AI as a Tool in Computational Astrophysics

Snigdaa S. Sethuram

October 21st, 2025

Astrophysics: A shared computational challenge

Credit: NASA JWST

The Data Challenge

Understanding the physical processes governing the universe using:

Observations & Simulations

- Simulations & Observations: Massive scale (TB-PB-scale outputs)
- Non-linear, statistically noisy data & extensive parameter spaces
- Traditional analysis methods overwhelmed
- Need: Automated intelligence processing

The Data Challenge → Why AI in Astro

Understanding the physical processes governing the universe using:

Observations, simulations, and theory

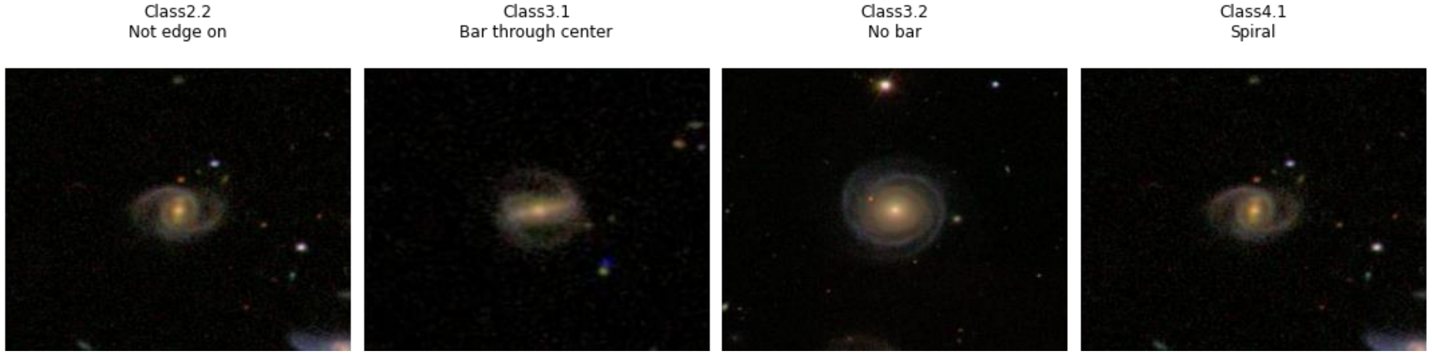
- Simulations & Observations: Massive scale (TB-PB-scale outputs)
- Highly correlated, high-dimensional data
- Traditional analysis methods overwhelmed
- **Need**: Automated intelligence processing

→

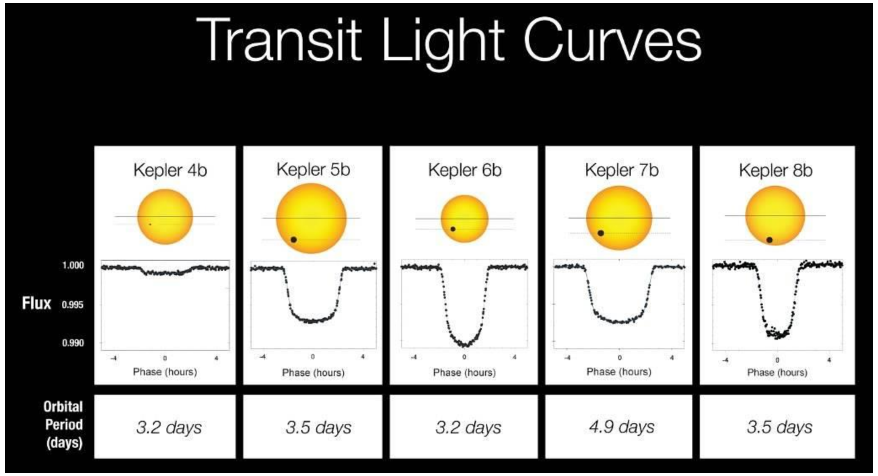
- Highly correlated, high-dimensional data
  - Pattern recognition
- Scalable to massive datasets
- Find non-obvious correlations
  - Self-/Un-supervised learning

Example AI applications in astrophysics

- Exoplanet detection in light curves
- Galaxy morphology classification
- Multi-modal searching – e.g. AstroCLIP
- Subgrid physics emulators



Jay Speidell, Galaxy Zoo



NASA/Kepler Mission

Traditional calculations vs. AI-enhanced calculations

Credit: JWST NIRCam image release (Chameleon I dark molecular cloud)

Strengths & weaknesses: Traditional algs

Strengths

- Physically motivated, interpretable
- Well-characterized uncertainties
- Extrapolates beyond training regimes
- Established validation/publication standards

Weaknesses

- ↑ complexity = ↑↑ computational expense
- Programmed and constrained physics
- Misses unexpected non-obvious patterns
- Difficult to scale

Strengths & weaknesses: Modern AI algs

Strengths

- Efficiently handles high-dimensional feature spaces
- Discovers non-obvious correlations
- Scales efficiently
- Learns complex nonlinear relationships
- Rapid inference after training

Weaknesses

- Needs large training datasets
  - Labeled data can be expensive
- Interpretability is often difficult
  - Uncertainty quantification imprecise
- Often fails when extrapolating beyond training parameter space

Science First: Planning Projects

Decision criteria: data volume, complexity, interpretability, computational resources

Best approach: hybrid implementation

Physics-informed networks, ML for bottlenecks and traditional for interpretation, Emulators validated against resolved simulations

Science First: Planning Projects

Decision criteria: data volume, complexity, interpretability, computational resources

Best approach: hybrid implementation

Physics-informed networks, ML for bottlenecks and traditional for interpretation, Emulators validated against resolved simulations

- Define clear science question first
  - Does your dataset provide a specific advantage?
- Determine where ML truly adds value
- Remain flexible on architecture choices

Key Takeaways

- AI transforms astrophysics through scale
- Know the strengths and weaknesses: traditional & AI
- Hybrid approaches often work best
- Science question drives method choice
- Apply ML strategically to different data bottlenecks

snigdaa.ram@gmail.com

Snigdaa S. Sethuram

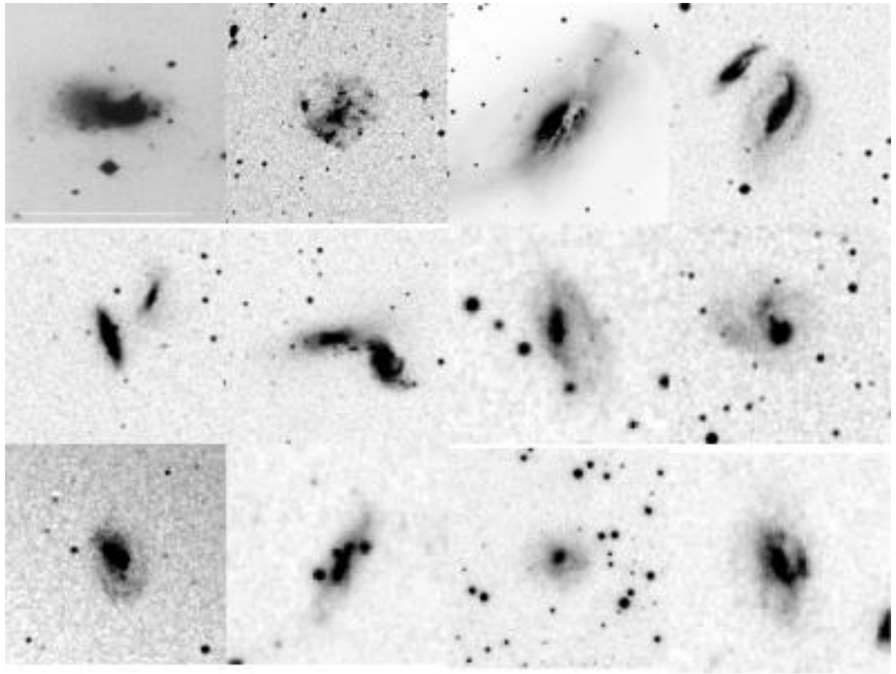
snigdaa.ram@gmail.com

www.snigdaasethuram.com

Thank you

Astrophysics overview

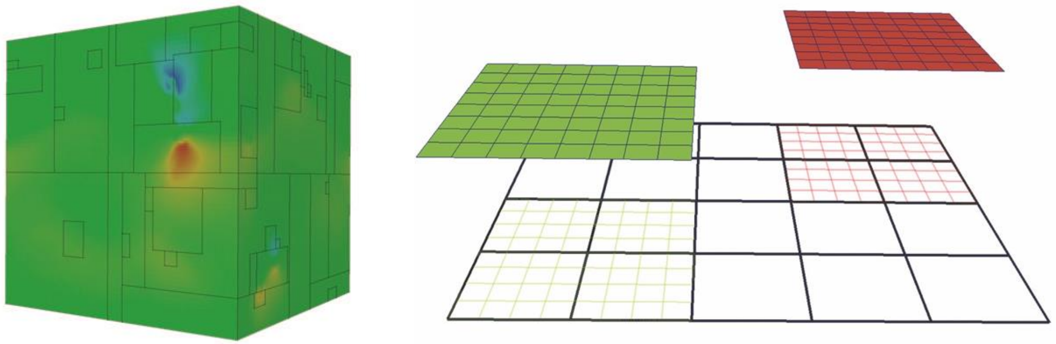
- Observation-specific challenges:
  - Instrumental noise and systematic errors
  - Incomplete sky coverage and selection effects
  - Limited time resolution
  - No control over astrophysical "experiments"



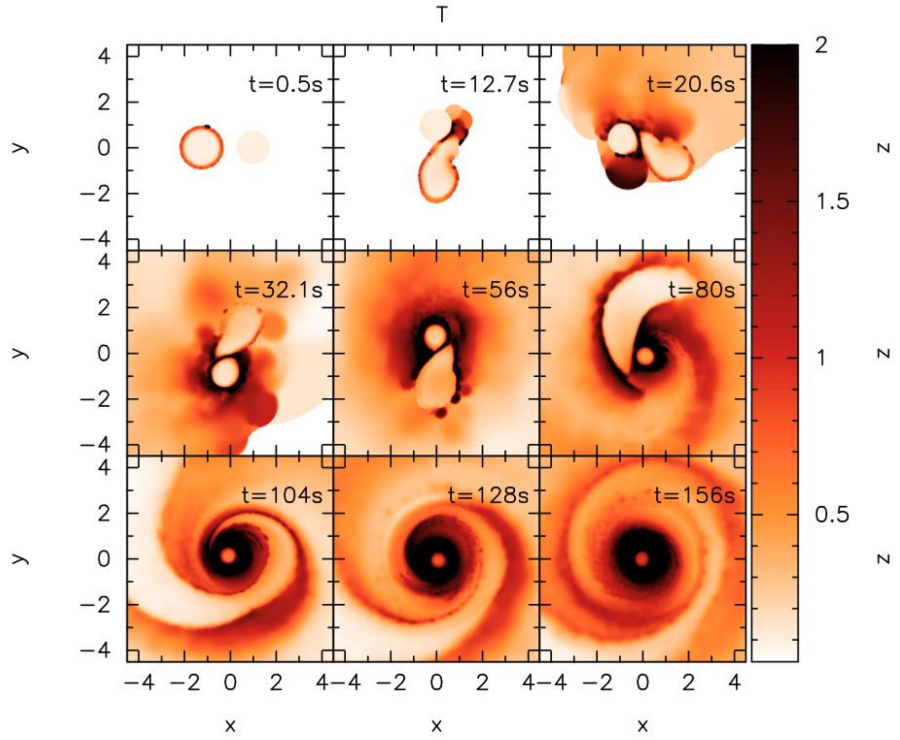
Conselice (2006)

Astrophysics overview

- Simulation-specific challenges:
  - Finite resolution limits resolved physics
  - Subgrid physics requires approximations
  - Computational cost limits parameter exploration



Sharkey et al. (2011)



Loren-Aguilar et al. (2009)

Data challenges made for AI

- Scale and Volume
  - PB+ scale datasets from surveys and simulations (LSST, Euclid, JWST; IllustrisTNG, FIRE, CAMELs)
  - Millions-billions of objects requiring classification (stars, galaxies, AGN, clusters, cosmic rays, binary systems, planetary systems, etc.)
  - Real-time processing of incoming data streams (telescope data; dynamic simulation modeling)
- Complexity
  - High-dimensional parameter spaces
  - Non-linear relationships between observables and physical properties
  - Complex analytical relationships in data (stellar history assumptions in observation, subgrid modeling in simulations)

Data challenges made for AI

- Computational bottlenecks
  - Computationally expensive to explore full parameter space + Multiphysics
  - Cross-cell flux calculations, subgrid approximations, implicit time-stepping schemes, self-gravity solvers replaceable with emulators "cheaply"
  - Creating mock observations (radiative transfer calculations like ray-tracing) requires thousands of iterations
- Pattern recognition
  - Identifying rare or unexpected phenomena in large datasets
  - Extracting weak signals from noisy observations
  - Morphological classification of complex structures

When to choose which tool

Decision criteria: data volume, complexity, interpretability, computational resources

Use traditional methods when:

- Studying simple well-understood processes
- Small datasets or rare phenomena
- Interpretability and transparency critical
- Extrapolation beyond data required
- Established methods work accurately and efficiently

Use AI methods when:

- Complex patterns in high dimensions
- Massive datasets requiring scalability
- No clear analytical model exists/is resolved
- Find unknown/missing correlations
- Large amounts training data available

Best approach: hybrid implementation

- Physics-informed networks with constraints
- ML for bottlenecks, traditional for interpretation
- Emulators validated against resolved simulations, real observations