Supporting Theory Simulations with Large Language Models

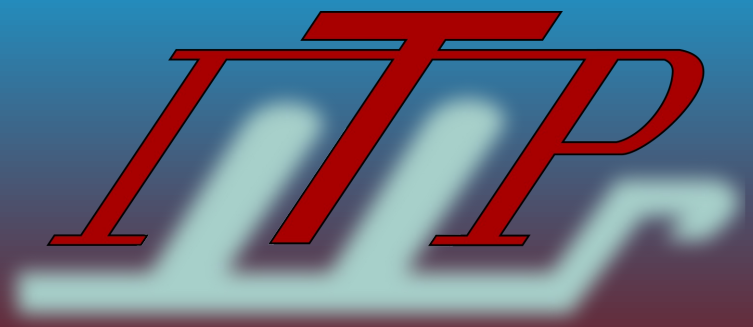
Daniel Schiller

15.05.2026, NPTwins, Trento

Based on **MadAgents** [arXiv:2601.21015]

With Tilman Plehn and Nikita Schmal



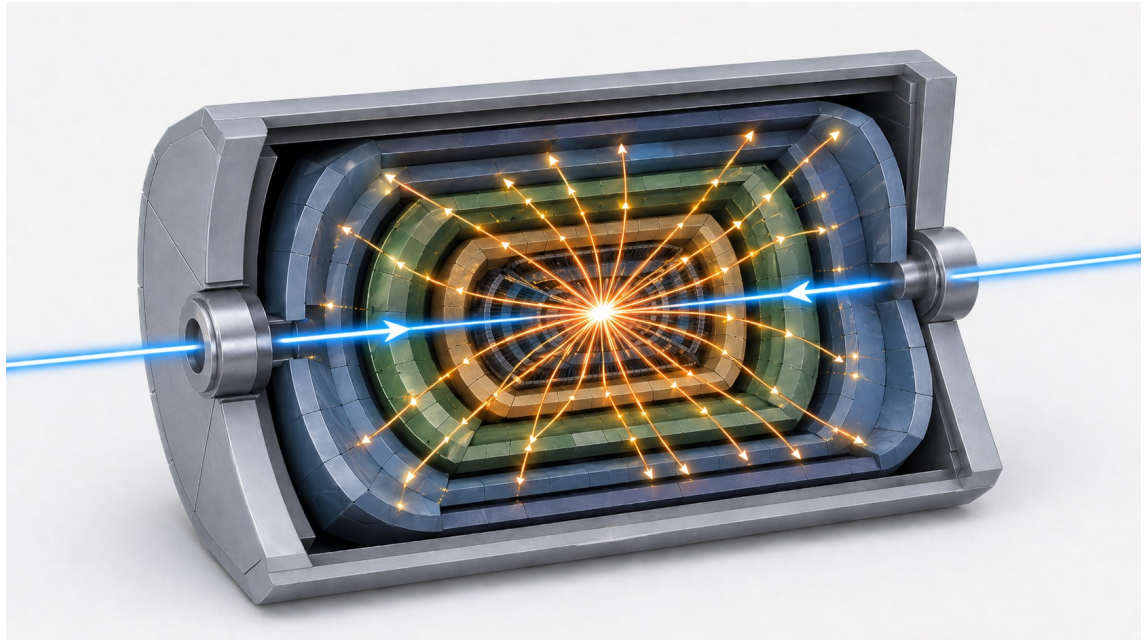




STRUCTURES
CLUSTER OF
EXCELLENCE

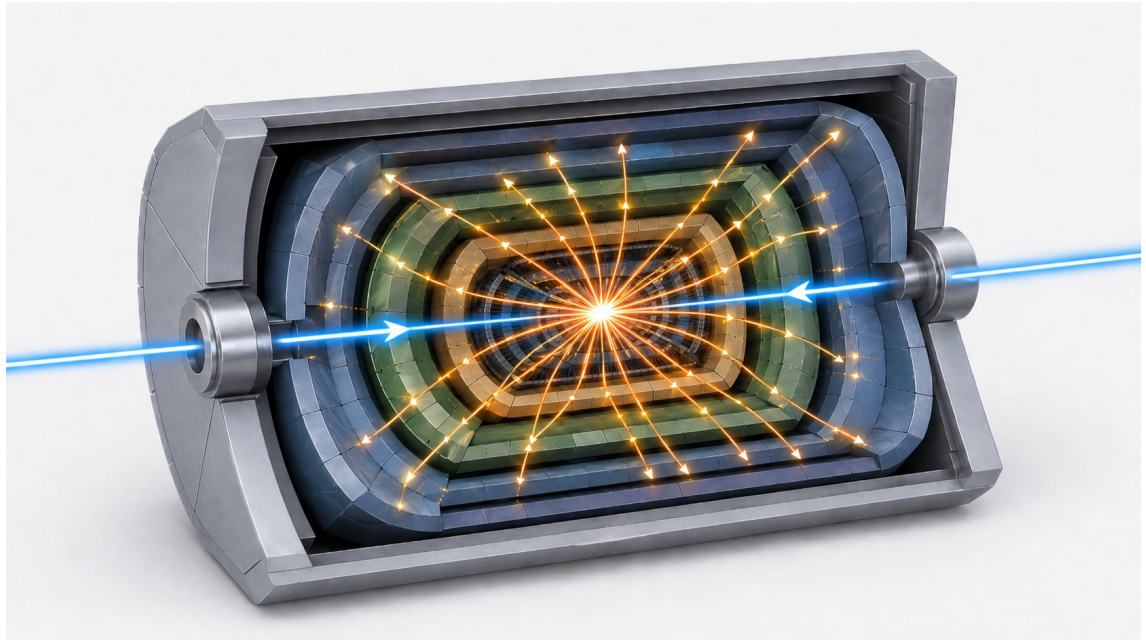
MadGraph

- Particle physics collider simulation
- Hard-scattering matrix-element generator
- Monte Carlo phase-space integrator



MadGraph

- Particle physics collider simulation
- Hard-scattering matrix-element generator
- Monte Carlo phase-space integrator
- Interfaces to
  - PDFs
  - Hadronization simulations
  - Detector simulations



MadGraph



- Based on first principles
- High accuracy within well-understood approximations

MadGraph



- Based on first principles
- High accuracy within well-understood approximations
- Over 30 years of development
- AI plugins

MadGraph



- Based on first principles
- High accuracy within well-understood approximations
- Over 30 years of development
- AI plugins
- **Powerful**, yet **challenging to use**…

```
Cross-section : nan +- nan pb
Number of events written: 0
```

```
./bin/madevent: line 123: gfortran: command not found
Error: Fortran compiler gfortran not found
```

The wall between two worlds

Developers:



Users:

The wall between two worlds

Developers:

- Add **advanced** capabilities



Users:

The wall between two worlds

Developers:

- Add **advanced** capabilities
- **Documentation** and **user interfaces** typically **limited**
- **Academic** development

Users:

The wall between two worlds

Developers:

- Add **advanced** capabilities
- **Documentation** and **user interfaces** typically **limited**
- **Academic** development

Users:

- **Use** MadGraph **for** their **research**
- Typically have **limited interest** in **technical details**

The wall between two worlds

Developers:

- Add **advanced** capabilities
- **Documentation** and **user interfaces** typically **limited**
- **Academic** development

MadAgents: Multi-Agent system as the **smart interface for MadGraph**

Users:

- **Use** MadGraph **for** their **research**
- Typically have **limited interest** in **technical details**

Large Language Model Agents

- LLMs can do more than answer questions. They can:
  - **reason** before providing an answer.
  - **inspect** images and PDF files.
  - **execute** tools, e.g. run bash commands, search the web, …
  - **interact** with other LLMs.
- A single LLM instance: **Agent**

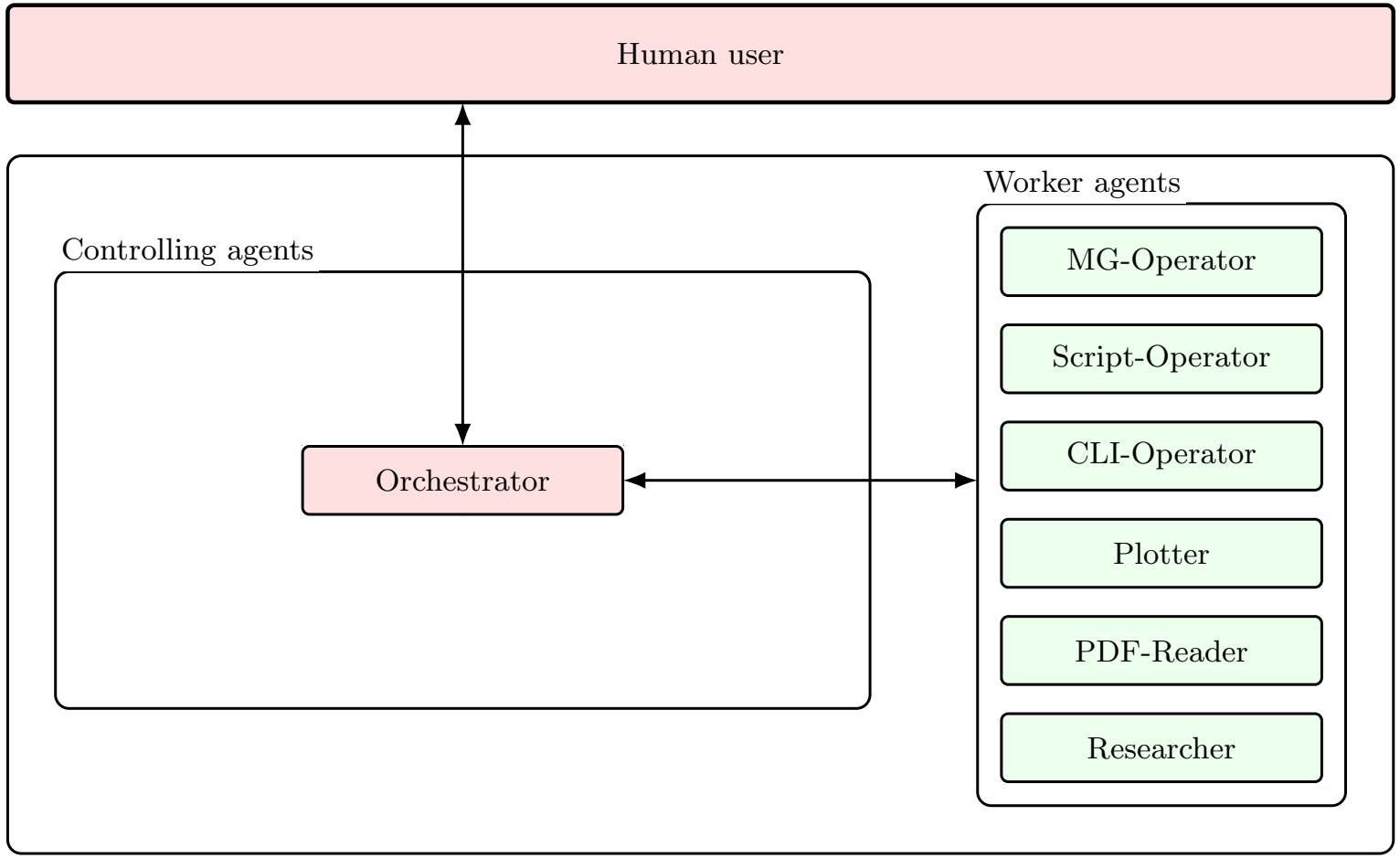
Large Language Model Agents

- LLMs can do more than answer questions. They can:
  - **reason** before providing an answer.
  - **inspect** images and PDF files.
  - **execute** tools, e.g. run bash commands, search the web, …
  - **interact** with other LLMs.
- A single LLM instance: **Agent**
- **MadAgents** = System of multiple **specialized agents**

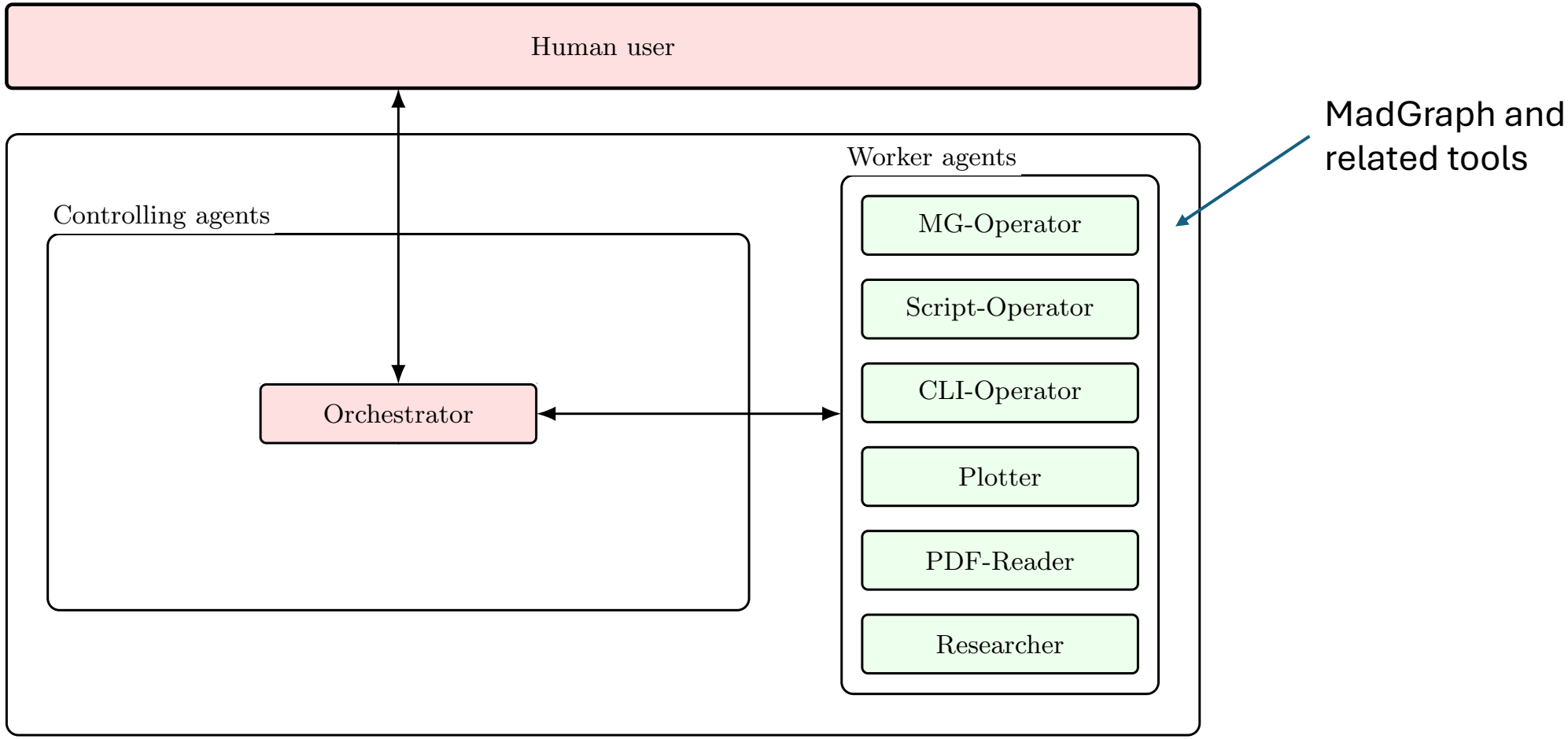
Large Language Model Agents

- LLMs can do more than answer questions. They can:
  - **reason** before providing an answer.
  - **inspect** images and PDF files.
  - **execute** tools, e.g. run bash commands, search the web, …
  - **interact** with other LLMs.
- A single LLM instance: **Agent**
- **MadAgents** = System of multiple **specialized agents**
- **Specialized** = instructions, tools, …

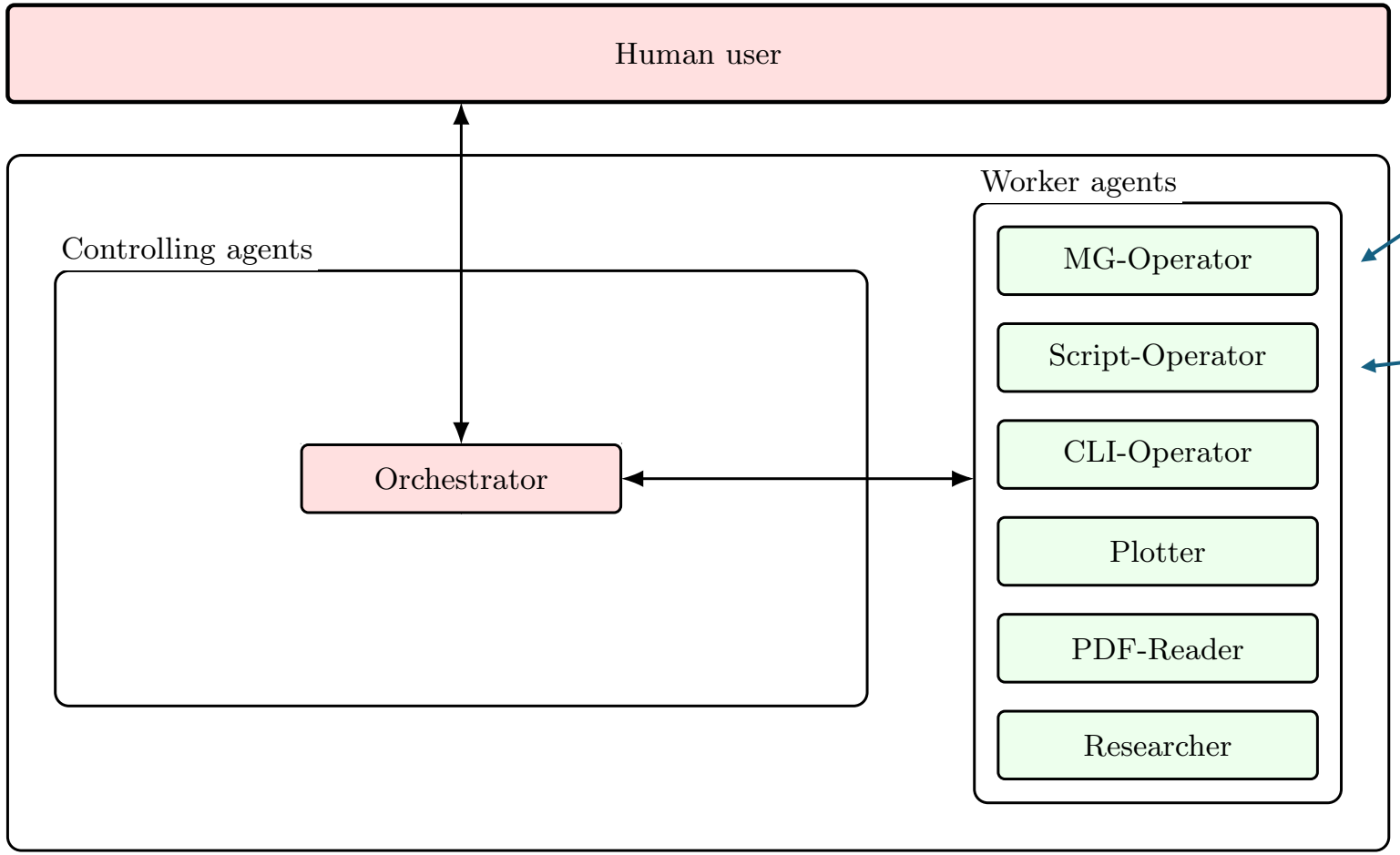
MadAgents


Human user
Worker agents
Controlling agents
MG-Operator
Script-Operator
CLI-Operator
Orchestrator
Plotter
PDF-Reader
Researcher

MadAgents


Human user
Controlling agents
Orchestrator
Worker agents
MG-Operator
Script-Operator
CLI-Operator
Plotter
PDF-Reader
Researcher
MadGraph and related tools

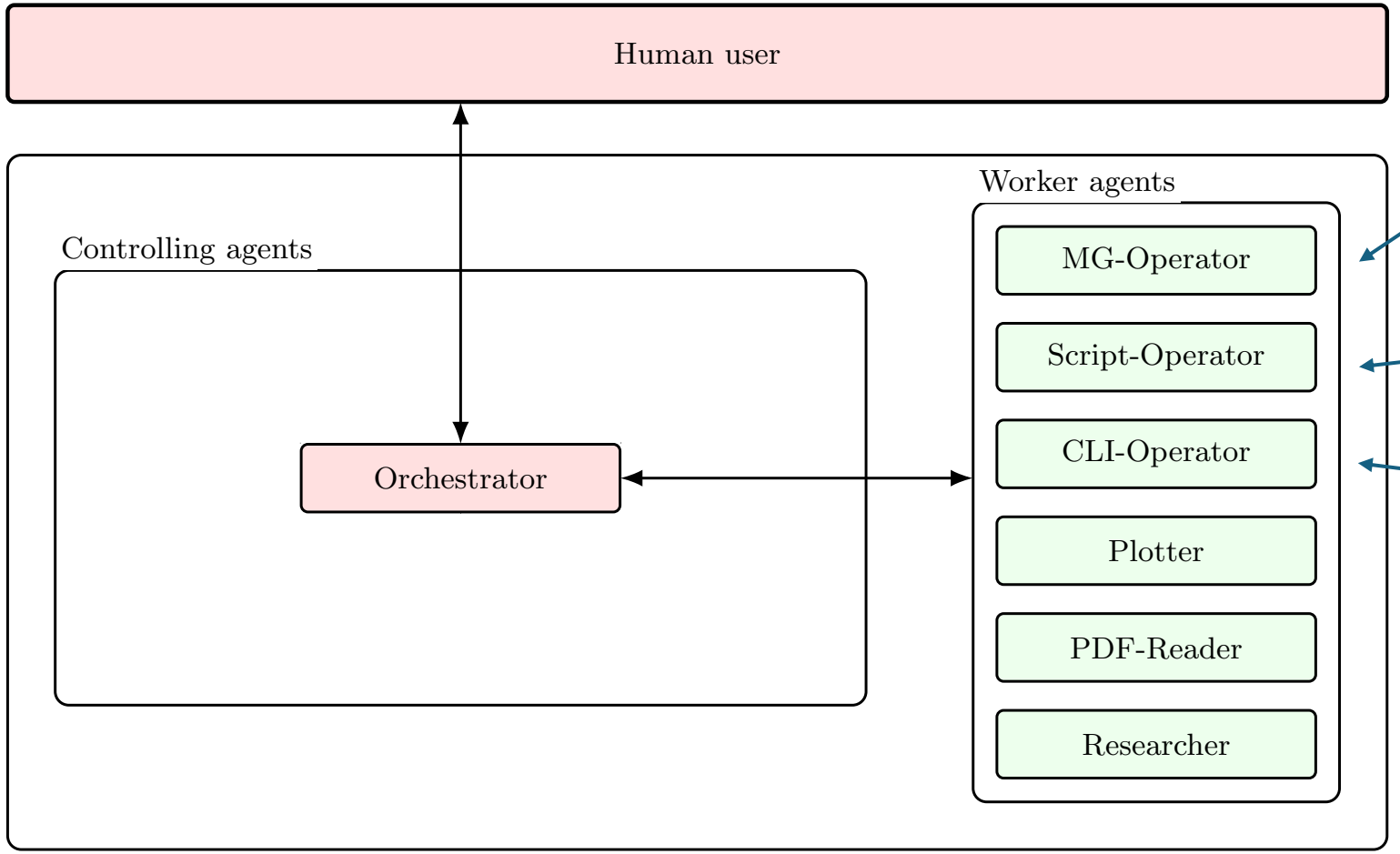
MadAgents



MadGraph and related tools

Generic bash and Python calls

MadAgents

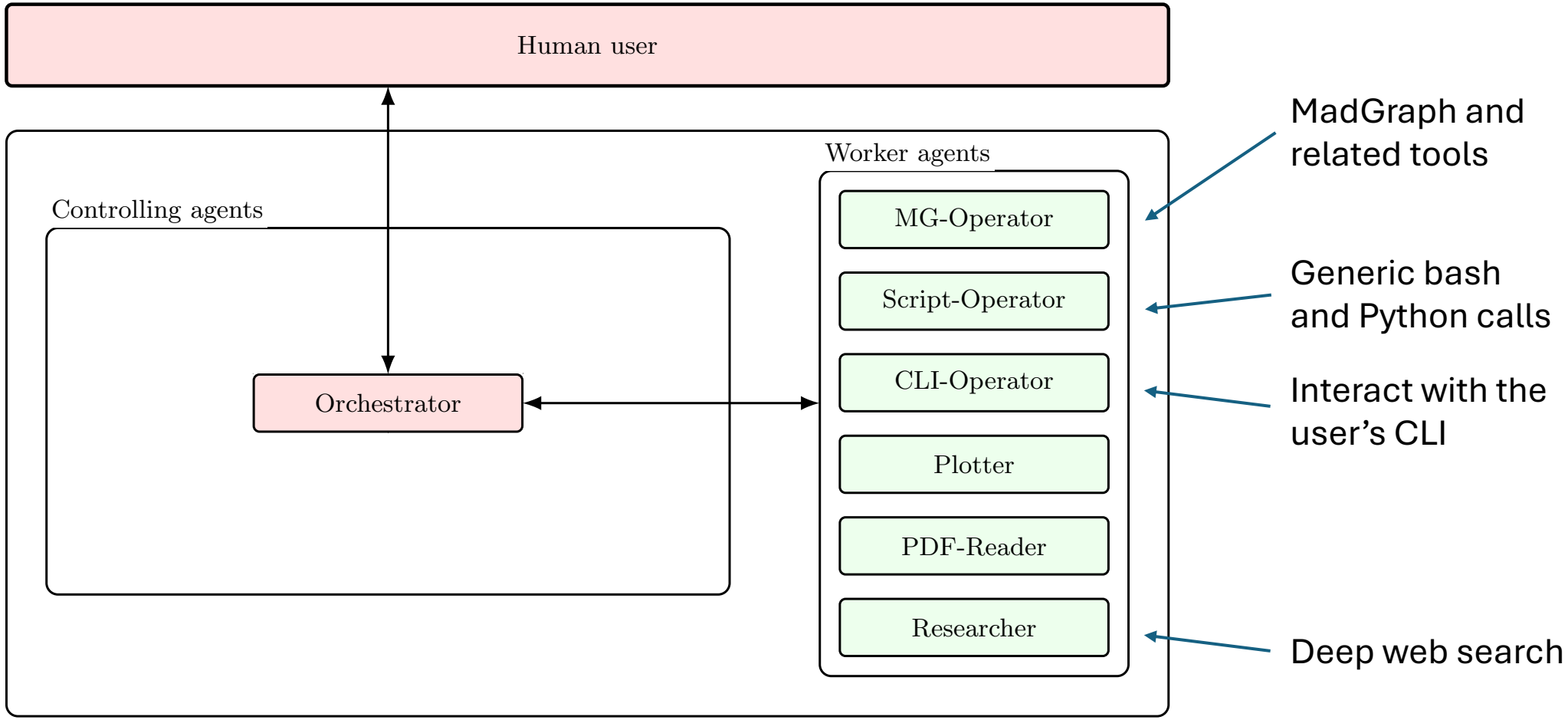


MadGraph and related tools

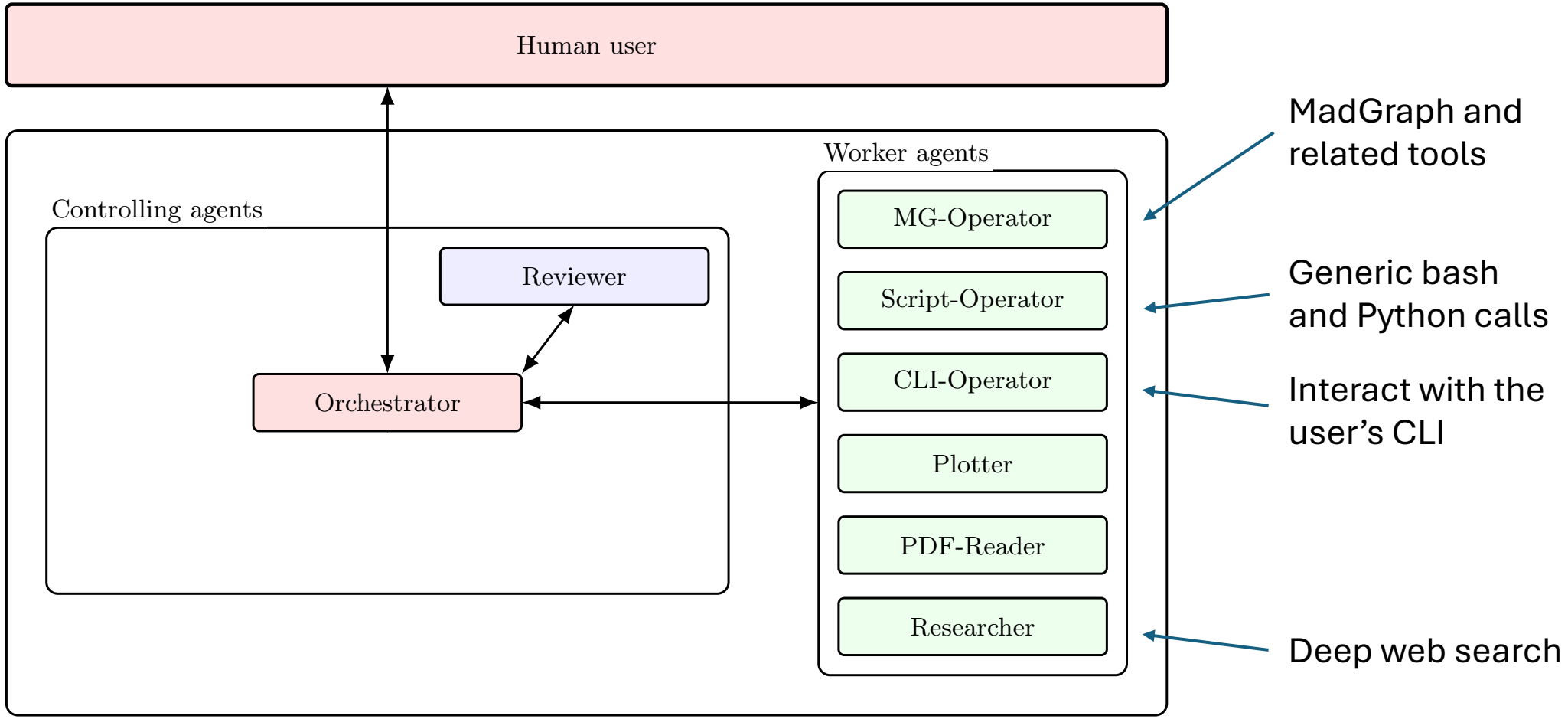
Generic bash and Python calls

Interact with the user's CLI

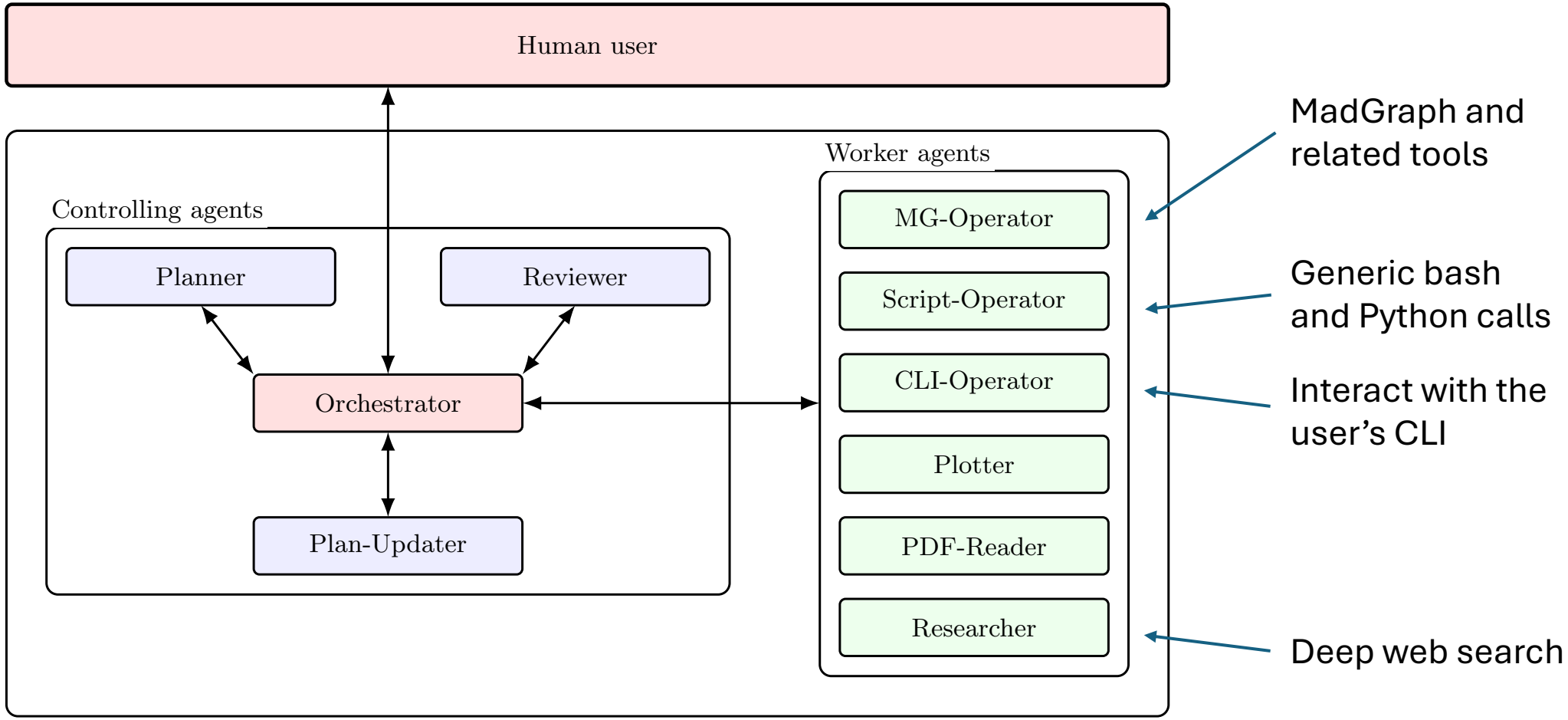
MadAgents


Human user
Controlling agents
Orchestrator
Worker agents
MG-Operator
Script-Operator
CLI-Operator
Plotter
PDF-Reader
Researcher
MadGraph and related tools
Generic bash and Python calls
Interact with the user's CLI
Deep web search

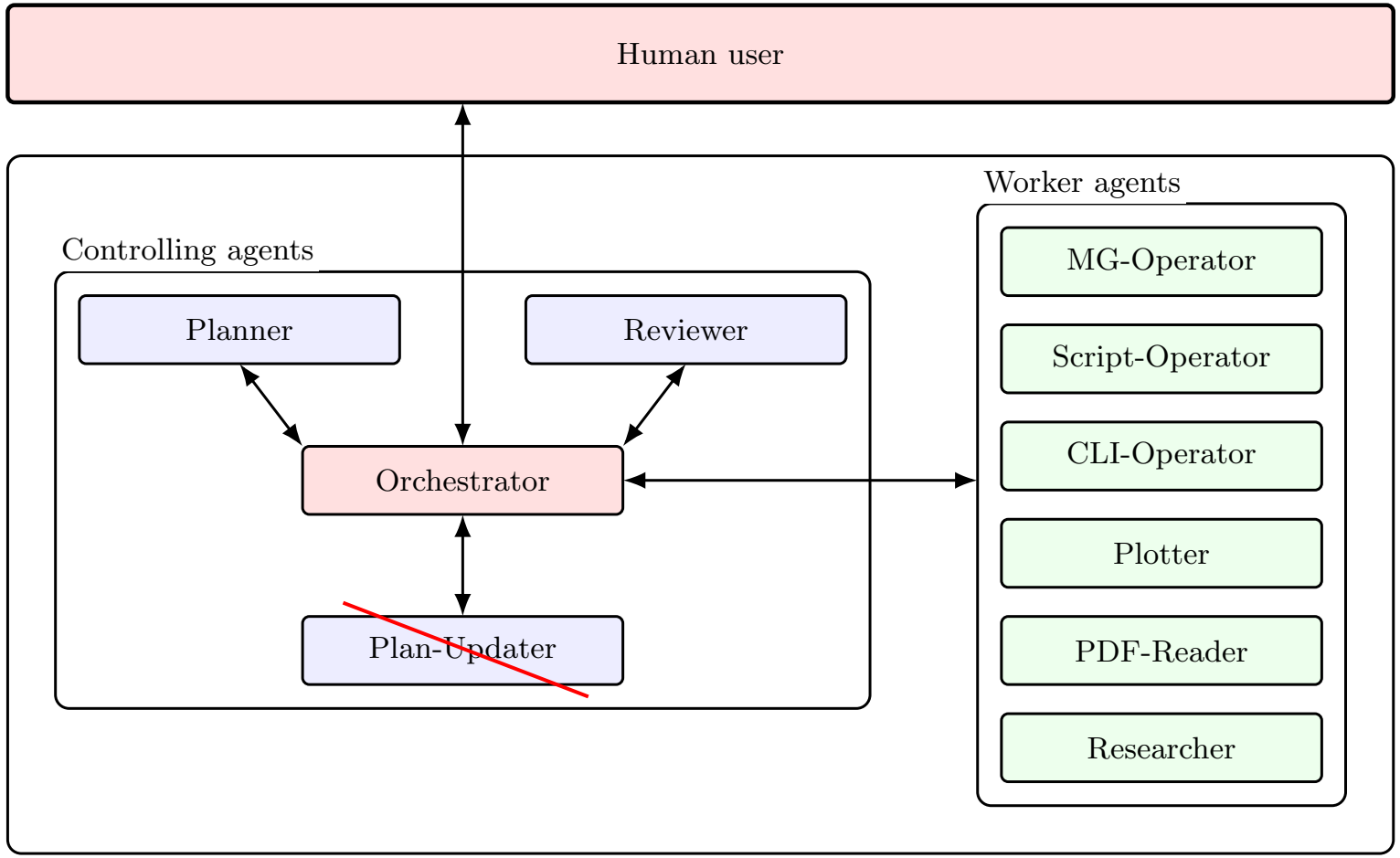
MadAgents


Human user
Worker agents
Controlling agents
Reviewer
Orchestrator
MG-Operator
Script-Operator
CLI-Operator
Plotter
PDF-Reader
Researcher
MadGraph and related tools
Generic bash and Python calls
Interact with the user's CLI
Deep web search

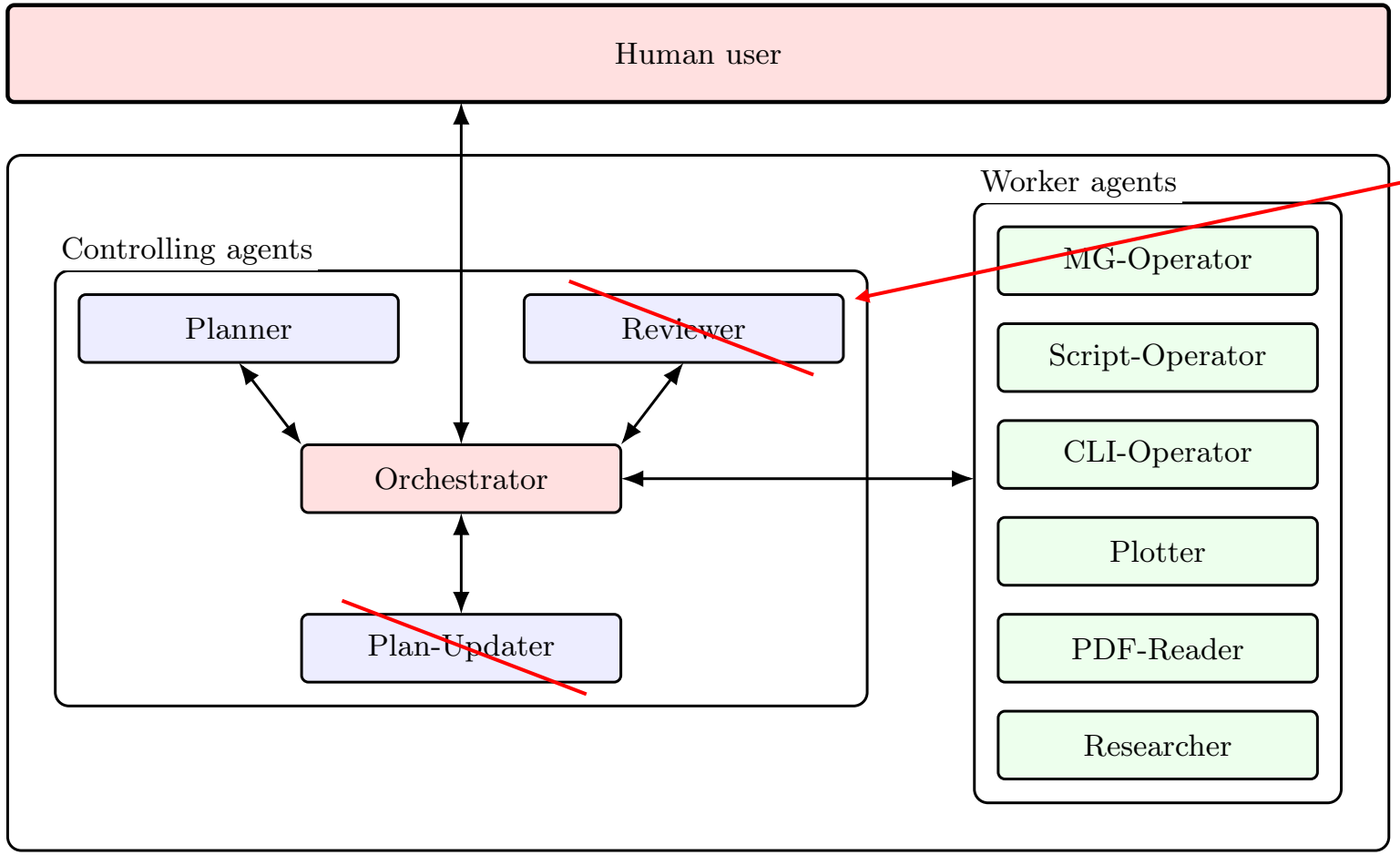
MadAgents


Human user
Worker agents
Controlling agents
Planner
Reviewer
Orchestrator
Plan-Updater
MG-Operator
Script-Operator
CLI-Operator
Plotter
PDF-Reader
Researcher
MadGraph and related tools
Generic bash and Python calls
Interact with the user's CLI
Deep web search

MadAgents Update



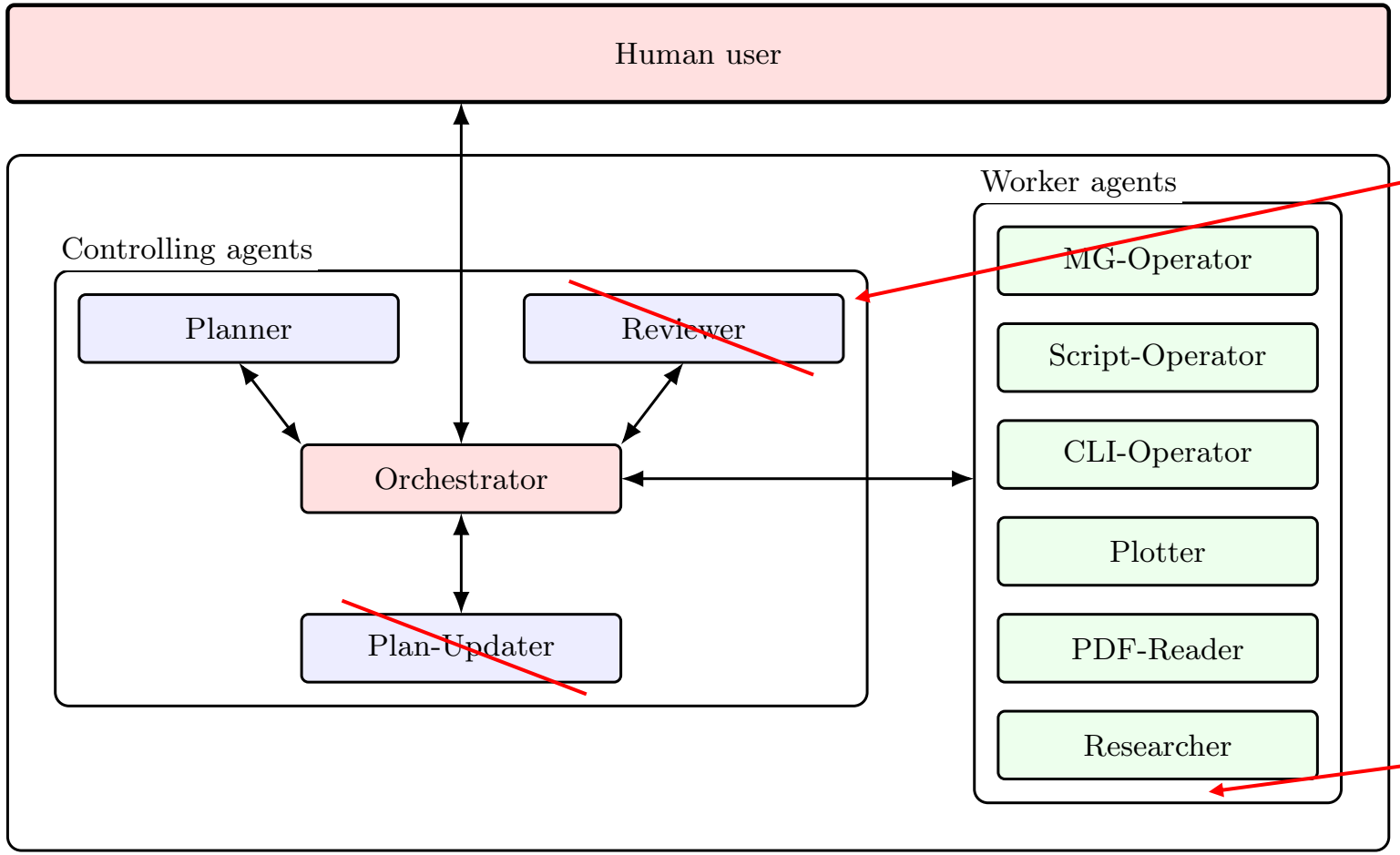
MadAgents Update



Specialized reviewers:

- Plan-Reviewer
- Verification-Reviewer
- Presentation-Reviewer

MadAgents Update



Specialized reviewers:

- Plan-Reviewer
- Verification-Reviewer
- Presentation-Reviewer

Physics-Expert

Flexible Orchestration

- No hard-coded series of agent calls – the Orchestrator decides
- Flexible user interaction

Flexible Orchestration

- No hard-coded series of agent calls – the Orchestrator decides
- Flexible user interaction
- MadAgents can support users at different stages, e.g.:
  - **Installation**
  - **Teaching inexperienced users**
  - **Advanced user-support**
  - **Autonomous execution**

Highlights

with GPT-5.1

MadGraph Installation

- **Task**: Install MadGraph+ Pythia8 + Delphes and compile ROOT

MadGraph Installation

- **Task**: Install MadGraph+ Pythia8 + Delphes and compile ROOT
- **Problem**: MadGraph did not find the ROOT executable

MadGraph Installation

- **Task**: Install MadGraph+ Pythia8 + Delphes and compile ROOT
- **Problem**: MadGraph did not find the ROOT executable
- MadAgents:
  - Inspected the MadGraph source code
  - Understood that MadGraph wanted a ROOT executable ending with “.exe”
  - Created a symbolic link

Tailored Tutorials

- **Task**: Create tutorials for a beginner worker on Simulation Based Inference

Tailored Tutorials

- **Task**: Create tutorials for a beginner worker on Simulation Based Inference
- **Outcome**: 6-step tutorial:
  1. MG basics
  2. SM parton level sim.
  3. Hadronization sim.
  4. Detector sim.
  5. BSM sim.
  6. SBI datasets

Tailored Tutorials

- **Task**: Create tutorials for a beginner worker on Simulation Based Inference
- **Outcome**: 6-step tutorial:
  1. MG basics
  2. SM parton level sim.
  3. Hadronization sim.
  4. Detector sim.
  5. BSM sim.
  6. SBI datasets

MadAgents act as a **teacher**:

- Offer explanations
- Provide hints
- Help users on the fly

Adaptive, real-time teaching

Dataset Documentation

- Generated $t\bar{t}$ events with at
  - LO and NLO with the SM.
  - LO with a BSM model.
- Applied K-factor reweighting
- **Task**: Create a detailed documentation

Dataset Documentation

- Generated $t\bar{t}$ events with at
  - LO and NLO with the SM.
  - LO with a BSM model.
- Applied K-factor reweighting
- **Task**: Create a detailed documentation

3.2 Top-quark parameters

From the param cards (`Cards/param_card.dat`):

- **SM LO and NLO samples**
  - Top mass: $m_t = 173\,\mathrm{GeV}$.
  - Yukawa: $y_t = 173\,\mathrm{GeV}$ (in SM convention).
  - Width: $\Gamma_t \approx 1.49\,\mathrm{GeV}$.
- **EFT UFO sample**
  - Top mass: $m_t = 172\,\mathrm{GeV}$.
  - Yukawa: $y_t = 172\,\mathrm{GeV}$.
  - Width: $\Gamma_t \approx 1.51\,\mathrm{GeV}$.
  - DIM6 block (excerpt): the EFT scale parameter is set to $\lambda = 1\,\mathrm{TeV}$, and among the listed Wilson coefficients only c_{tG} is non-zero; all other entries in that block are set to zero.

This introduces a small mismatch between the SM reference samples and the EFT UFO sample in m_t and Γ_t, which is acceptable for many use cases but should be kept in mind for precision studies.

Provide Expert Physics Insight

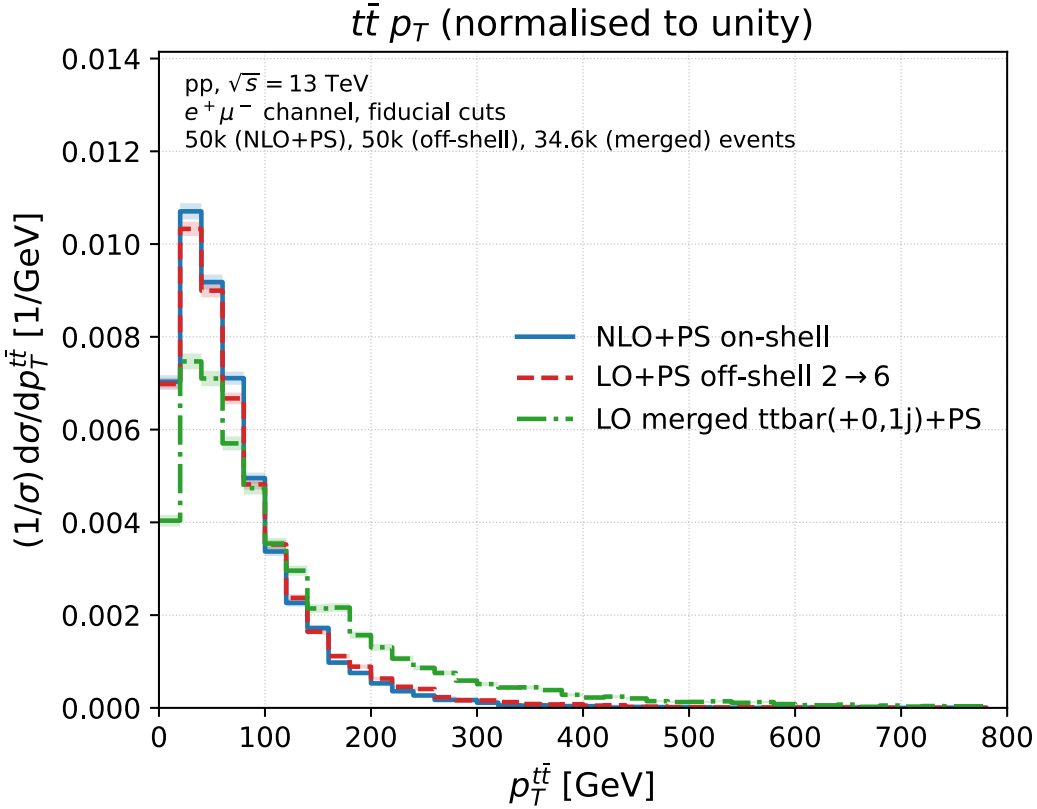
- Experienced user has LO, on-shell $\mathrm{pp} > t\bar{t}$ production
- **Task**: precision improvements

Provide Expert Physics Insight

- Experienced user has LO, on-shell $\mathrm{pp} > t\bar{t}$ production
- **Task**: precision improvements
- MadAgents discuss:
  - Perturbative corrections
  - Off-shell effects
  - Multi-jet merging

Provide Expert Physics Insight

- Experienced user has LO, on-shell $pp > t\bar{t}$ production
- **Task**: precision improvements
- MadAgents discuss:
  - Perturbative corrections
  - Off-shell effects
  - Multi-jet merging
- Creates a plot to visualize the effects

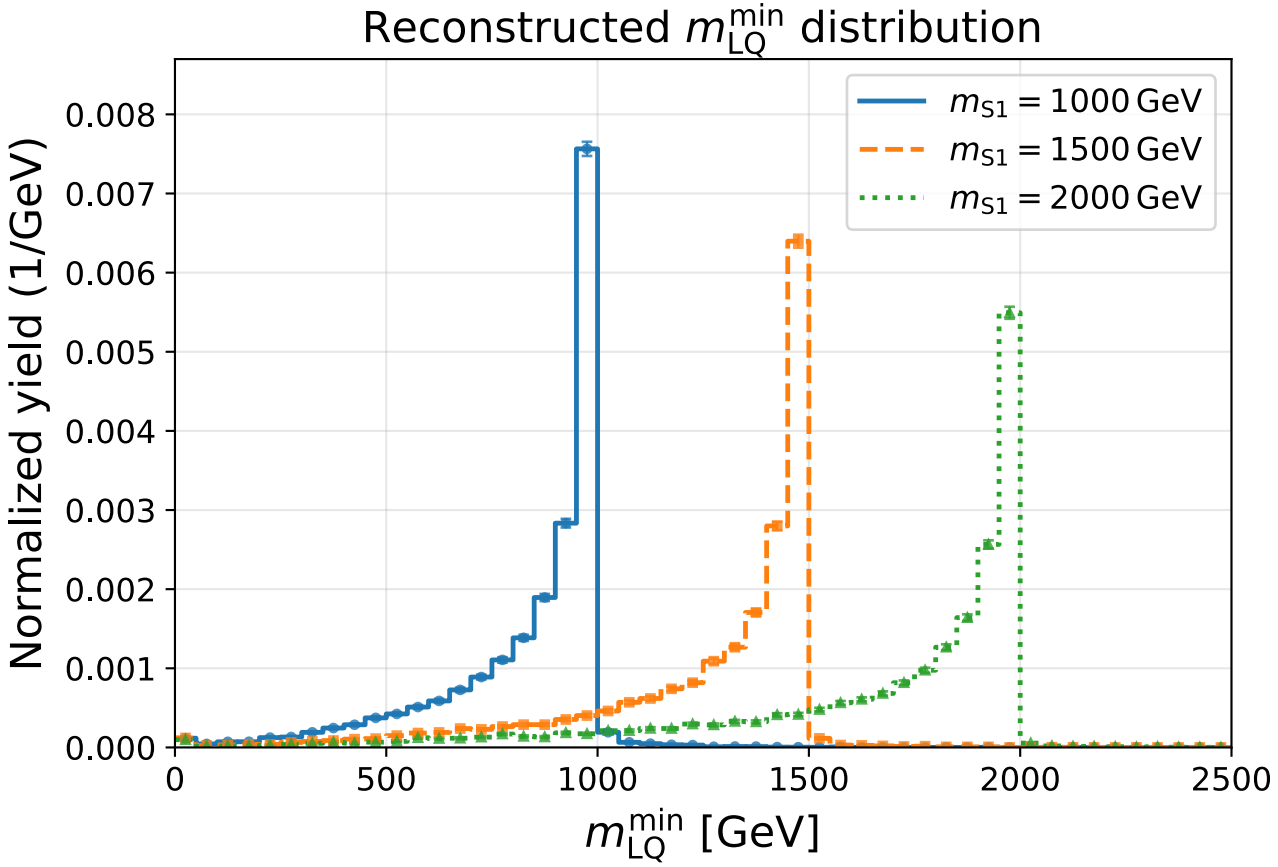


Autonomous Event Generation

- **Task**: Generate data from PDF file (BSM scalar leptoquark)
- Searched the web for a suitable BSM model

Autonomous Event Generation

- **Task**: Generate data from PDF file (BSM scalar leptoquark)
- Searched the web for a suitable BSM model
- **Outcome**: events and plot



Self-Improvement

- Extend the **internal documentation** via **self-improvement**

Self-Improvement

- Extend the **internal documentation** via **self-improvement**



Self-Improvement

- Extend the **internal documentation** via **self-improvement**



Verification:

1. Split into a list of claims
2. Verify each claim with unambiguous evidence

Self-Improvement

- Extend the **internal documentation** via **self-improvement**

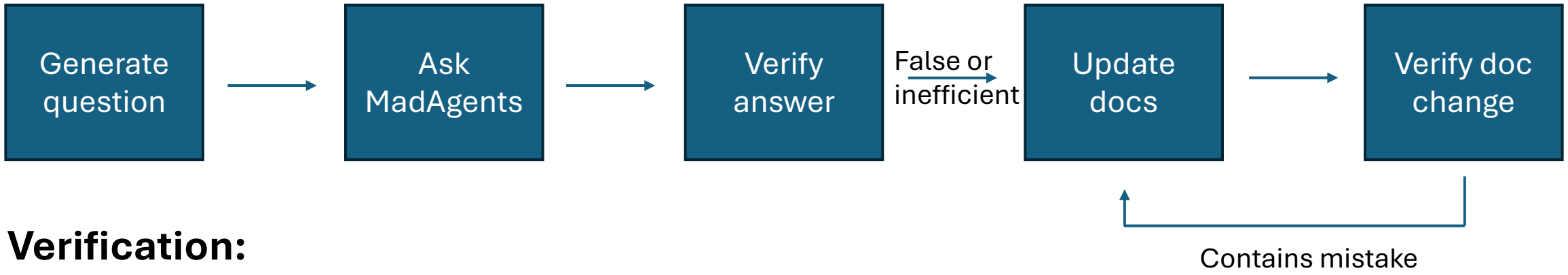


Verification:
1. Split into a list of claims
2. Verify each claim with unambiguous evidence

Forces a **frontier model** to **reason** and **explore deeply**!

Self-Improvement

- Extend the **internal documentation** via **self-improvement**



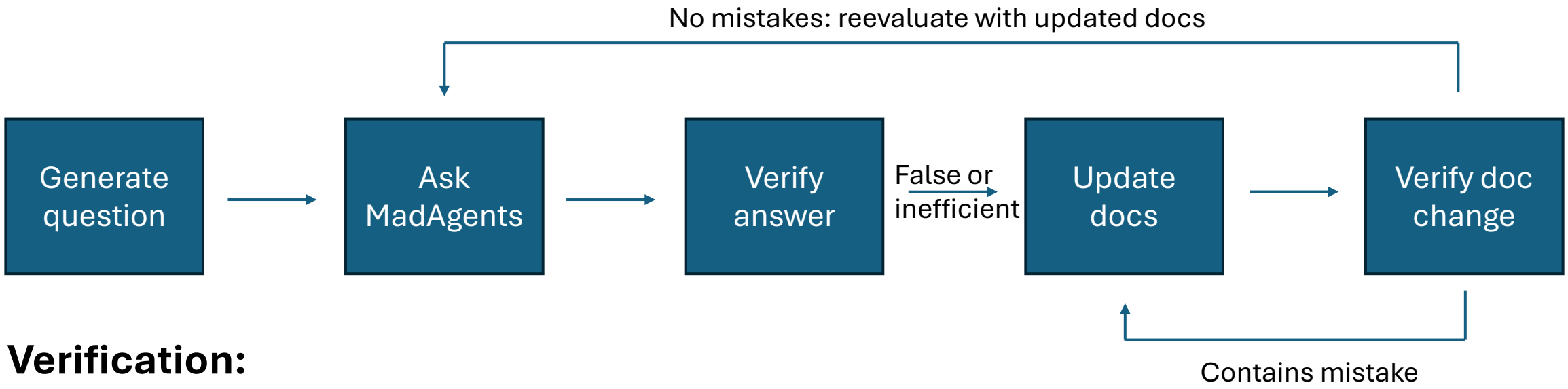
Verification:

1. Split into a list of claims
2. Verify each claim with unambiguous evidence

Forces a **frontier model** to **reason** and **explore deeply**!

Self-Improvement

- Extend the **internal documentation** via **self-improvement**



Verification:

1. Split into a list of claims
2. Verify each claim with unambiguous evidence

Forces a **frontier model** to **reason** and **explore deeply**!

Other work: LLMs for numerical data

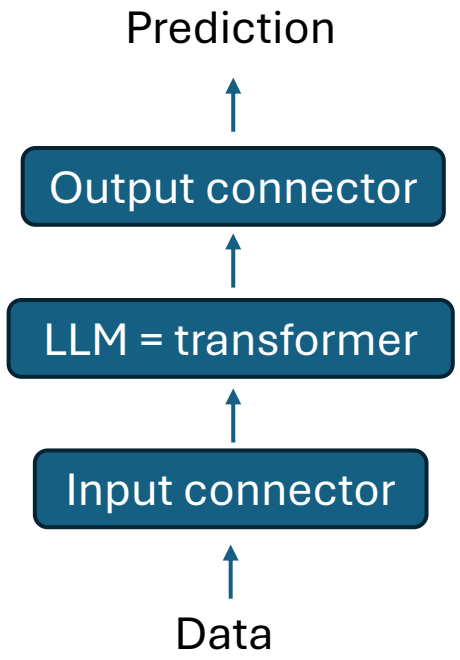
LLM: transformer pretrained on **text**

Other work: LLMs for numerical data

LLM: transformer pretrained on **text**
Can **LLMs** also be used as **foundation models** for **physics data** [2506.14757]

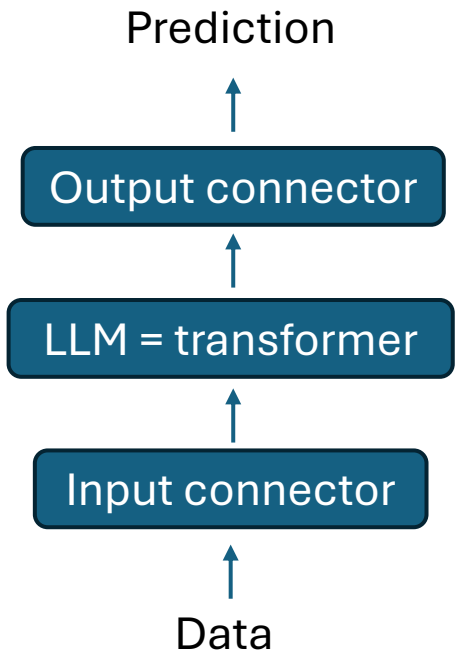
Other work: LLMs for numerical data

LLM: transformer pretrained on **text**
Can **LLMs** also be used as **foundation models** for **physics data** [2506.14757]

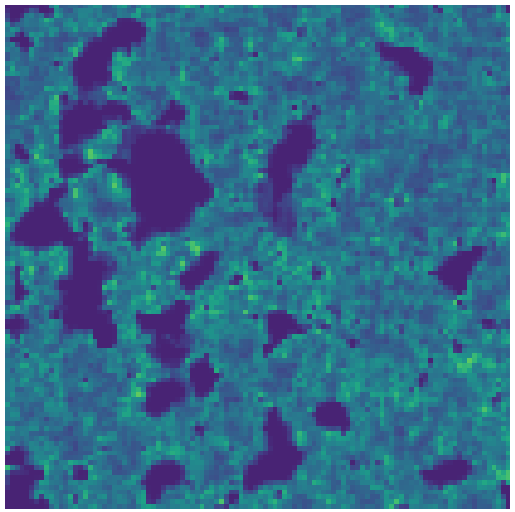


Other work: LLMs for numerical data

LLM: transformer pretrained on **text**
Can **LLMs** also be used as **foundation models** for **physics data** [2506.14757]

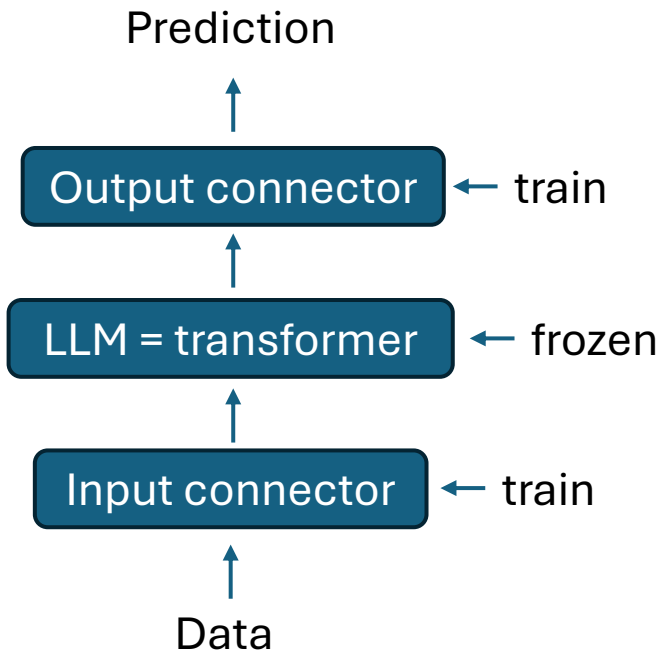


2d cosmological data

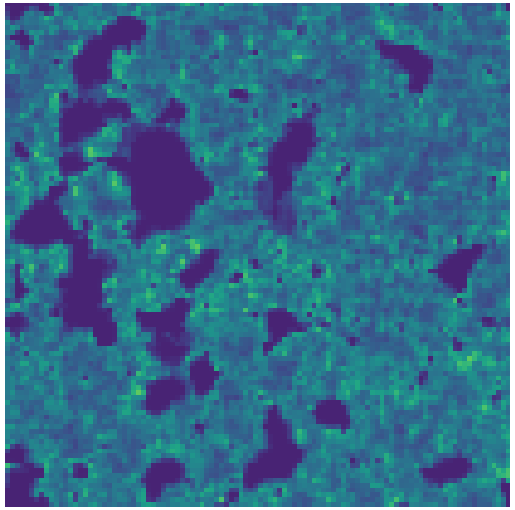


Other work: LLMs for numerical data

LLM: transformer pretrained on **text**
Can **LLMs** also be used as **foundation models** for **physics data** [2506.14757]

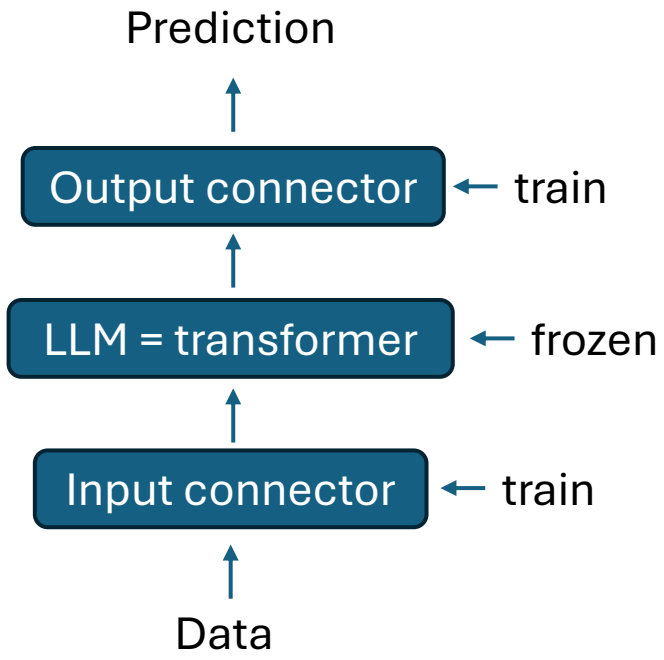


2d cosmological data

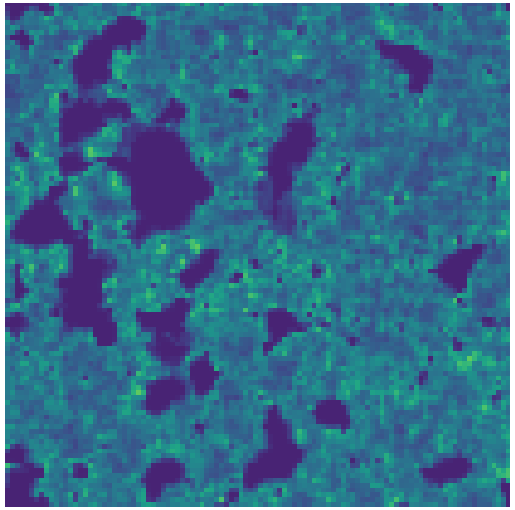


Other work: LLMs for numerical data

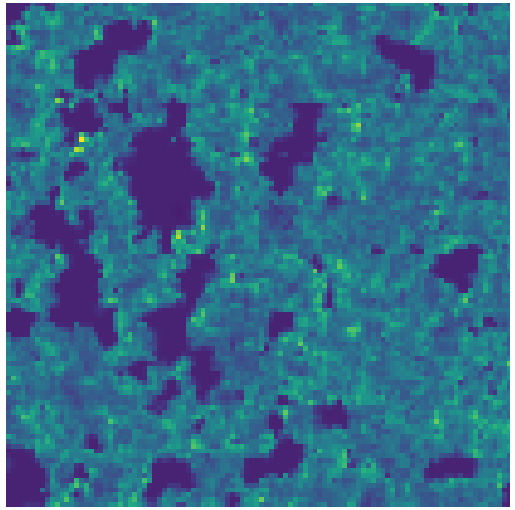
LLM: transformer pretrained on **text**
Can **LLMs** also be used as **foundation models** for **physics data** [2506.14757]



2d cosmological data

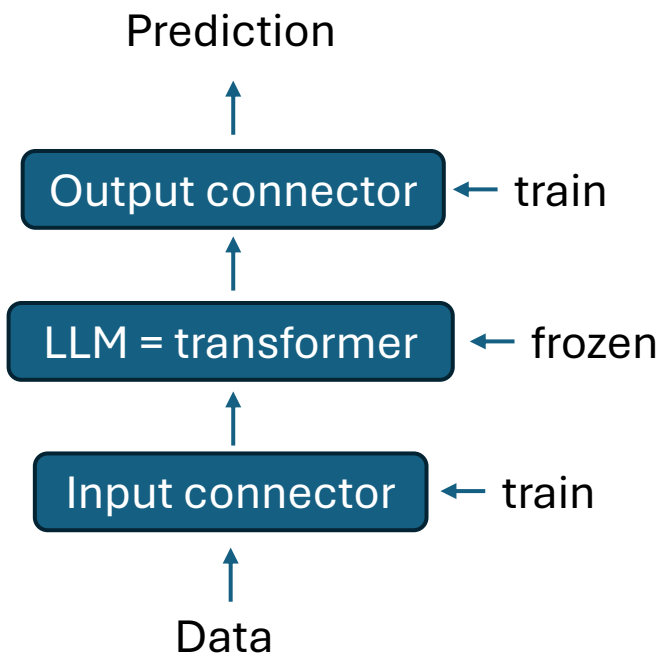


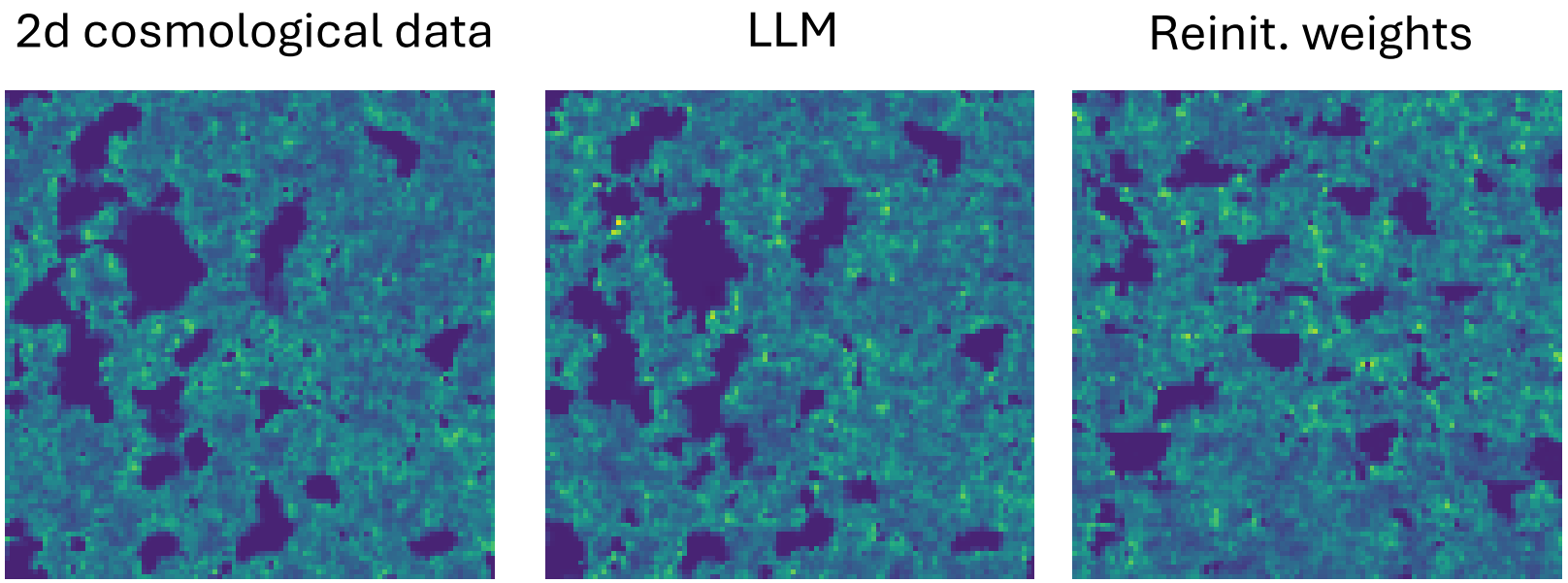
LLM



Other work: LLMs for numerical data

LLM: transformer pretrained on **text**
Can **LLMs** also be used as **foundation models** for **physics data** [2506.14757]





Conclusion



MadAgents: **smart interface** for **MadGraph**

- **Teaching** MadGraph
- Provide **expert support**
- Increase the **speed of scientific progress**

Conclusion



MadAgents: **smart interface** for **MadGraph**

- **Teaching** MadGraph
- Provide **expert support**
- Increase the **speed of scientific progress**

Paper: arXiv:2601.21015
Code Repo: github.com/MadGraphTeam/MadAgents
A **Claude Code** implementation exists!
Soon: **CERN** integration