Syllabus – Data Science and Machine Learning

Master's in Physics with a focus on Fundamental Physics

Instructor: Dr. Hernán Andrés Morales-Navarrete

An intensive course designed to bridge theoretical physics with practical data science applications for graduate-level physics research.

Course Overview

This intensive master's-level course combines theoretical foundations with hands-on applications of data science in physics research contexts.

1 Credits

4 credits (48h class, 48h practical, 96h self-study)

2 Duration

16 weeks of intensive in-person instruction (3 hours/week)

3 Language

Python-based implementation with scientific computing libraries

All course materials will be provided through our learning management system, with supplementary resources available through the university library's digital collections.

Learning Outcomes

Upon successful completion of this course, students will be able to:

Statistical Methods

Apply advanced statistical methods to complex physics problems requiring nuanced data analysis

Data Analysis

Analyze time-series and scientific images using specialized computational techniques

Simulation Techniques

Implement Fourier analysis and Monte Carlo simulations for physical systems

AI Foundations

Understand core AI and ML principles with specific application to physics challenges

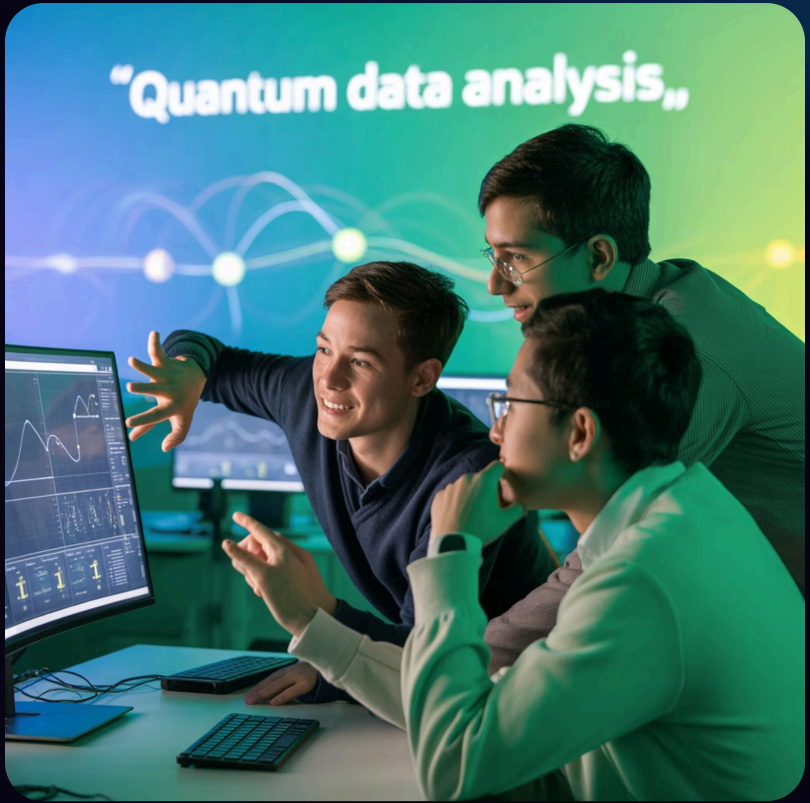
Applied Machine Learning

Apply machine learning algorithms to solve real-world physical problems

Scientific Publishing

Write publication-quality scientific papers at the intersection of ML and Physics

Methodology



This course employs an immersive project-based learning approach that mirrors actual research environments in computational physics.

Project-Based Learning

Students develop solutions to real physics problems using data science techniques

Collaborative Teams

Work in pairs to simulate research partnerships and peer review processes

Publication Focus

Produce two collaborative scientific papers of publication quality

Weekly Tasks

Regular presentations and progress reports to develop communication skills

📅 Weekly Plan

The course progresses from foundational concepts to advanced applications, culminating in research-quality outputs.

Week	Topic	Description	Paper Link
1	Introduction	Course overview, Python review	-
2	Dimensionality Reduction	PCA, SVD, UMAP techniques	Paper 1 - M1
5	Neural Networks	Fundamentals, backpropagation	Paper 1 - M3
7	PINNs	PDEs and boundary conditions	Paper 2 - A2
10	Generative Models	VAEs, GANs for physics	Paper 2 - A4
14-16	Project Completion	Final analysis and documentation	Final Submissions

The complete 16-week schedule with detailed readings will be provided in the course materials.

Key Deliverables

The course is structured around progressive deliverables that build toward publication-quality research papers.

1. **Week 4**

   Project Proposal (2 pages)

   Define research question, methodology, and anticipated results. Includes literature review and preliminary data exploration.

2. **Week 8**

   Midterm Report (4-5 pages)

   Present preliminary results, refined methods, and address challenges encountered. Includes peer feedback session.

3. **Week 10**

   Draft Figures & Outlines

   Complete data visualization suite and detailed paper structure. Includes caption writing and methodological documentation.

4. **Week 16**

   Final Submission (arXiv-ready)

   Complete research paper formatted to publication standards with comprehensive supplementary materials and code repository.

🧩 Project Groups

Paper 1 – Methods

Group 1

Autoencoders for Dimensionality Reduction in Physical Systems

Group 2

Performance Metrics and Validation Methods for Physics ML Models

Group 3

Generative Models for Simulating Physical Phenomena

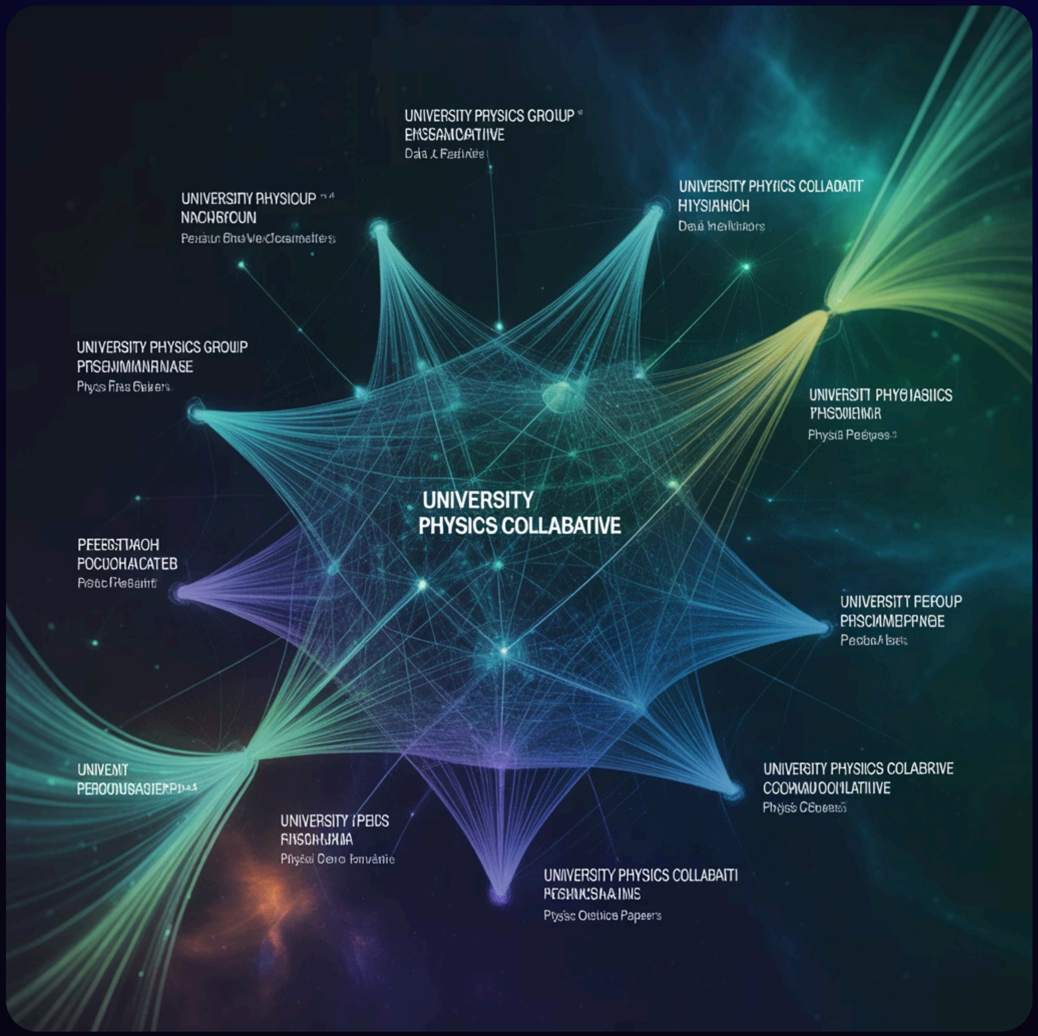
Paper 2 – Applications

Group 4

Image-based Physics Analysis for Experimental Data

Group 5

Physics-Informed Neural Networks for PDE Modeling



Groups will collaborate across papers to ensure methodological consistency and application relevance.

Evaluation Criteria

Assessment is designed to evaluate both technical proficiency and scientific communication skills, mirroring the evaluation process for actual research publications.

20%

Weekly Participation

Active engagement in discussions, code reviews, and collaborative problem-solving sessions

30%

Deliverables

Quality and timeliness of project milestones, including proposals, reports, and code documentation

20%

Presentations

Clarity, technical accuracy, and effectiveness of regular progress presentations

30%

Final Paper

Scientific rigor, originality, and publication readiness of the final research paper

All assessment components include both individual and team evaluation elements to ensure comprehensive skills development.

Papers will undergo a simulated peer-review process mirroring actual scientific publishing procedures.

Tools & Libraries

Students will work with industry-standard tools used in both academic and industrial physics research settings.

Core Programming

- Python 3.x
- NumPy & SciPy
- Pandas & Matplotlib

ML Frameworks

- scikit-learn
- TensorFlow & Keras
- PyTorch
- DeepXDE for PINNs

Development

- GitHub for version control
- Jupyter Notebooks
- Google Colab for GPU access

Publication

- Overleaf (LaTeX)
- Markdown for documentation
- Zenodo for data archiving

Students will receive guided introductions to specialized physics libraries including QuTiP, ASE, and other domain-specific tools as needed for their projects.

Research Integration

Beyond the Classroom

This course is designed to directly contribute to the scientific community through authentic research experiences:

- Papers target actual scientific journals and arXiv submission
- Projects address open problems in computational physics
- Emphasis on reproducibility and open science practices
- Integration with departmental research initiatives

> "The most effective learning happens when students engage in genuine scientific inquiry rather than simulated exercises."

Ethical Considerations

Students will engage with critical issues including:

- Data ethics in scientific research
- Energy costs of large-scale simulations
- Transparency in ML/AI methodology
- Reproducibility crisis in computational science