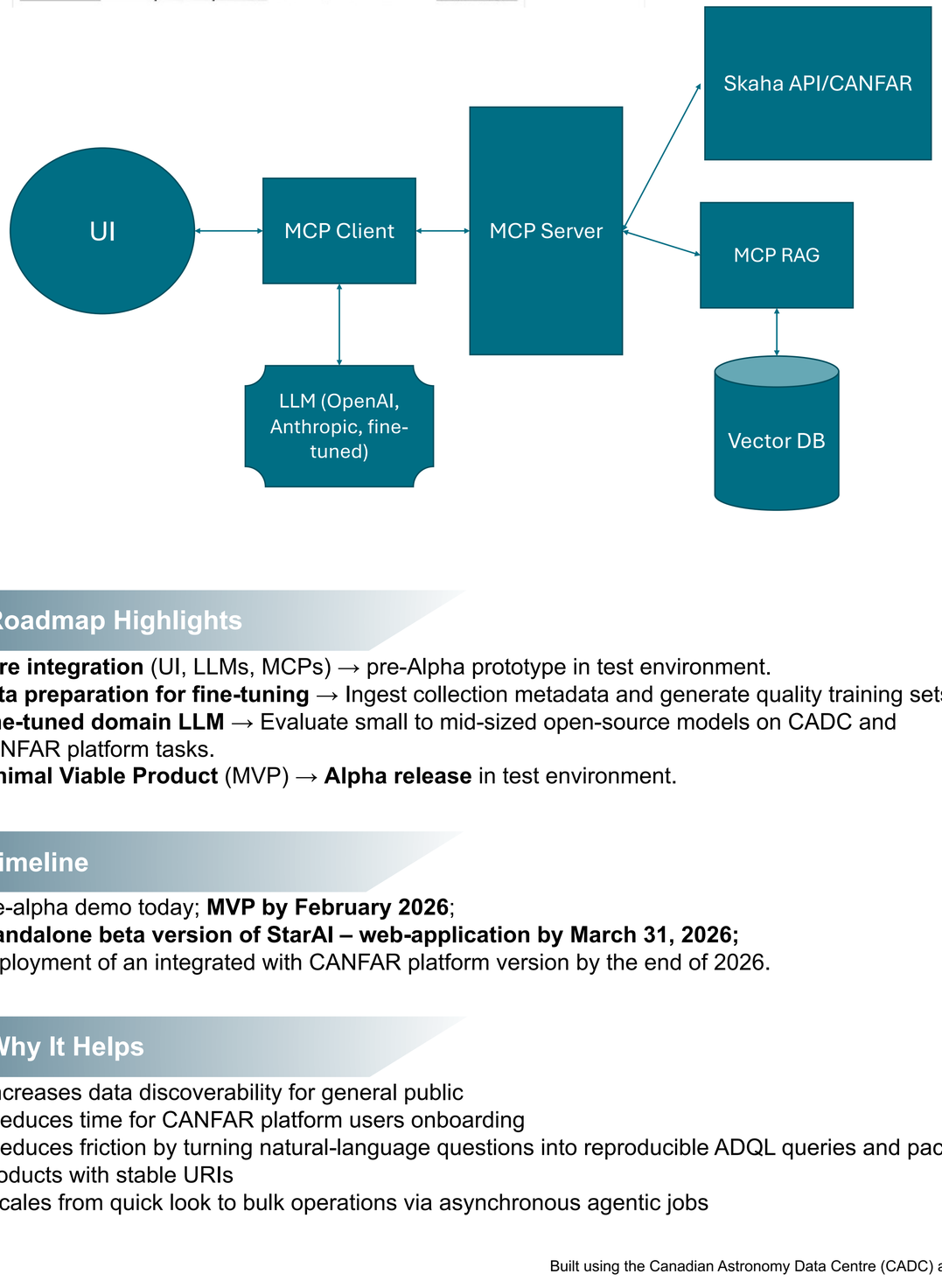
- "Any quasar observations?"
- "Infrared images of M31"
- "Iris collection images"







Architecture at a Glance



UI: a Nextjs, React, Typescript web-application

MCP Client: TypeScript/Express service bridge that connects with LLM providers (API or local), prompts models to evaluate user queries, analyse and use tools provided by MCP Server, handles and evaluates responses, and returns HTML rendering for results
MCP Server: TypeScript/Express service exposing ADQL, DataLink, coordinate resolution, CADC tables discovery, and RAG based search tools
MCP RAG: TypeScript/Express service connected to the Qdrant vector database of research proposals (JWST and HST) that helps to find relevant observations based on the plain language request as well as the CANFAR user knowledge base to facilitate the work of the StarAI chatbot
LLMs models accessed via APIs (OpenAI, Anthropic) or using locally hosted open-sourced models.
Deployment: Dockerized services, CI/CD; Pre-alpha version hosted at `https://canfar.testapp.ca/star-ai`.

Query Flow

User enters a query in the **UI**; the **MCP Client** receives it with any model settings/credentials.
The **MCP Client** connects to the **MCP server**, establishes a session, discovers available tools/resources, and constructs the system guidance/context passed to the model.
An iterative loop runs: the model requests operations; the **MCP Client** executes them via the **MCP server's** tools (including reading resources) and feeds results back until the model produces a final response.
MCP server tools may chain into internal services/APIs (e.g., Skaha and CANFAR), including Argus TAP endpoints exposing CAOM via TAP.
The **MCP Client** validates/normalizes the final HTML output and returns it to the UI.

Roadmap Highlights

Core integration (UI, LLMs, MCPs) → pre-Alpha prototype in test environment.
Data preparation for fine-tuning → Ingest collection metadata and generate quality training sets.
Fine-tuned domain LLM → Evaluate small to mid-sized open-source models on CADC and CANFAR platform tasks.
Minimal Viable Product (MVP) → **Alpha release** in test environment.

Timeline

Pre-alpha demo today; **MVP by February 2026**;
Standalone beta version of StarAI – web-application by March 31, 2026;
Deployment of an integrated with CANFAR platform version by the end of 2026.

How to Try It

Scan the QR code to open the live test app at **canfar.testapp.ca/star-ai**.

Try asking:
"Give me 10 observations of M31"
"Any quasar observations?"

Review the response, preview available images, download provided FITS files

Post your comments and feature requests.

Why It Helps

- Increases data discoverability for general public
- Reduces time for CANFAR platform users onboarding
- Reduces friction by turning natural-language questions into reproducible ADQL queries and packaged products with stable URIs
- Scales from quick look to bulk operations via asynchronous agentic jobs

Attribution

Built using the Canadian Astronomy Data Centre (CADC) archives and the CANFAR platform. Prototype queries and UI screens are derived from internal test runs. Please acknowledge CADC/CANFAR when using retrieved data.

Contact: Serhii.Zautkin@nrc-cnrc.gc.ca
Poster Size: A0 (84.1 × 118.9 cm)

National Research Council Canada
Conseil national de recherches Canada
Canada