LNS GENESIS Proposals:

March 31st Workshop Talks
Lina Necib

Charting the Inner Darkness: AI-Driven Dark Matter Density Profiles Across Cosmic Scales

Focus Area: 14A — Foundation Models of Particle Interactions and Cosmic Physics (HEP, NP) **[Most natural one]**

Focus Area: 14C — Expedited Discovery from High Complexity and Petabyte-Scale Datasets (HEP, NP) **[Also for consideration]**

MIT (IHE — Lead Institution): Lina Necib (PI)

DOE/NNSA Lab (Category 1 — Required):

Alex Drilca-Wagner, Fermilab, Confirmed

Industry (Category 2 — Required): None

Project Overview

Science Vision: Measure inner dark matter density slopes as a function of halo mass by training ML models on cosmological simulations and applying them to multi-survey observational data. These measurements directly test CDM vs. SIDM predictions and constrain signals for indirect detection experiments.

Why AI is essential: Traditional Jeans-equation and rotation-curve methods assume axisymmetry and equilibrium, which real galaxies violate.

- [Dwarfs:] Our GNN (GraphNPE) model (Nguyen et al. 2023, 2025) tested on dwarf galaxies outperforms traditional Jeans analysis with fewer stars.
- [External Galaxies:] Currently testing a diffusion-model methods trained on realistic, disequilibrium-inclusive simulations (FIRE, DREAMS) to recover density profiles from galactic images. This is not possible at all without AI.
- [Milky Way:] Build a new framework to obtain the density profile of the Milky Way

AI enables inference from sparse data (as few as 100 stars in the dwarf galaxy systems) and scales to thousands of galaxies from imaging surveys.

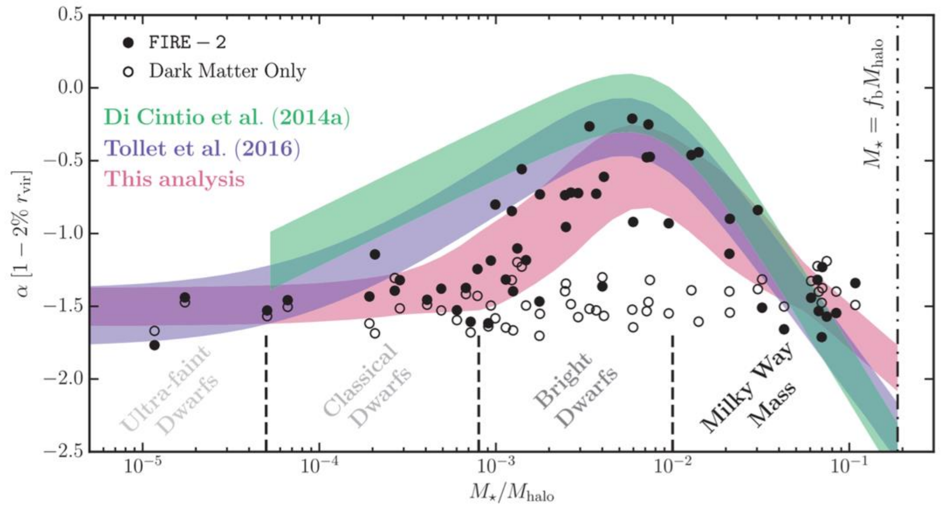

Figure 1. Figure from Lazar et al. (2020), which compares the inner slope of the DM density profiles α, defined in the range of $[1\% - 2\%]r_{vir}$ as a function of the stellar mass to halo mass **in simulations**. The PI will work on obtaining **robust measurements** on this theoretical figure, combining stellar data, machine learning, and cosmological simulations. The open circles are simulations of Dark Matter only, which follow an NFW profile (Navarro et al. 1997). The NFW profile corresponds to $\alpha \sim -1.5$ in this definition of the slope. The dark circles are hydrodynamic simulations from the FIRE-2 physics (Hopkins et al. 2018), showing that stellar feedback is most efficient at the bright dwarf mass range, in which in the inner profile gets close to a core ($\alpha \sim 0$).

Project Overview

14A Alignment: Synthesizes multiple data modalities (spectroscopic kinematics from PFS/DEIMOS, phase-space from Gaia, photometric images from Euclid/LSST, rotation curves from JWST) with simulation-trained ML to address dark matter — a core question named in 14A. Sparse-data inference and multi-survey fusion are technical challenges highlighted in the focus area.

14C Alignment: 14C targets AI methods for extracting robust scientific insight from high-complexity, petabyte-scale datasets. Euclid DR1 alone will contain images of thousands of dwarf galaxies; LSST will produce ~20 TB/night of imaging data; Gaia DR4 delivers 6D phase-space for billions of stars. Our simulation-trained ML methods (GraphNPE, diffusion models, CNNs) provide uncertainty-aware posterior inference that directly connects these observational datasets to theoretical DM parameters — exactly the uncertainty-aware reasoning that 14C calls for. The domain-shift challenge (training on simulations, deploying on real survey data) is a core complexity problem highlighted in the focus area.

Data Pipeline:

<u>Training:</u> generated idealized simulations for dwarfs, FIRE, DREAMS, Illustris-TNG simulations (CDM, SIDM, WDM) with synthetic observations being produced via GalaxyGenius and Synthesizer.

<u>Inference:</u> DEIMOS/PFS spectroscopy, Gaia DR4, Euclid Q1+DR1, LSST DR1, JWST archival rotation curves. All training data publicly available (except PFS, which has a proprietary period of 18 months, though some of it will be public in Fall 2026); pipelines and mock catalogs will be released.

9-months work plan

Months 1–3 (Jul–Sep 2026): Data Curation and Baseline Infrastructure

- [Dwarfs:] Test robustness to measurement errors and selection function (the pipeline was already tested on simulations). Establish baseline Jeans-method performance on these mocks for head-to-head AI advantage comparison.
- [External galaxies:] Build unified training dataset from FIRE, DREAMS, and Illustris-TNG simulations under CDM, SIDM, and WDM physics. Generate synthetic observations via GalaxyGenius with survey-specific instrumental effects (Euclid, LSST).
- [Milky Way:] Build a non parametric density profile estimate of the Milky Way inner profile.
- **Months 3–6 (Oct–Dec 2026): AI Advantage Demonstration [Go/No-Go Review]**
- [Dwarfs:] Deploy GraphNPE on dwarf galaxy mocks with realistic selection functions and noise; quantify inner-slope recovery vs. Jeans baseline.
- [Dwarfs:] Apply GraphNPE to two example of galaxies.
- [External galaxies:] Train diffusion model prototype on DREAMS phase-space data for inner galaxy profile from external galactic images.
- [Milky Way:] Test Milky Way pipeline on Gaia DR3, and prep for Gaia DR4.
- Present AI advantage metrics at 6-month go/no-go review.

Months 6–9 (Jan–Mar 2027): Gaia DR4 Application and Phase II Roadmap

- [Dwarfs:] Apply GraphNPE to more galaxies, study the distributions of density profiles across (DESI, PFS, Deimos).
- [Milky Way:] Gaia DR4 releases Dec 2026, delivering 6D phase-space for billions of stars. Apply validated diffusion model to DR4 data for inner Milky Way DM profile, first measurement that accounts for bar-driven disequilibrium. The bulk of this work will be moved to Phase II.
- [External galaxies:] Test developed methods on LSST DR1 and Euclid DR1.
- [External galaxies:] Evaluate scaling trajectory for Phase II deployment to full Euclid/LSST catalogs (thousands of galaxies).

9-months deliverables & Metrics

Deliverables

- D1: [External Galaxies:] Curated multi-simulation training dataset (FIRE + DREAMS + TNG) with synthetic observations across survey configurations, publicly released
- D2: [Dwarfs:] GraphNPE applied to real dwarfs — first DM inner-slope measurements with full posterior uncertainties for Milky Way satellites from GraphNPE
- D3: [Milky Way:] Pipeline for the density profile of the inner Milky Way DM profile from stellar phase-space, validated on simulations.

Go/No-Go Metrics for AI Advantage

- M1: [Dwarfs:] Inner-slope recovery accuracy: GraphNPE achieves lower bias and tighter posteriors than Jeans methods on matched mock datasets, quantified by fractional error
- M2: [Dwarfs:] Sparse-data robustness: accurate profile recovery with 100 tracers where Jeans methods require 500+
- M3: [Milky Way:] Disequilibrium resilience: maintained accuracy of the Milky Way density profile accuracy on simulations.
- M4: [External Galaxies:] Scaling trajectory: demonstrated path from single-galaxy inference to batch processing of 100+ galaxy images for Phase II survey-scale deployment

Workforce Development Plan

Graduate Students and Postdocs

- Phase I supports 1 graduate student and partial postdoc effort, trained at the intersection of ML methods and dark matter physics.

Undergraduate Researchers

- MSRP students (Summer: 10 weeks) working on mock data stress-testing, Jeans-method benchmarking, and data pipeline construction.

IAIFI Integration

- PI is a senior investigator at IAIFI, enabling cross-pollination with the broader AI-for-physics community. Project results and methods will be disseminated through IAIFI workshops, summer schools, and public code releases.

Open Science

- All training pipelines, mock catalogs, and trained model weights will be publicly released to maximize community impact and reproducibility.