Contribution ID: **80** Type: **not specified**

Local and Specialized AI Infrastructure for Next-Generation Astronomical Workflows

Tuesday, 10 March 2026 11:30 (20 minutes)

"The exponential growth of astronomical datasets, driven by next-generation observatories such as the Square Kilometre Array (SKA), demands a paradigm shift in how the community approaches data analysis and scientific workflows. Artificial intelligence, and large language models (LLMs) in particular, are emerging as transformative tools in this landscape. Here we describe the deployment of a local AI infrastructure based on five NVIDIA GB10 Grace-Blackwell Superchip nodes, each equipped with 128 GB of unified memory, designed to serve a research community with domain-specific AI capabilities while preserving full data sovereignty.
Our setup runs two classes of specialized open-weight models: AstroSage, a Llama-3.1-based LLM fine-tuned on the complete corpus of astronomical literature and achieving state-of-the-art performance on astronomy benchmarks —surpassing frontier proprietary models at a fraction of the inference cost —and Qwen3-Coder, a Mixture-of-Experts architecture optimized for agentic coding tasks. This dual-model approach reflects the complementary nature of scientific work: deep domain reasoning on one side, and automated code generation and pipeline management on the other.
Looking ahead, we argue that agentic AI workflows will become the dominant paradigm for managing complex astronomical pipelines. Tasks such as automated flagging, calibration, imaging, and source extraction in radio astronomy —exemplified by SKA data reduction pipelines —are natural candidates for LLM-driven agents capable of orchestrating multi-step processes, recovering from failures, and adapting to heterogeneous data conditions. Local, specialized infrastructure of the kind described here will be essential to support such workflows in a reproducible, resource-efficient, and privacy-preserving manner."

Author: LACOPO, Giovanni (Istituto Nazionale di Astrofisica (INAF))

Presenter: LACOPO, Giovanni (Istituto Nazionale di Astrofisica (INAF))

Session Classification: Artificial Intelligence