

## PhD position

### “Unveiling the faintest HESS gamma-ray sources with an AI-based data analysis”

#### CALL

The University of Paris Cité calls for applications for a PhD position in Data Intensive Astroparticle Physics to work under the supervision of Prof. Yvonne Becherini. The candidate will develop a new analysis scheme for Imaging Atmospheric Cherenkov Telescopes (IACTs), implemented for ground-based detection of ( $E > 50$  GeV) cosmic gamma-rays using Deep Learning techniques. The new analysis strategy is expected to be faster and more sensitive than existing analysis schemes and will be used to analyse data from extragalactic sources, especially from those sources whose detection and characterization require long observing times. The analysis scheme will be developed in the framework of the HESS project. The research project will benefit both from the Astroparticule et Cosmologie laboratory (APC) and from the Data Intelligence Institute of Paris (diiP) environments.

**Subject field of the position:** Physics with specialization in Data Intensive Astroparticle Physics

**Placement:** [University of Paris Cité](#), [Astroparticule et Cosmologie laboratory \(APC\)](#) and [Data Intelligence Institute of Paris \(diiP\)](#)

**Extent:** 100%

**Duration of appointment:** 3 years

**Research project title:** Unveiling the faintest HESS gamma-ray sources with an AI-based data analysis

**Data access:** [HESS](#) simulations and data

**Doctoral school:** [STEP'UP \(Earth and Environment Science and Physics of the Universe in Paris\)](#)

#### CONTEXT AND GOAL OF THE RESEARCH PROJECT

The [High Energy Stereoscopic System](#) (HESS) telescopes in Namibia have helped to unveil and to understand the most energetic phenomena in the Universe since 2002. Before the operation of HESS, we knew only a handful of sources capable of emitting very-high-energy (VHE  $> 100$  GeV) gamma rays. But after more than 20 years of operation of HESS (plus the [Magic](#) and [Veritas](#) experiments), the number of cosmic accelerators detected is approaching 300. HESS is the only array of telescopes operating in the Southern Hemisphere, and will remain the only one operating until 2028. In 2028, it is expected that telescopes of the southern part of the [Cherenkov Telescope Array](#) (CTA-South) will start taking data and, and therefore, that CTA will start to be competitive with the sensitivity of HESS.

For this reason, the HESS steering committee is currently discussing the possibility to prolong the operation of the telescopes until 2028 and a decision is expected soon, with first discussions between funding institutions before the end of 2023. The possibility of the prolongation of HESS opens new perspectives in the coming years concerning the development of more sensitive analyses, with the aim of getting the most out of the already very successful experience. A termination of the HESS operations in 2024 or 2025 would not be optimal from a scientific perspective, because: 1) it would leave the Southern VHE sky completely uncovered in this energy band, with the possibility of missing interesting transient events and of missing the participation to global observation campaigns with other telescopes, and 2) it would completely miss the possibility to simultaneously cover the Southern sky together with the [Fermi](#) telescope (covering the 100 MeV-300 GeV range, started operations in 2008) which is still in operation, but it is unclear until when. Probably, in 5 years, Fermi will not be in operation when CTA-South will start to take data, and it is still uncertain about what kind of mission will be substituting Fermi in the future.

Even if a prolongation of HESS is highly probable for all the reasons mentioned above, one should be ready to accept also a different scenario, where the funding agencies cease to



support the project in 2025. The current proposal would also fit very well that scenario, because the data analysis scheme proposed in this application would be applied to all HESS data from 2002 until the shutdown with the goal of revisiting all the undetected targets, and to release more precise measurements on the spectra, the light curves and the morphology of the already-published datasets.

### ***HESS needs a second sensitivity improvement by analysis***

A first need for a sensitivity improvement was already deeply felt by the HESS collaboration already in 2008. In 2008, all the brightest gamma-ray sources visible by HESS had already been detected in the Galaxy and beyond, therefore, in order to try to increase the sensitivity of the weaker sources with no hardware upgrade, new, more sensitive, analysis strategies had to be developed. A number of analysis schemes have been developed and published at that stage [1, 2, 3, 4] opening the possibility to publish important detections and results in all papers published after 2010.

Until 2012, HESS has consisted of four 15-m spherical-mirror telescopes. But with the advent of the second phase of HESS (called HESS-II, started in 2012), where a 30-m telescope had been added at the centre of the 4-telescope HESS array with the goal of lowering the energy threshold, a new era in the analysis development had started. The fifth telescope in the middle of the array opened a new visibility window in the energy range (30 GeV-150GeV) by exploiting 1) the large telescope alone (“Mono observations/analysis”), or/and 2) the large telescope together with the previously-existing 4 (“Hybrid observations/ analysis”). The Mono analysis has given already a number of results, while no observations have yet been published with the Hybrid analysis.

HESS has observed a number of sources where no signal of gamma-rays could be found, probably given the limited sensitivity. The HESS sensitivity can be limited by two main factors: 1) the gamma-ray sources can have very soft spectra, with the bulk of the flux at the lowest energies reachable by HESS (i.e. low energy threshold needed), or 2) the source flux is extremely low, of the order than a handful of gamma-ray events per 28-minute observation, making it challenging to find those particular events in a “sea” of proton-induced events (i.e. a gamma-ray identification problem).

Given the probable prolongation of the HESS operations up to 2028, a second important opportunity for a final boost in the analysis is now, opening the possibility to revisit the old observations and to fully exploit the new coming data, with the added value of refining the observation strategies, given the upgraded sensitivity.

*The main motivation of this proposal is to increase the sensitivity of HESS by a factor of two in the 4-telescope analysis and optimize the Hybrid (5-tel) analysis in order to revisit the non-detections obtained in the past on several targets, some examples in the extragalactic field being: the composite and near Seyfert-starburst galaxies (e.g. NGC 1068, NGC 4945), the ultra-luminous infrared galaxies (e.g. the “merger” Arp 220), the Fields-of-View of the IceCube neutrino alerts, etc. and to be ready for the probable new 5-year observations with a renewed analysis sensitivity.*

Improving the sensitivity by a factor of 2, needs an enhancement of the current angular resolution of HESS and a further suppression capability of the proton background, by also giving the possibility, at the same time, to speed-up significantly the processing time of the analysis pipeline from the calibrated data to the gamma-ray signal extraction, opening the possibility to perform complete re-analysis of 20 years of observations in just a few hours.

### ***A new analysis strategy for HESS based on Artificial Intelligence***

The new analysis strategy, to be developed in this context, takes advantage of the latest developments in Artificial Intelligence and of the close cooperation with the [Data Intelligence Institute of Paris \(diiP\)](#) and the [Data Intensive Sciences and Applications \(DISA\)](#) centre. The analysis procedure proposed contains several innovative approaches: an AI filter, a Deep Learning shower reconstruction technique and a final signal extraction procedure.



1. The AI filter, based on *Unsupervised Graph Neural Network Clustering*, optimized on real data, aims to reduce the background contamination from proton showers just after the calibration procedures, with the goal of minimizing the number of gamma-like events to be considered as input in the following shower reconstruction procedure (goal: speed-up of analysis). Ideally, only 10-20% of the whole dataset should be the input of the subsequent shower reconstruction step, but even a 50% reduction from the whole sample will be considered acceptable. The filtering approach will also use ultra-modern data visualization techniques as [T-SNE](#), [UMAP](#), [PCA](#) and data clustering algorithms as [hdbscan](#).
2. The Deep Learning shower reconstruction method will be based on *Supervised Graph Neural Network Regression Techniques*, where the supervision is coming from Monte Carlo simulations of gamma-rays. The main goal at this step is to ensure an optimal reconstruction of the arrival direction and of the energy of the gamma-ray-like events.
3. Since we expect the sample to be still contaminated by proton-induced showers even after steps 1. and 2., a final signal-over-background separation procedure based on *Supervised Graph Neural Network Classification Techniques* can be implemented in order to increase the identification power of gamma-ray events.

The goal would be to dedicate one year to the implementation and testing of the new strategy and to the necessary publication of the method in a refereed Journal, and the following time to the application of the analysis to the HESS datasets. Several publications, with the results obtained with the new method on the HESS datasets, are expected.

#### **DESCRIPTION OF GROUP/LABORATORY/SUPERVISION**

This cross-disciplinary project in Physics and Computer Science is developed within the diiP and the [APC](#) research environments. In order to best exploit all the existing and near-future Machine Learning architectures, the proposed project will be supported by high-expertise Computer Scientists from the diiP.

This PhD thesis will be supervised by Yvonne Becherini, Professor at the University of Paris Cité, and will take place within the High-Energy Astrophysics (AHE) group of the AstroParticule and Cosmologie Laboratory (APC) and the Data Intelligence Institute of Paris (diiP). The APC is an ideal laboratory for carrying out such a research project, as the lab participates and therefore has access to the data of several VHE observatories. The PhD student will become a member of the HESS collaboration.

#### **PROPOSED WORK**

- Deep Learning for IACTs, with a focus on faint extragalactic sources
- Active participation in proposals for, and decisions on HESS observation campaigns
- Analysis based on Python/C++ programming
- Writing of scientific articles
- Oral presentations at national and international workshops/conferences

#### **DUTIES**

- Attend doctoral school courses for a total of 15 Academic credits, more information may be found at this address: <https://ed560.ed.univ-paris-diderot.fr/en/rules-for-training/>
- Work on the research subject proposed in this document
- Regularly presentations of intermediate research results to the supervisor
- Active participation in the HESS Collaboration, with responsibility to be undertaken on a technical aspect of data analysis and/or data calibration
- Work in close collaboration with the other project members in an interdisciplinary research environment as well as with domain experts
- Presentation and publishing of intermediate results in conference proceedings
- Presentation and publishing of more mature results in journal articles
- Preparation of the thesis manuscript
- Participation to the annual "Congrès des Doctorants"



## TRAINING AND SKILLS REQUIRED

- Master in Astronomy and Astrophysics or Master in Astroparticle Physics
- Ability to work in a team
- Python/C++ programming
- Good command of English

## ACQUIRED SKILLS

Various skills acquired and developed during this PhD thesis will be valuable and transferable to other fields: data analysis at different wavelengths, numerical simulations, data processing, data analysis, machine learning, writing of articles and of observation proposals, teamwork, oral presentations at national and international workshops and conferences.

## ASSESSMENT CRITERIA

The selection of candidates is made with regard to the applicant's ability to successfully complete and benefit from their studies at the graduate level. The assessment takes into account academic skills documented in scientific works, especially focused on the quality of the essays at the undergraduate level, any advanced work and other scientific or scholarly works. The assessment also takes into account breadth and composition of the undergraduate degree.

The successful candidate has excellent analytical and problem-solving skills, and is a committed researcher with a drive for excellence. Prior research experience concerning the subject is a significant advantage. Excellent written and oral communication skills in English are essential to publish and present results at international conferences and in international journals. Advanced skills in computing are a key requirement, as all activities are carried out in Linux/Unix environments and using the Python programming language. Interpersonal skills and flexibility are of key importance since the work is done in a research group.

## REQUIRED DOCUMENTS

A cover letter, a CV, links to the Master thesis and previous works, and contact information of two referees in only one file to [yvonne.becherini@apc.in2p3.fr](mailto:yvonne.becherini@apc.in2p3.fr).

## REFERENCES

1. [\*A new analysis strategy for detection of faint gamma-ray sources with Imaging Atmospheric Cherenkov Telescopes\*, Y. Becherini et al., \*Astroparticle Physics\*, 2011](#)
2.  [\*\$\gamma\$ /hadron separation in very-high-energy-ray astronomy using a multivariate analysis method\*, S. Ohm et al., \*Astroparticle Physics\*, 2009](#)
3. [\*A high performance likelihood reconstruction of gamma-rays for Imaging Atmospheric Cherenkov Telescopes\*, M. De Naurois & L. Rolland, \*Astroparticle Physics\*, 2009](#)
4. [\*A Monte Carlo Template based analysis for Air-Cherenkov Arrays\*, R.D. Parsons & J.A. Hinton, \*Astroparticle Physics\*, 2014](#)
5. [\*Signal extraction in atmospheric shower arrays designed for 200 GeV – 50 TeV gamma-ray astronomy\*, M.Senniappan et al., \*JINST\*, 2021](#)