



Using AI in Open Data Context

DCP – KD session

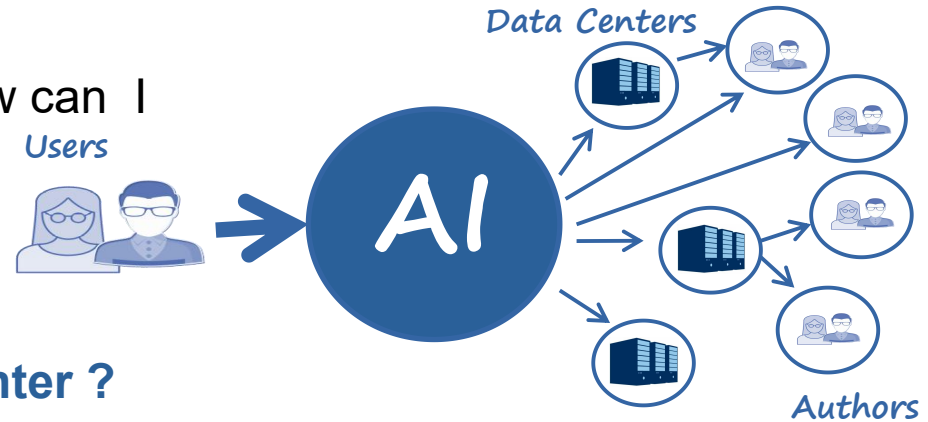
G.Landais, A.Muench

Y.Tao, A.Schaaff

DCP motivation – providing Trustworthy AI

In OpenScience context, when I consume AI, How can I

- trust the (Provenance) ?
- check copyrights ?
- credits authors ?



What can I do when I am working in a DataCenter ?

Solution1: Protection

- filter the AI/bots harvesting : do without powerfull tools ?
- Add licence/copyrights : is it consumed by GAFAM-AI ?

Solution2: Can I influence AI harvesting ?

How can I provide a trustworthy AI ?

- Am I subject to any rules?
- Are there ethic ? good practice guideline ?

From KD perspective

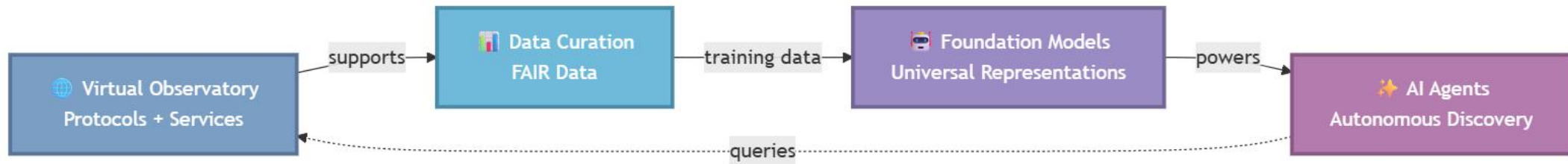
VO as the essential infrastructure connecting AI and astronomical data

Given the Emerging paradigms of AI for Astronomy

- Foundation Models: Universal Representations
- AI Agents: From Workflow Orchestration to Autonomous Discovery

How to provide FAIR data for AI?

How could VO tools and protocols be applied and integrated with agent-based workflows?



Discussion

Making AI from FAIR data

- AI is more cool with FAIR data, But,
- Can trust the (Provenance) ? check copyrights ? credits authors ?
- How Data Center can influence (GAFAM) harvesting ?
- Linking AI and VO services ? (MCP)
- Challenges and best practices for curating multi-modal multi-survey datasets for AI?

... and Making FAIR AI

- Experience in AI usage to populate metadata/data ?