Deeper Skies

The Intersection of AI and Cosmology at Fermilab

B. Nord
On behalf of A. Ciprijanovic, the Deep Skies Lab, the Cosmic Physics Center
Oct 30, 2023

Goals Today

Define AI and Review History

Describe Current Work at Fermilab in Cosmology and AI

Plan Vision, Collaboration, and Proposals

Connect People for Collaboration

Questions to consider during this talk

- What are the **major challenges** in your research?
  Consider modeling, hardware, operations, science.

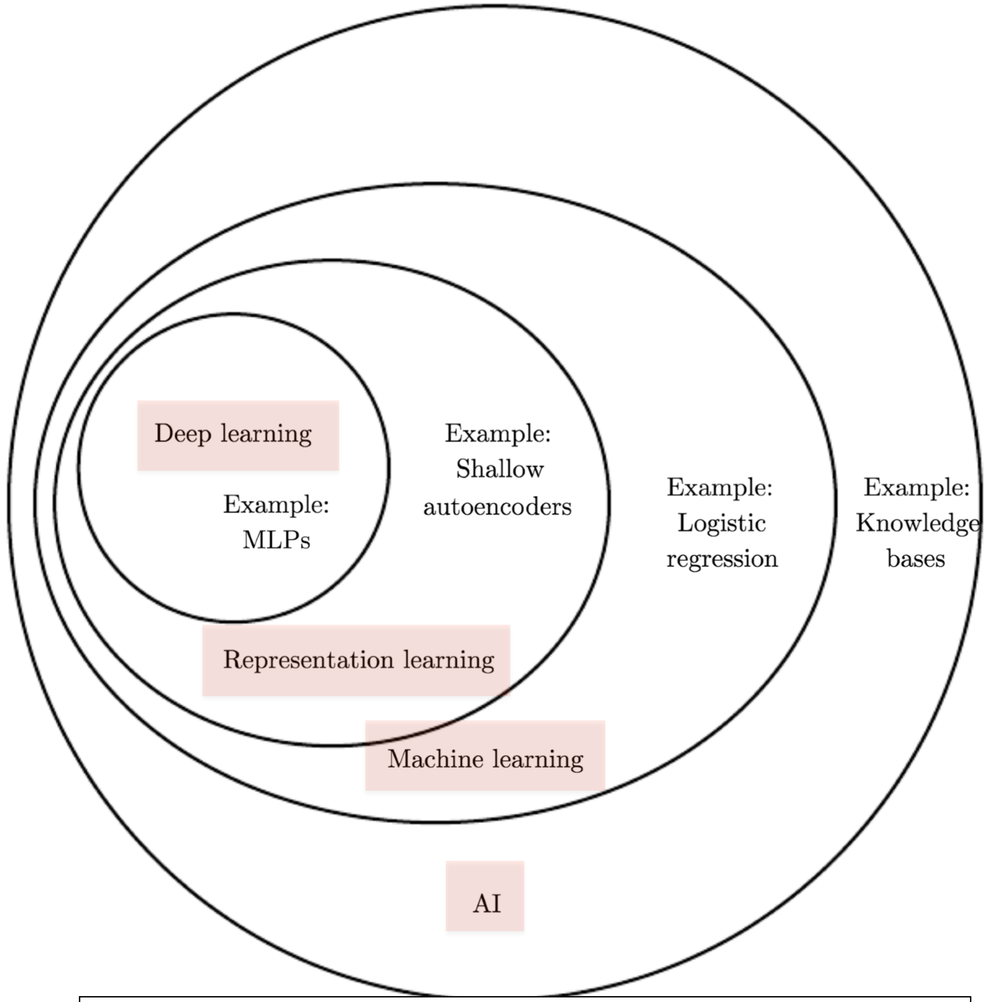
- What **kind of data** do you have?
  Consider type, amount, format.

- What **collaboration and information** do you seek to further consider AI?
  Consider papers, blogs, codes, people.

- Is AI for Cosmology **more** **science** **OR** **computer science?**
  It's an opportunity for both and for more.

What is AI?

- Definition
  - A **class of algorithms** used to build models driven primarily by the data
  - A fitting paradigm for large-parameter models
- Modern Synonyms
  - Data science
  - Machine learning
  - Deep learning
- It is not …
  - learning, understanding, or thinking.



deeplearningbook.org: Goodfellow, Bengio, Courville

AI: Benefits, Risks, and Opportunities

- Pro's
  - **Classification** of objects (and then go do science and stats)
  - **Imagining possibilities** for next stages of science applications
  - **Clarification** of statistical methods
- Con's
  - **Uninterpretable** and unexplainable (~uncertainty quantification poorly defined)
  - Requires **large** amounts of data and specific computing resource
  - **No concrete** statistical theoretical underpinning (~loss landscape is non-convex)
  - **Jargon is unsettled** and overly anthropomorphizing

~80 years of AI

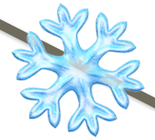
1943: McCulloch-Pitts **neuron**

1956: Dartmouth **conference**

1973: Lighthill **Report**

AI Winter 1

- Architectures not general.
- **Over-hype**

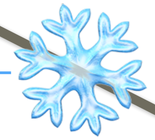


1980: Fukushima **neocognitron**

1988: LeCun's Network*

- Algorithms inefficient
- **Over-hype**

AI Winter 2



Current Summer

2012: **ImageNet** and CNNs

2015-17: **AI for Science begins**; DeepSkiesLab starts

2018: **Commercialization Permeates**

2022: **Generative AI and LLMs**

Funding from private and public sources is significant

Is AI for Cosmology ...

more about science OR computer science?

Both: the intersection offers major potential for advancement for both fields.

Funding: This view is also held by many program managers.

When did Fermilab get started?

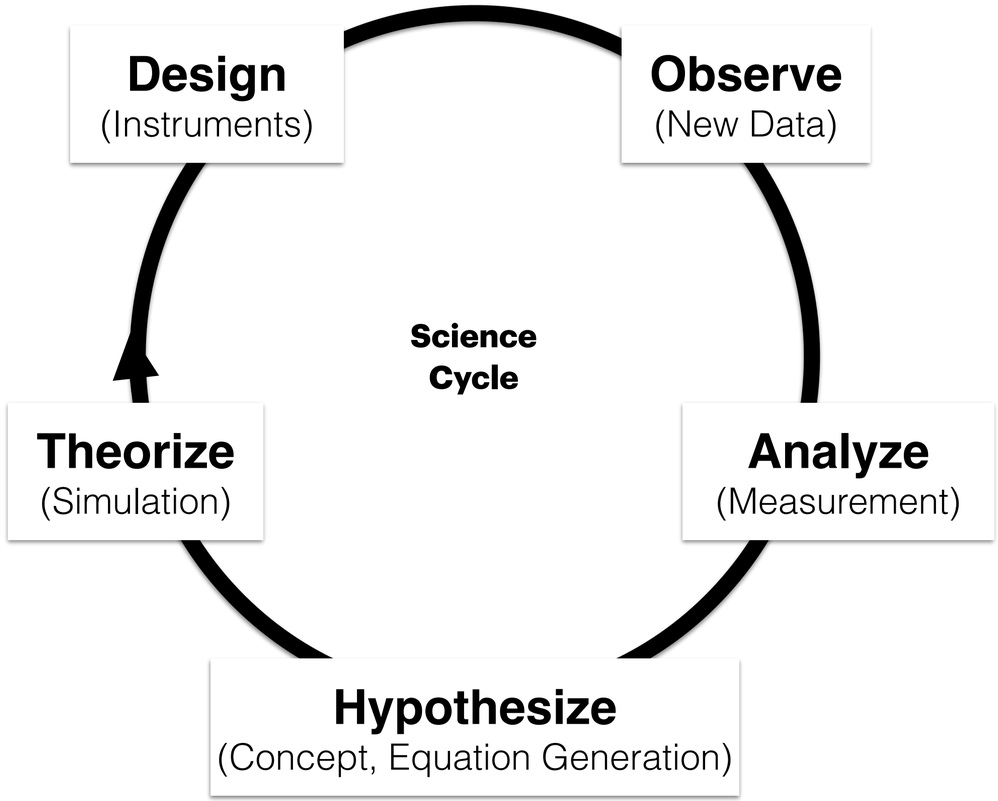
- **2016**: ***DeepLensing***
  Early applications of AI to strong lens classification (Nord, 2016)

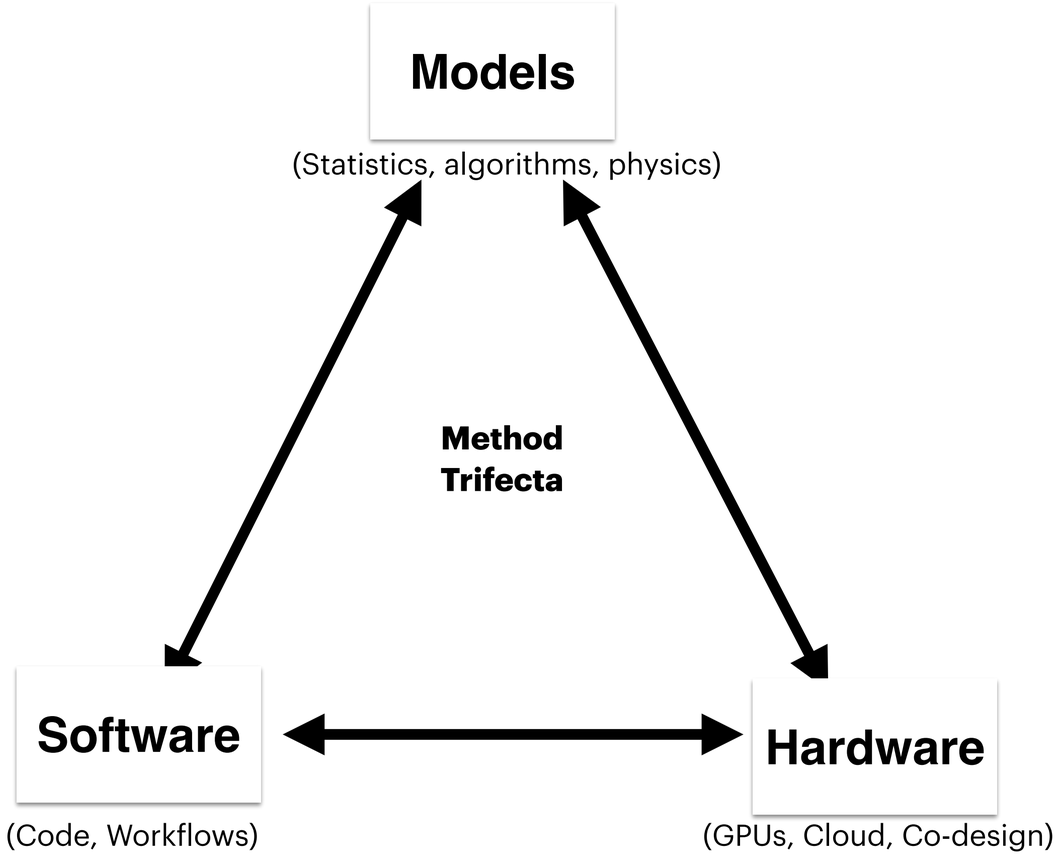
- **2017: Deep Skies lab Inaugurated**
  Multi-institutional, multi-generational, transnational community/collaboration of scholars in both cosmology and AI (Nord, Peek, Avestruz, 2017); full community now has >550 people.

- **2019: Hired Co-PI Ciprijanovic** (Wilson Fellow)
  Pursuing methods, applications, and community-building

Current Work in CPC

Scientific Cycle: Where can AI be used in Cosmology ?





Examples of experiments where this research could be applied

DES | LSST | JWST | DESI | SpecS5 | CMB-S4 | LIGO | FollowUp

DES | CMB-S4 | LSST | JWST | DESI

Analyze

Classification, Finding, Prediction, Inference

The earliest and still most common area of application for AI.

A. Drlica-Wagner

- Artifact (ghosts) detection, masking
- Low-surface-brightness galaxy finding, modeling
- Star-galaxy classification with tabular machine learning
- Strong lens identification
- Labeling data sets with Zooniverse and other human inspection tools

J. Frieman

- Supernova light curve calibration
- Strong lens galaxy inference

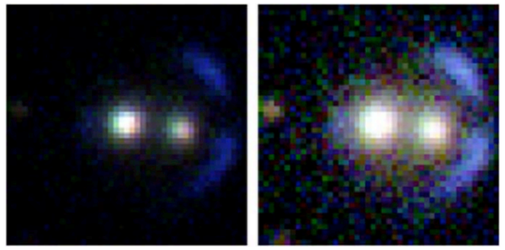
N. Gnedin

- Fit rotation curves of galaxies with double-axion models with Actor Advantage Critic (A2C)

Deep Skies Lab

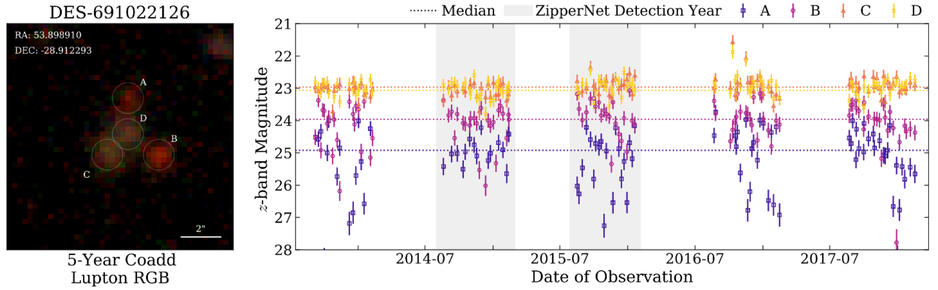
- Strong lens galaxy Finding/Inference
- Double source plane lens Finding
- Strong lens Cluster Inference
- Strong lens Cosmology Inference
- Lensed supernovae Finding
- Galaxy Merger Finding/Inference
- Galaxy Morphology Classification
- Galaxy Evolution Inference
- Large-scale structure Inference
- Axion Inference w/ lensing
- CMB de-lensing and r Inference
- Cluster (SZ/Optical/Xray) Mass Inference

Zabrowski + 2022



SL Candidates in DECam

Morgan + 2022



Lensed SNe in DES

DES | LIGO | LSST | SpecS5 | CMB-S4 | DESI

AI is faster than N-body, but not interpretable/trustworthy.

Theorize

Simulations, Surrogates, Generative Models, Mechanistic Models

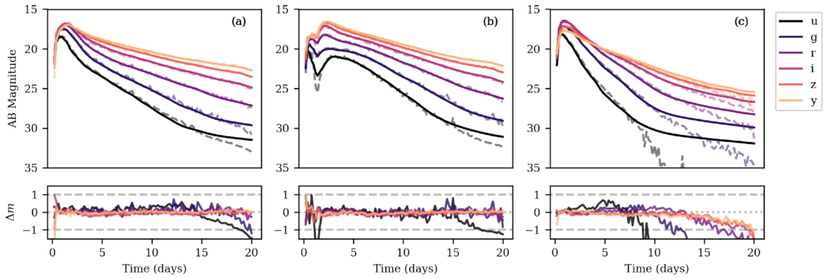
N. Gnedin

- Improve gas cooling and heating functions for cosmological simulations with XGBoost

Deep Skies Lab

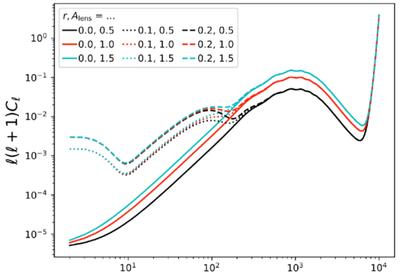
- Strong lens simulation: Galaxies, Clusters, Supernovae, Quasars, Fuzzy dark matter
- Kilonovae and Gravitational Waves
- Lensed CMB with B modes
- Clusters - SZ, Optical, Xray, Lensing
- Optical telescope surveys
- Toy physics models for benchmarking

Lukosiute + 2022



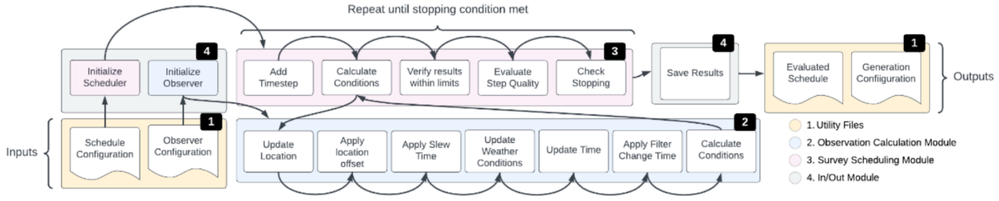
Kilonovae Light Curves Sims for Grav Waves

McDermott + 2023, in prep



Lensed CMB Spectra

Voetberg + 2023



Telescope Survey Simulator Workflow

LSST | DESI | JWST | SpecS5 | FollowUp | DES

This is resource allocation with multiple competing rewards and constraints.

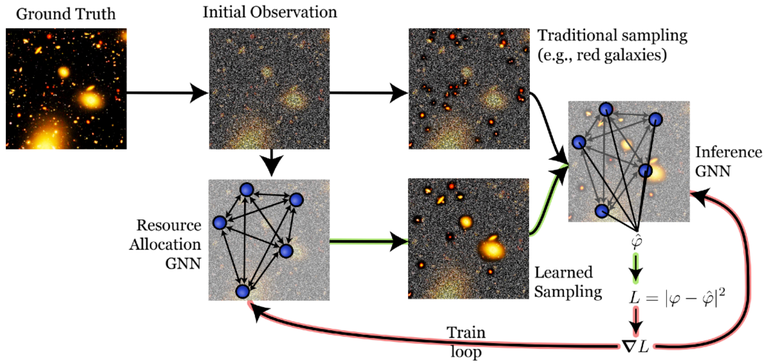
Observe

Telescope Control, Reinforcement Learning, Scheduling

Deep Skies Lab

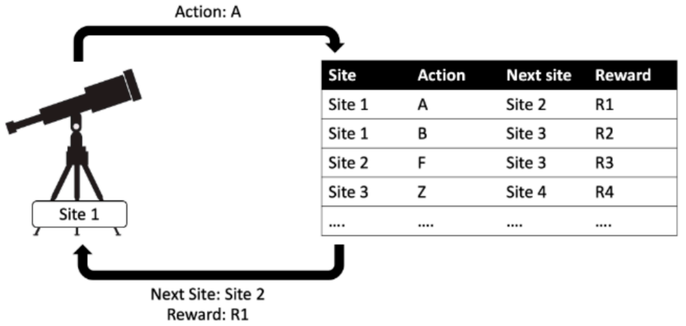
- Reinforcement Learning for Pre-Scheduling
- Reinforcement Learning for Adaptive Observing
- Scheduling and adaptive tests at educational telescope facilities
- Simulation software for pre-scheduling and adaptive Reinforcement Learning
- Unsupervised learning for spectroscopic fiber allocation

Cranmer + 2022



Unsupervised learning for spectroscopic fiber allocation

Terranova + 2023



Site	Action	Next site	Reward
Site 1	A	Site 2	R1
Site 1	B	Site 3	R2
Site 2	F	Site 3	R3
Site 3	Z	Site 4	R4
....	....	....	....

Reinforcement Learning for Pre-Scheduling

DES | SpecS5 | LSST | JWST | CMB-S4 | DESI

Design

Instruments, Hardware

There is very little literature in this area; could be open space for contribution/proposals

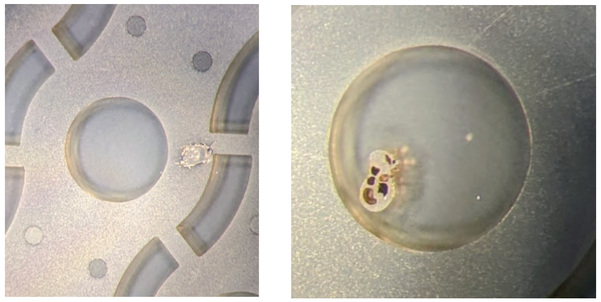
S. Simon

- Segmentation of instrument images to identify defects in silicon optical coupling wafers for CMB-S4

Deep Skies Lab

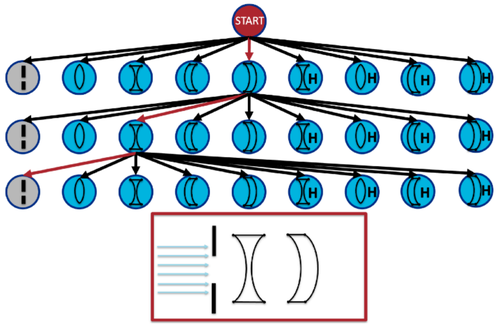
- Arrangement and tolerance-setting of optical elements for telescope design
- Design spectroscopic experiments for optimized data-taking (SPOKeS Pipeline)

Simon and Team



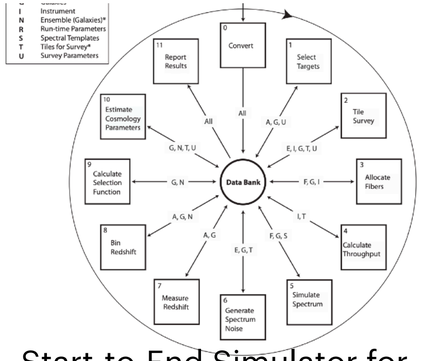
Wafer Defect Detection

Cohen+2023, in prep.



Simulation-based Inference for optics design

Nord+2016



Start-to-End Simulator for Spectroscopic Survey

DES | SpecS5 | LSST | JWST | CMB-S4 | DESI | LIGO

Hypothesize

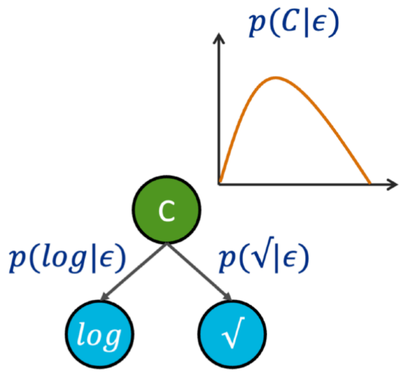
Symbolic Regression, Semantic Concept Generation, Interpretation

The most challenging task to automate, but perhaps the most time-saving.

Deep Skies Lab

- Generate equations from data
- Identify concept, links and paths in the literature to predict new ideas
- Mutual Information for neural network parameter interpretation

Cohen+2023, in prep.



Automated symbolic equation generation

Critical approaches for enabling trustworthiness of AI.

Methods

Development of Foundational AI and Statistics

Domain Adaptation

- Model generalization when training on simulations and predicting on real data

Hierarchical Bayes

- Inference of population-level or summary statistics from large data sets

Simulation-based Inference

- Fast inference with robust error bars.

Uncertainty Quantification

- Determining how to estimate error bars with neural networks

Benchmarking Classification

- Comparing neural network architectures for strong lens classification (LSST Data Science Internships for Undergraduates, 2023)

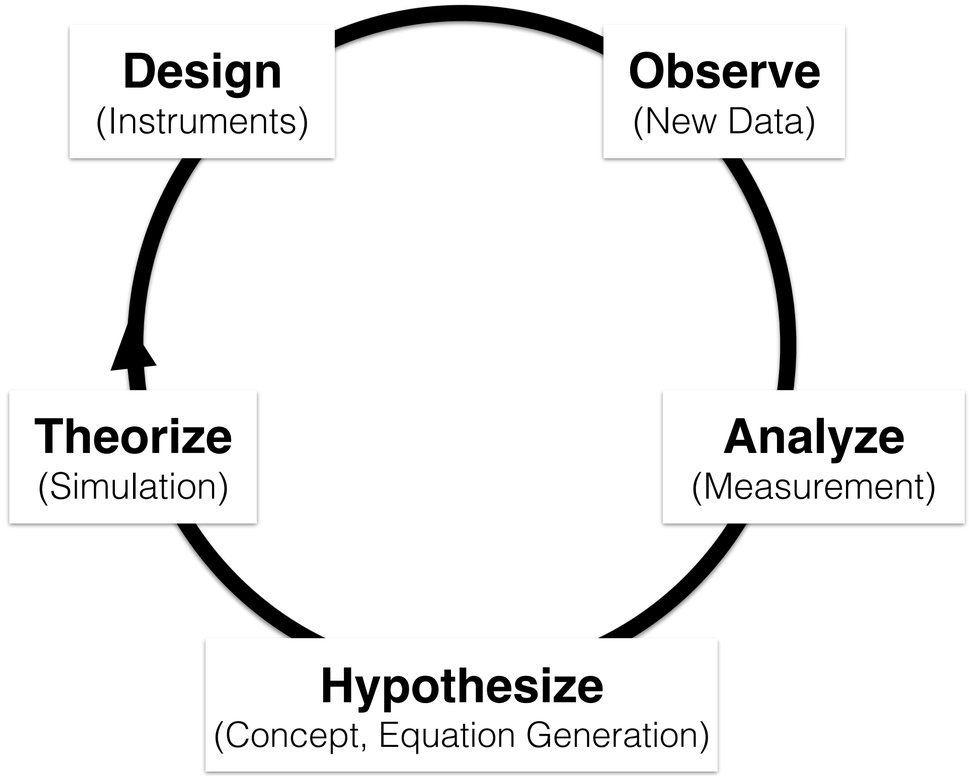
Workflow Automator

- *DeepUtils m*akes it trivially easy to run deep learning experiments on science data.

One of the major goals of Deep Skies

Automate the Scientific Cycle

Improve accuracy of and reduce time for design



- Science is expensive.
  - Instrument design costs money.
  - Decision-making and idea generation costs time.
  - Observations are precious.
- DOE is pursuing self-driving laboratories
  - 2019 AI Town Hall: *Automated Cosmic Experiment*
  - 2020 Future Scientific Methodologies Workshop
- Uses of Automated Design:
  - Connect the separate elements
  - Make design choices and solve debates more quickly
  - Provide evidence of project efficacy to program reviewers
- Requirements:
  - Improve interpretability of algorithms and models.
  - Keep humans in the loop.
  - Only automate what you can trust.

People in AI and Cosmology

47 People in FY23 affiliated with CPC

6	PI	Ciprijanovic, Drlica-Wagner, Frieman, Gnedin, Neilsen, Nord, Simon
5	Postdoc	Awan, Jarugula, McDermott, Nevin, Paul
12	Undergrad	Banker, Durojaiye, Kumar, Malagon, Ran, Silva, Swierc, Tsiane, Venkat, Willis, Youssef, Zhao
5	Grad	Dunn, Poh, Reza, Roncoli, Samudre, Schechter, Terranova
2	AI Associate	Lewis-Livaudais, Voetberg
2	Post-bac	Pensamiento, Tamargo
7	High school	Cohen, Grimm, Padhi, Sarkar, Shevchuk, Wu, Zhou
7	Affiliate	Avestruz, Khullar, Paul, Speagle, Trivedi, Vavagiakis, Zhang

(Most affiliations are through the Deep Skies Lab)

Opportunities Conversation and Collaboration

Journal Clubs, Education, Community Discussion

Deep Skies Lab Offerings

Journal Club

- Weekly
- Papers, blogs, videos
- Late-breaking topics:
  AI, stats, physics, ethics

Community Hour

- Weekly
- Consultation
- Any topic:
  science, career, educational

Edu Club

- Bi-weekly
- Presentations
- Introductory topics:
  Stats, AI, Computing

Seminar

- Monthly
- Presentations
- Advanced topics:
  Stats, AI, Computing

April 2023: First Deep Skies Lab Retreat (at Fermilab)

Considerations for Future Projects and Proposals

- New avenues and ideas?
  - Investigate areas that are not yet popular
  - Consider key problems in the AI space
  - Grand challenges for focus
- Papers and Conferences
  - AI publication schedule is very different: >5 opportunities per year.
  - Conferences and Workshops are like proceedings
  - Popular long-form venues for AI + Cosmic: MLST, OJA(?), MNRAS, JCAP, Astron. and Comp.
- Proposals and FoA
  - Off-the-shelf solutions to existing problems is not interesting: program managers
  - Finding appropriate FoA is difficult
  - Building collaboration long before FoA
  - Planning now for Spring LDRD and DOE FoA
- NSF AI + Astronomy Institutes (proposal in progress)
  - California
  - Maybe east coast
  - Midwest is Northwestern+UChicago+UIUC+Fermilab

Solving problems in **AI + Cosmology** requires ...

- **Data** from principled, mathematical origins (cosmology)
- **Algorithms** and models that can process large, complex data (AI)