



# More than meets the eye (AI): Towards an Artificial Intelligence Observatory

---

**Matias Carrasco Kind**

Senior Research Scientist, National Center for Supercomputing Applications

Director, Data Science Research Services, Gies College of Business

Assistant Research Professor, Astronomy

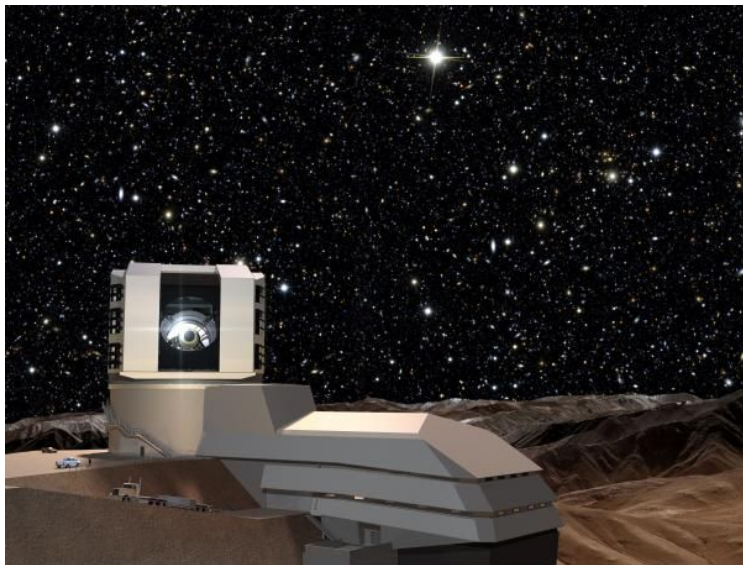
Data Release Scientist, Dark Energy Survey

University of Illinois at Urbana-Champaign

[go.ncsa.illinois.edu/delta21](https://go.ncsa.illinois.edu/delta21)

2nd International Conference in Deep Learning Theory and Applications, DeLTA 2021,  
July 7-9, 2021

# Observatories generate data to understand the Universe



- We can ask questions → Define what and how to observe
- We can use different instruments → Different science cases
- Many discoveries beyond design → Low hanging fruit

# What If?

Thought experiment

- Extremely Artificial Intelligence
- Infinite Memory (physical and analytical)
- Infinite compute resources
- Feed ALL possible observations
- Feed Simulations
- 1000s of interconnected AI models
- Able to answer any question, find hidden patterns, generate new data

# What If?

Thought experiment

- Extremely Artificial Intelligence
- Infinite Memory (physical and analytical)
- Infinite compute resources
- Feed ALL possible observations
- Feed Simulations
- 1000s of interconnected AI models
- Able to answer any question, find hidden patterns, generate new data

# Answer: 42

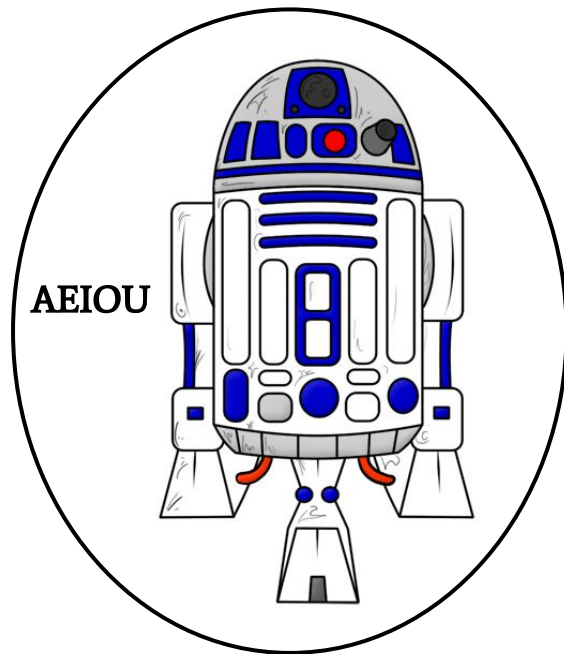
# Artificial & Extremely Intelligent Observatory Unit (AEIOU)



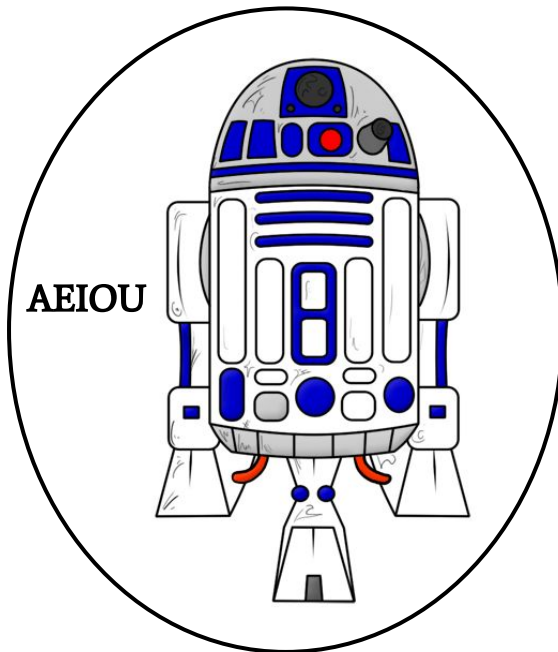
# Artificial & Extremely Intelligent Observatory Unit (AEIOU)



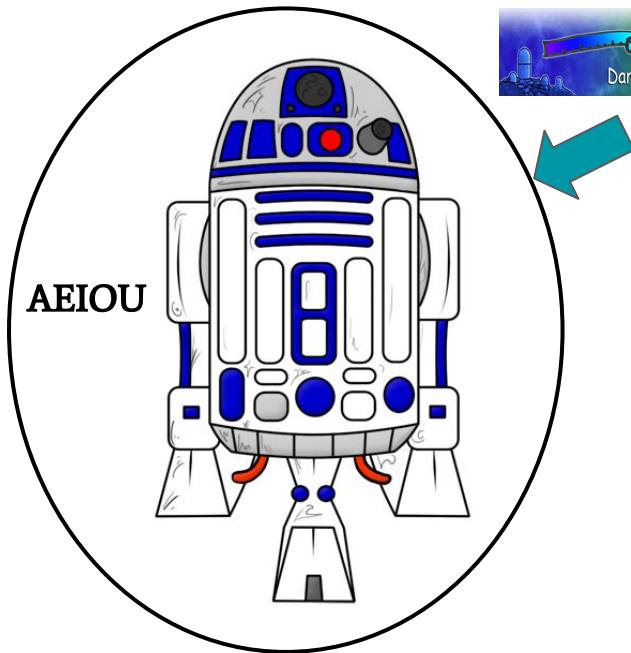
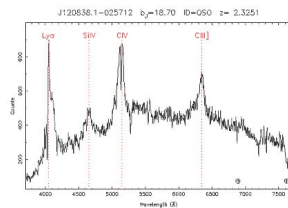
# What if?



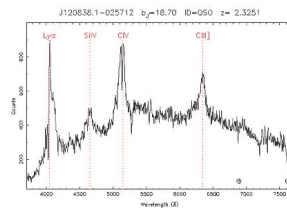
# What if?



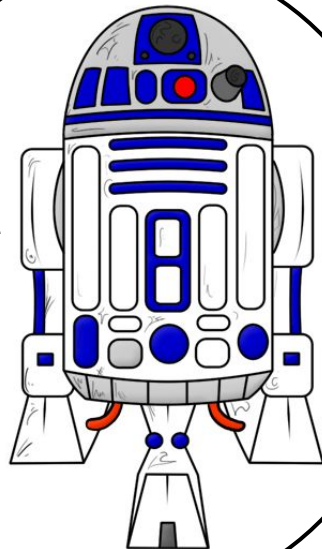
# What if?



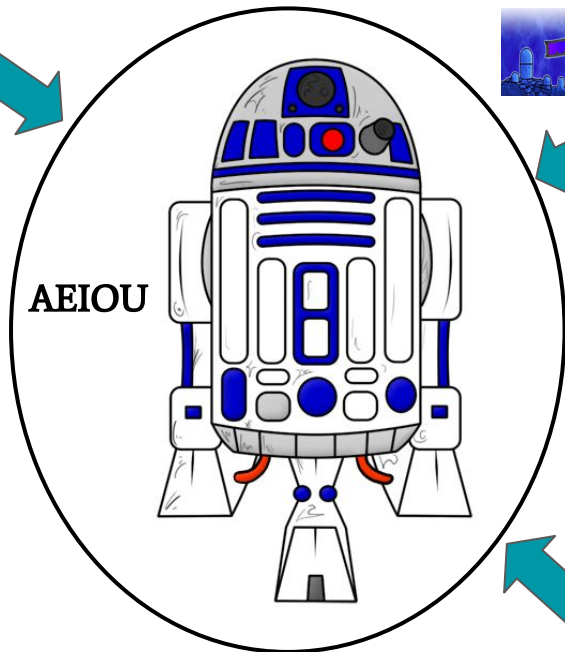
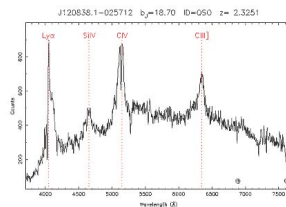
# What if?



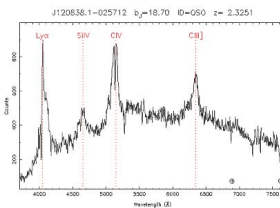
AEIOU



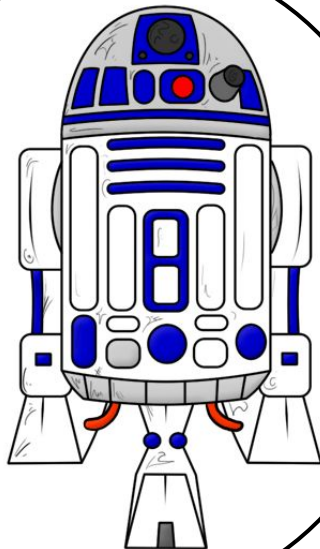
# What if?



# What if?



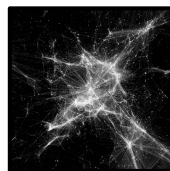
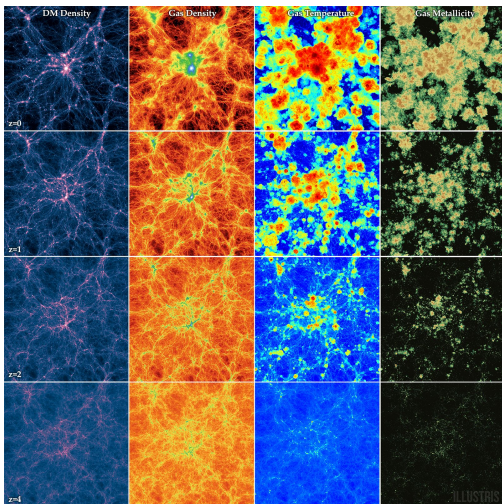
AEIOU



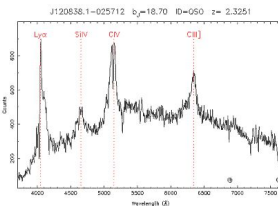
gaia



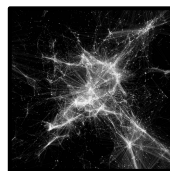
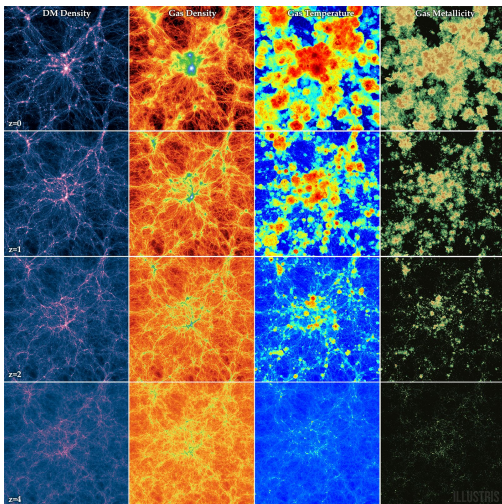
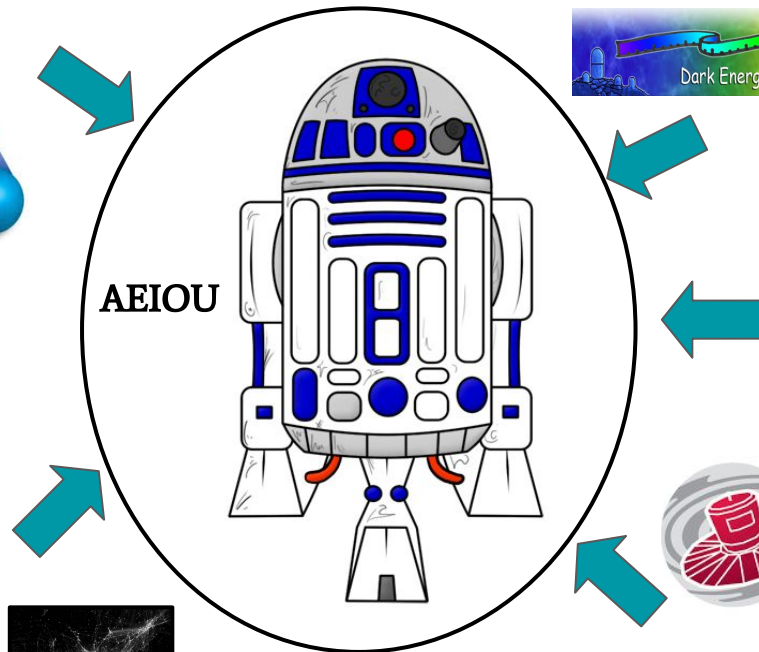
PLANCK



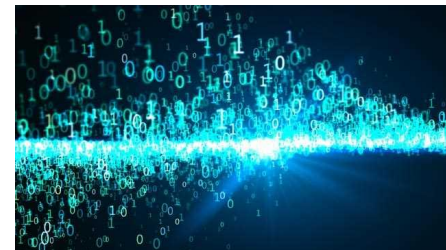
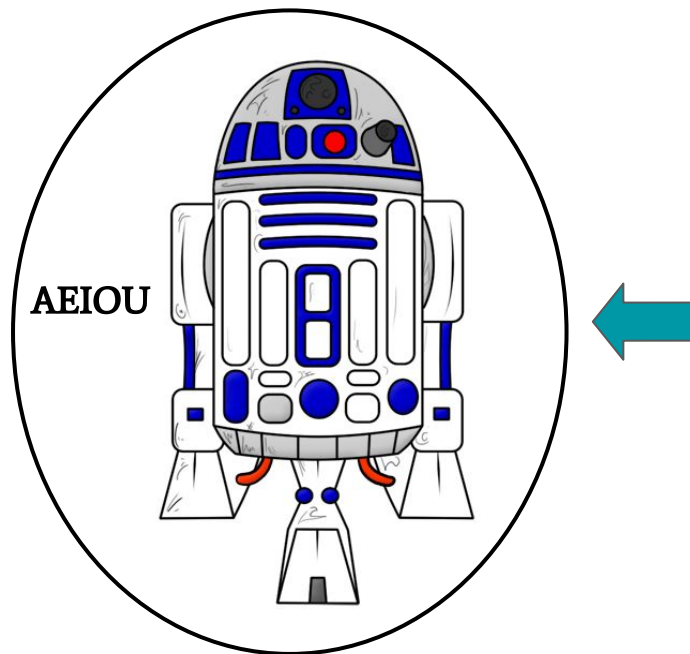
# What if?



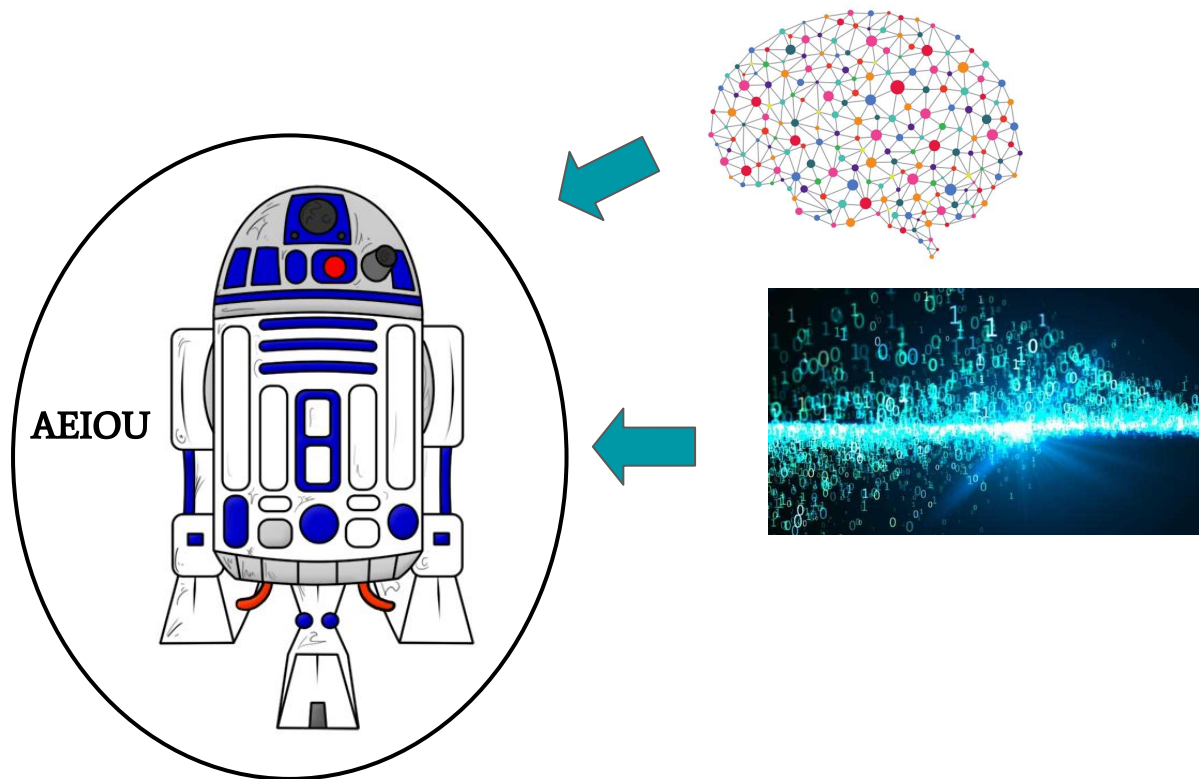
gaia



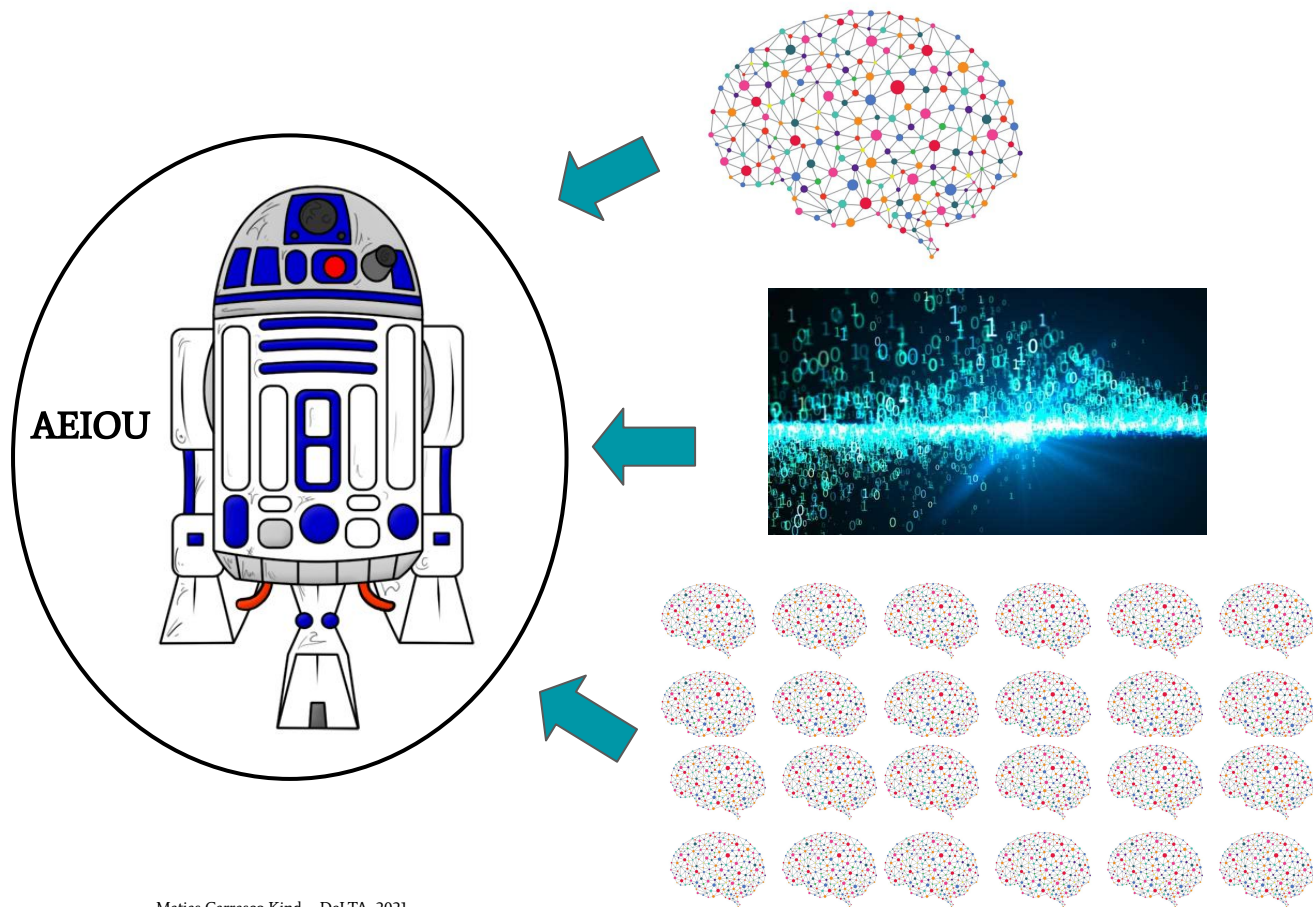
# What if?



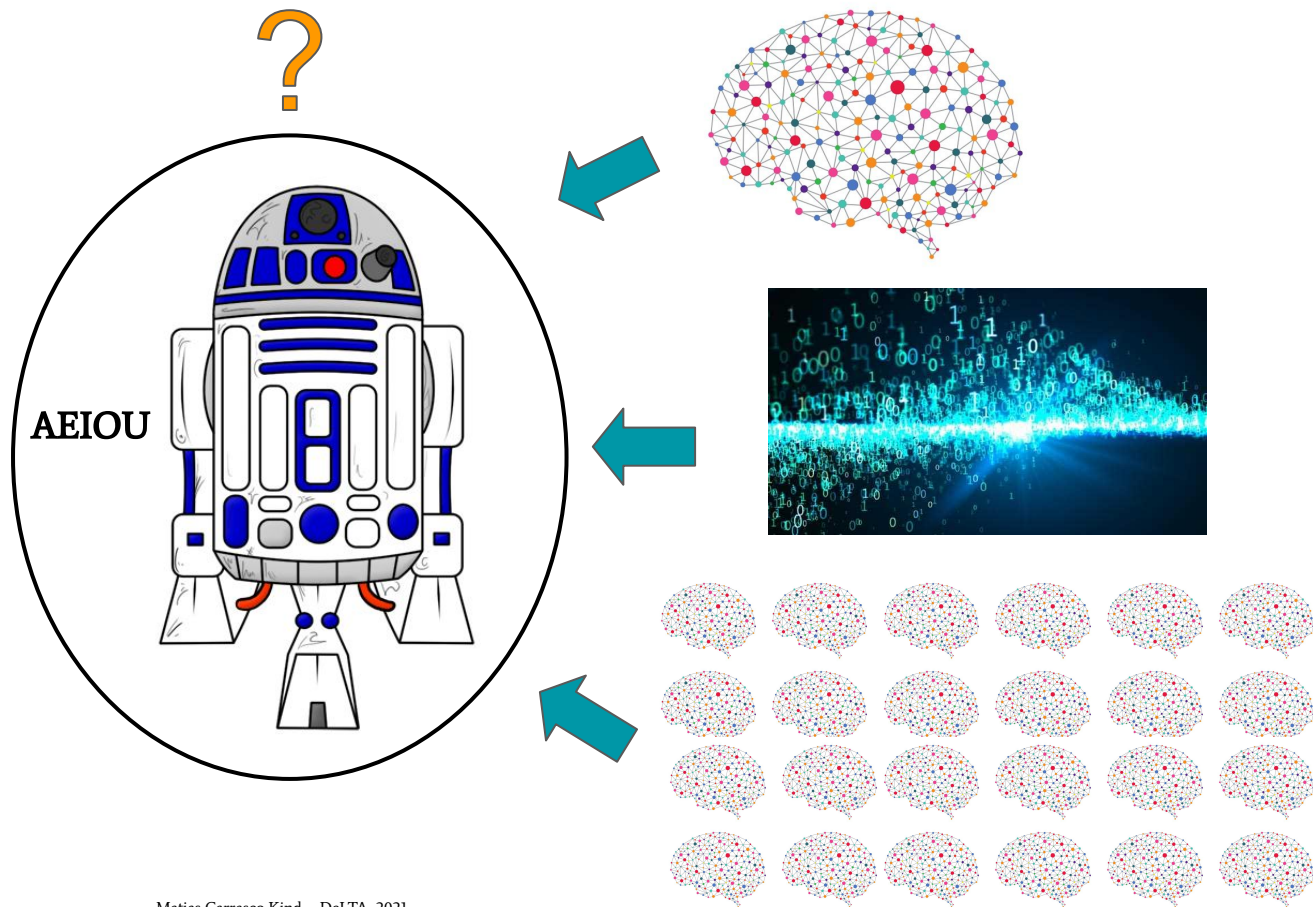
# What if?



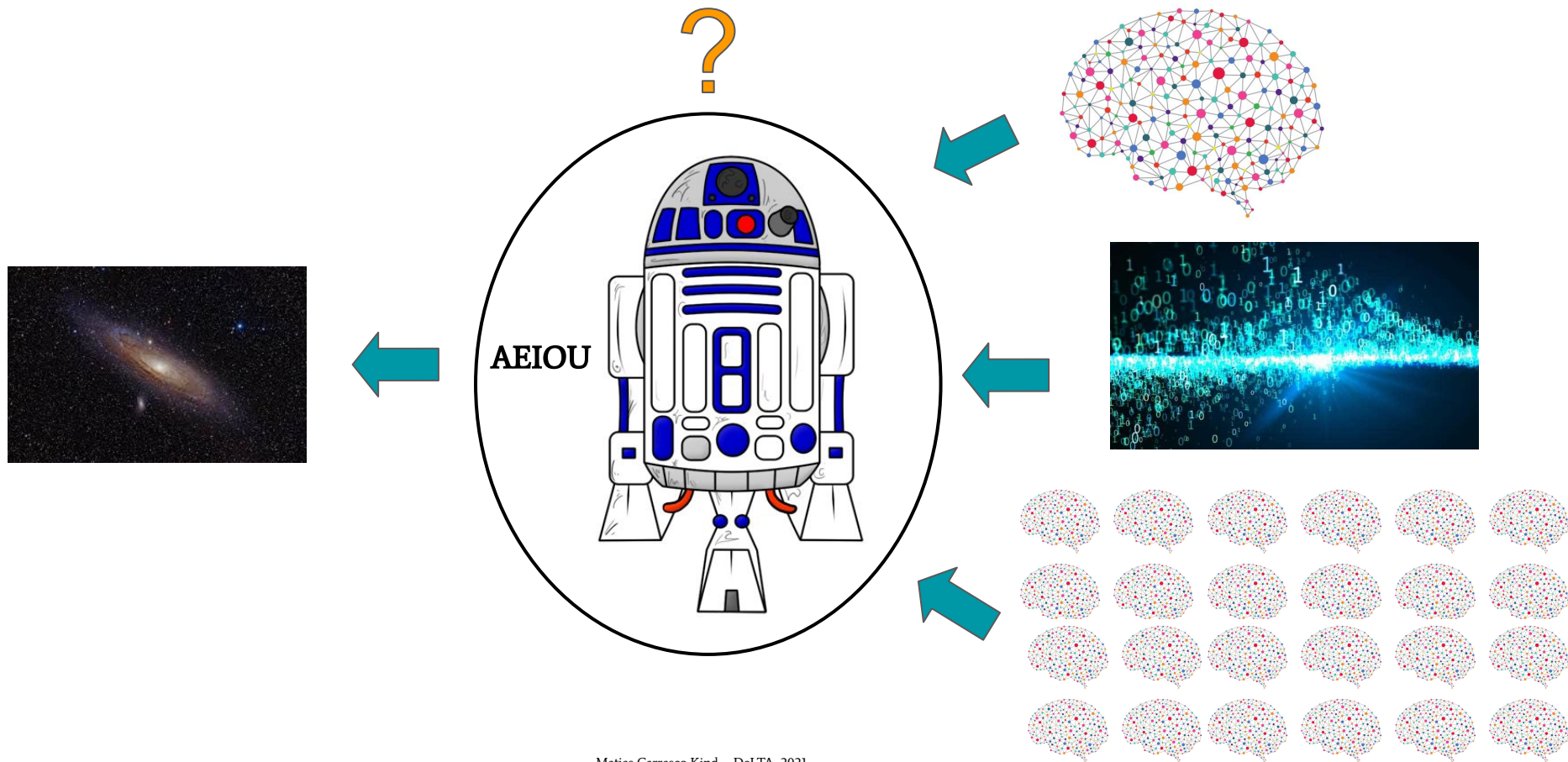
# What if?



# What if?



# What if?



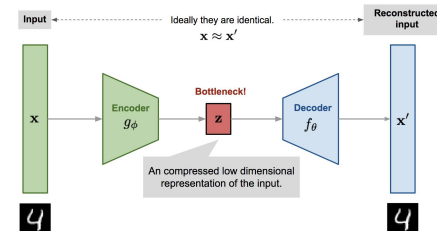
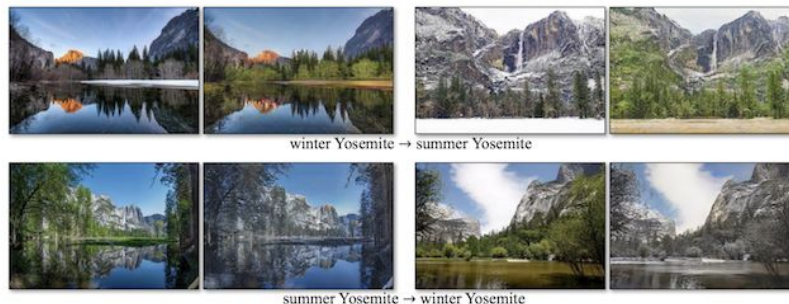
# Deep Learning Advances



<https://thispersondoesnotexist.com/>

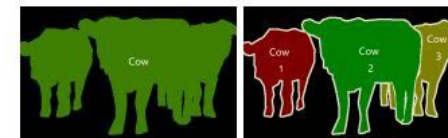
<https://arxiv.org/abs/1812.04948>

- CycleGan
- CVAE
- MVAE
- Noise2Noise
- Pix2Pix
- Transfer Learning
- Bayesian Framework
- ... and so many others



(a) Image Classification

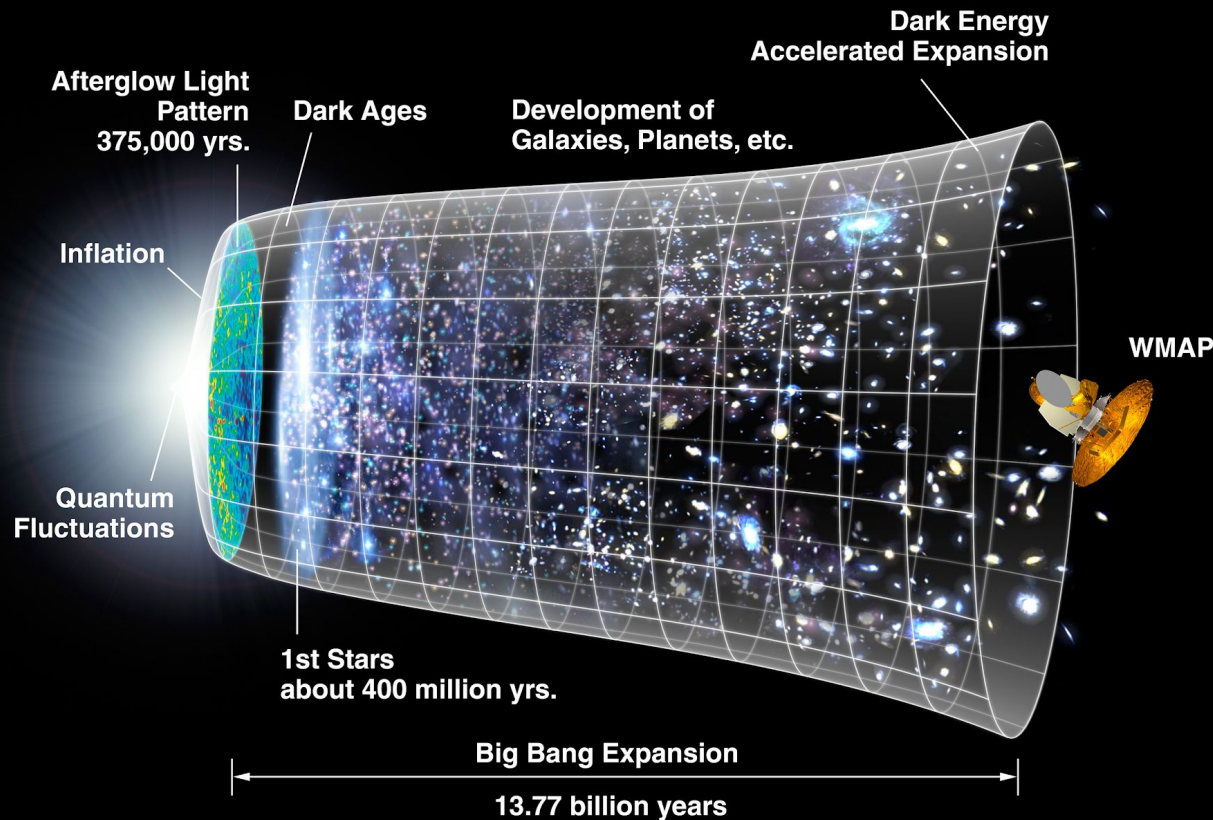
(b) Object Detection



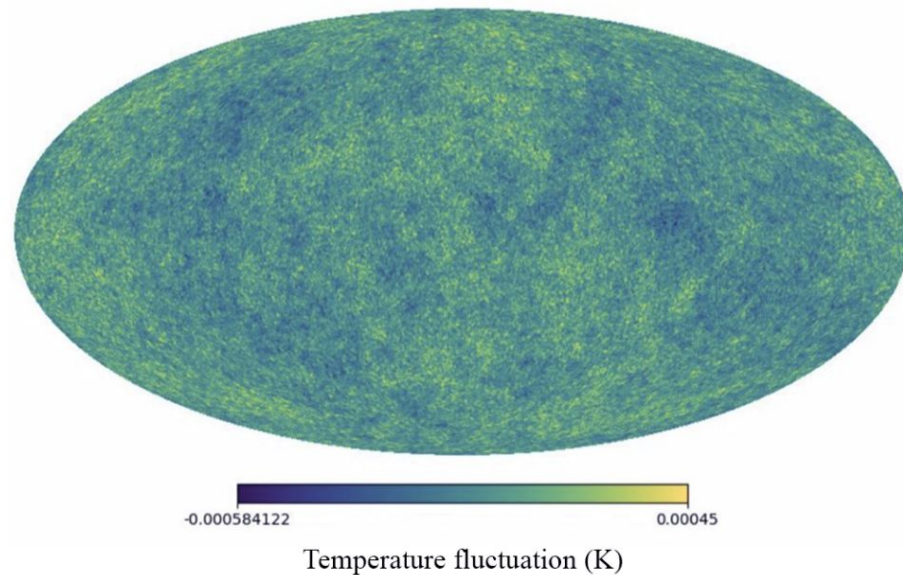
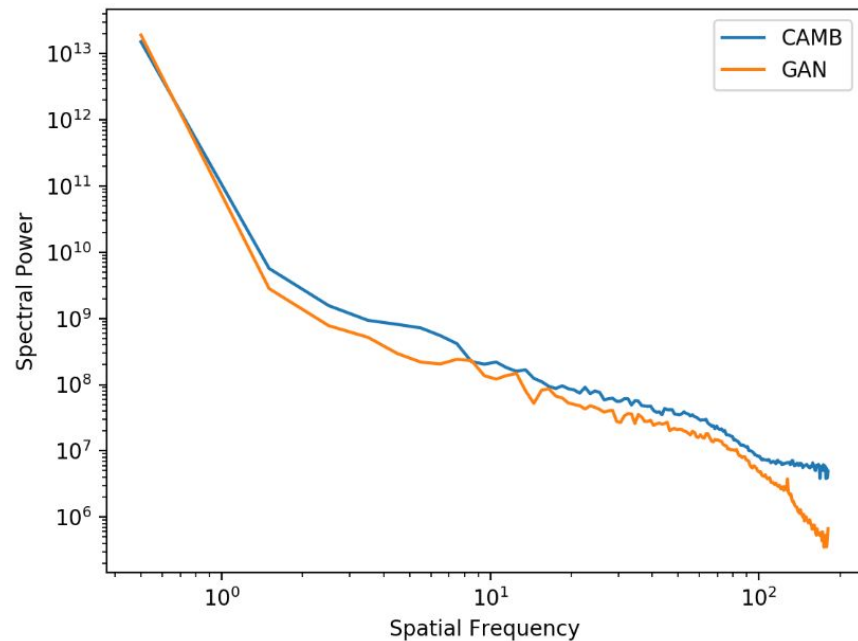
(c) Semantic Segmentation

(d) Instance Segmentation

# The History of the Universe in a nutshell

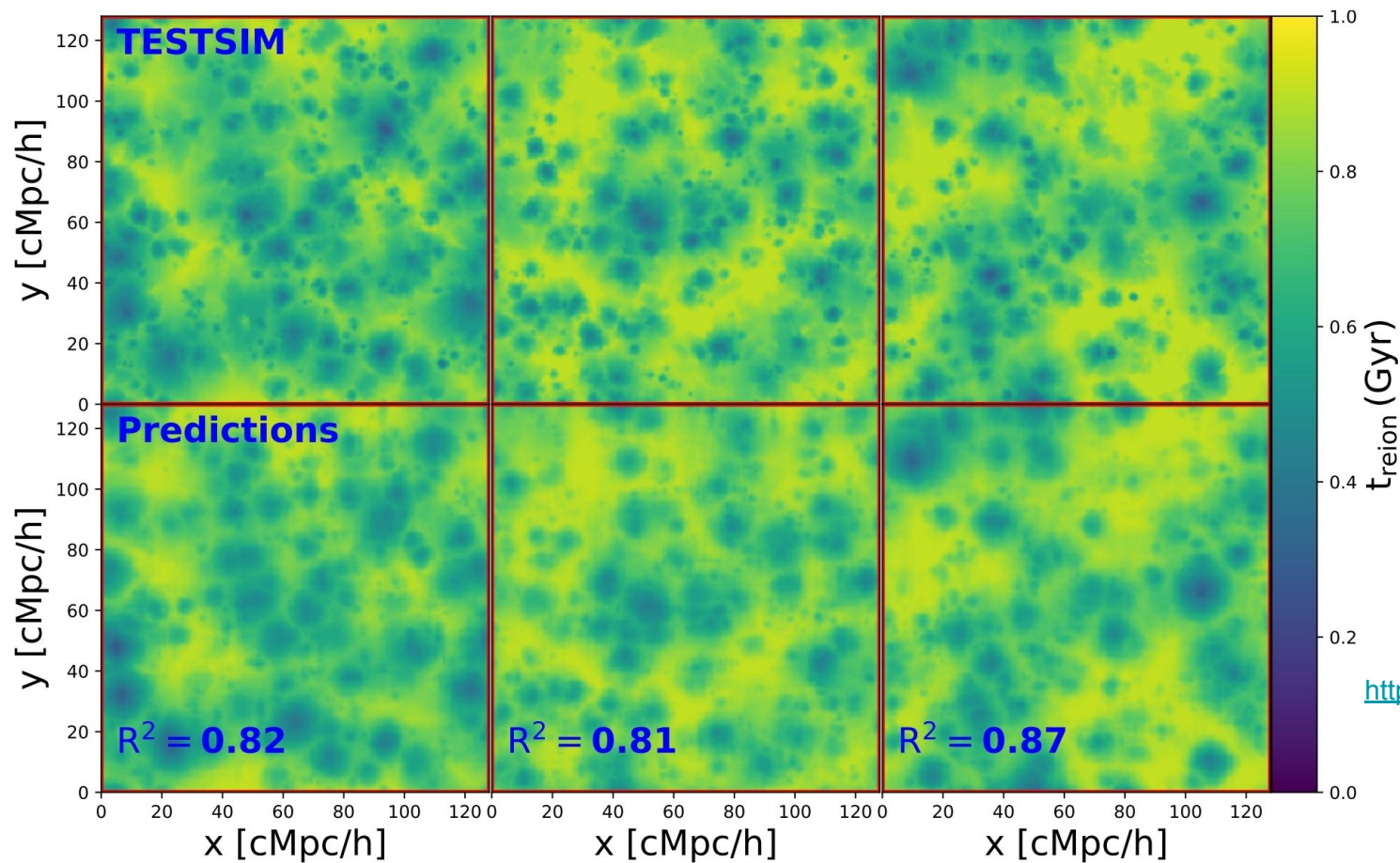


# Cosmic Microwave Background



<https://arxiv.org/abs/1908.04682>

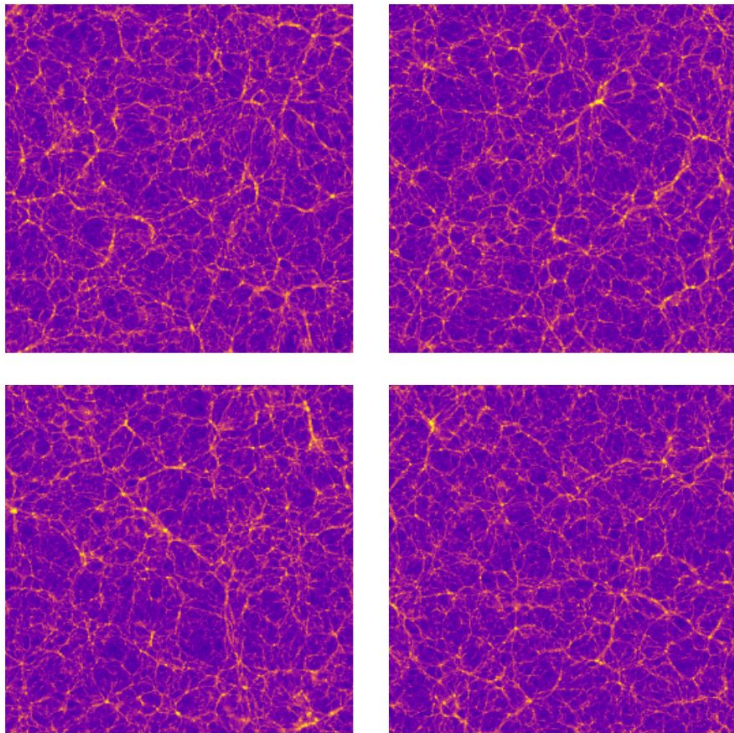
# Cosmic Reionization



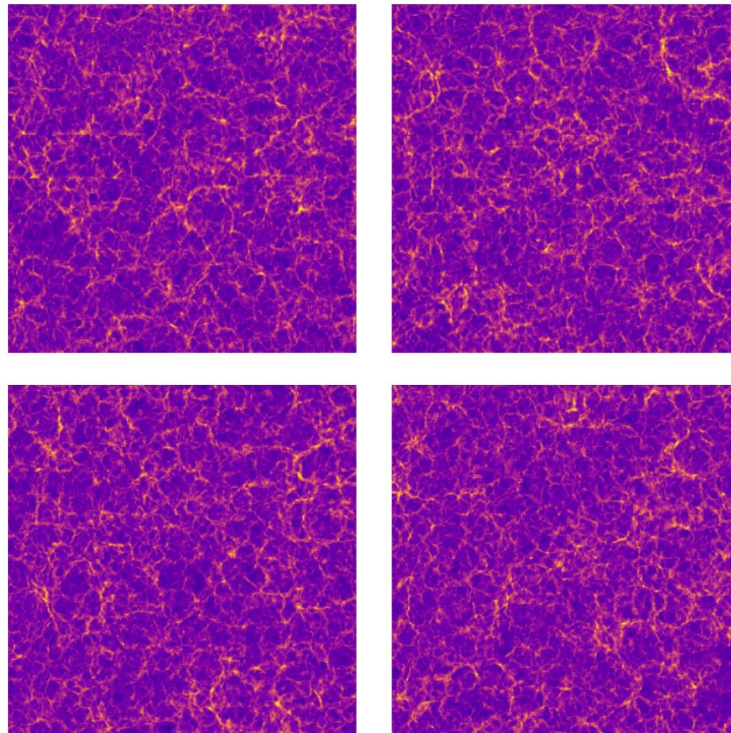
<https://arxiv.org/abs/1905.06958>

# Large Scale Structure

Real  $256^3$

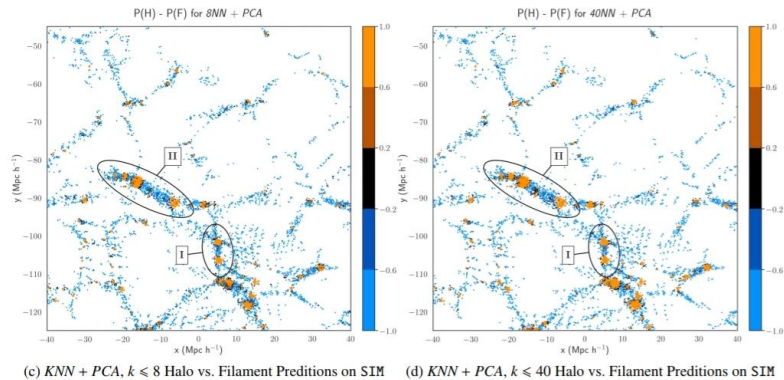
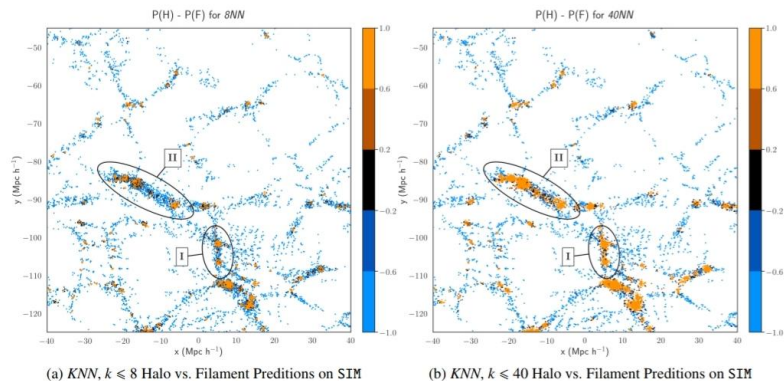


Fake  $256^3$

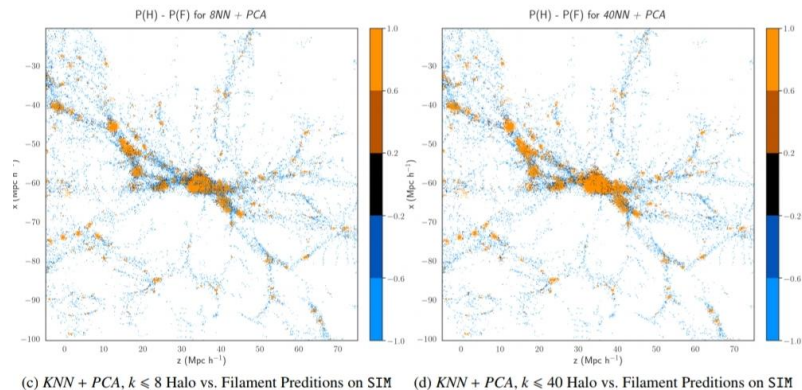
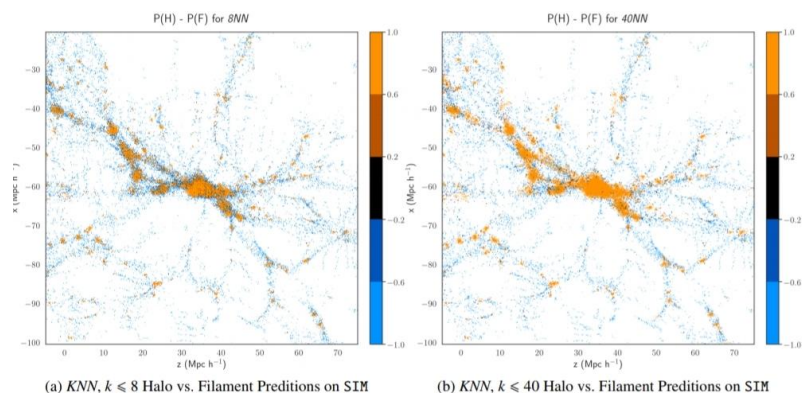


<https://arxiv.org/abs/1908.05519>

# Large Scale Structure



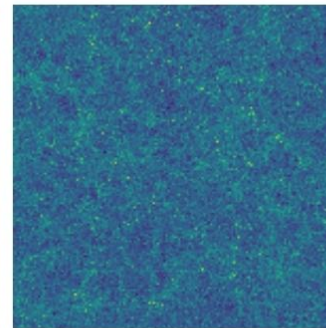
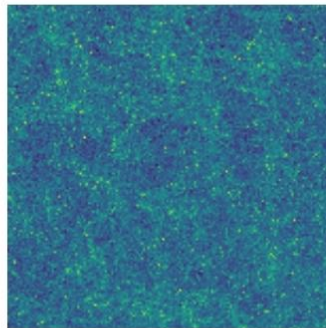
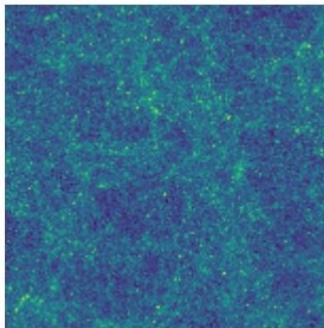
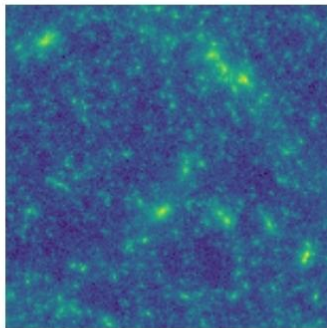
vs. filament probability scatterplots for  $KNN$  ((a) and (b)) and  $KNN + PCA$  ((c) and (d)) for  $k \leq 8, 40$ . Particles are from a slice of  $v$  region, most halos had radii that were not substantially larger than the maximum radius used for  $PCA$  calculations ( $r = 2.0$  Mpc).



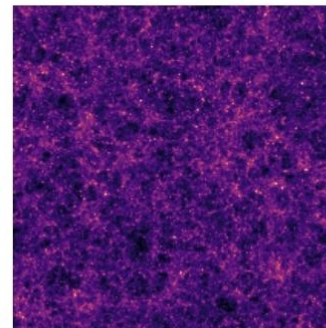
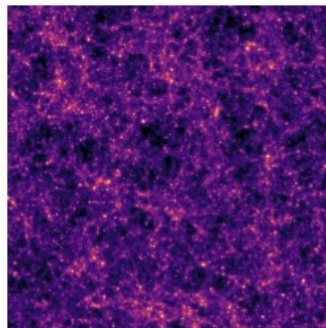
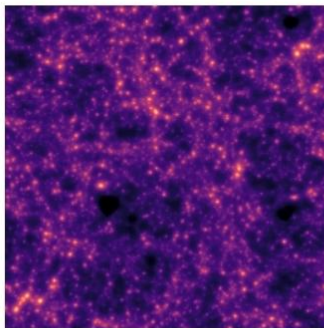
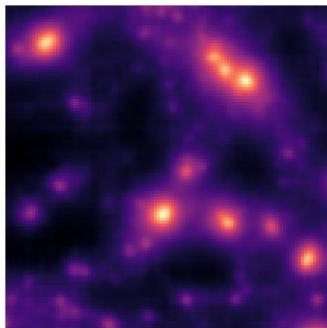
# Adding baryons to N-body simulations

 $z = 0.0$  $z = 0.3$  $z = 0.6$  $z = 1.0$ 

Dark matter  
(SLICS)

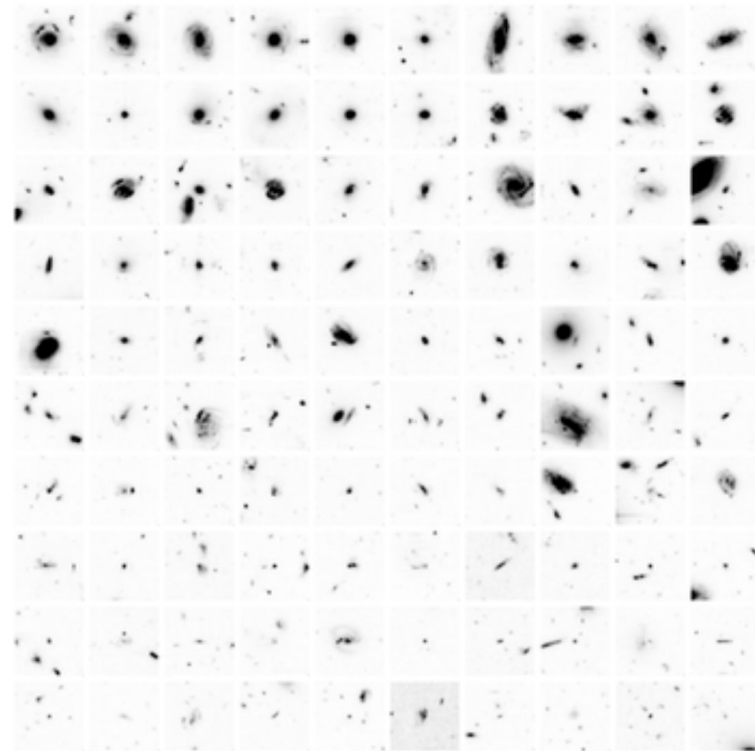
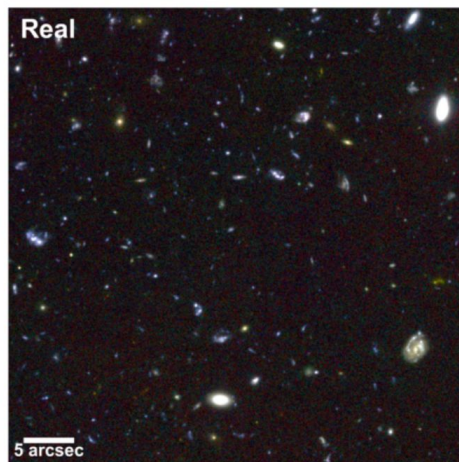
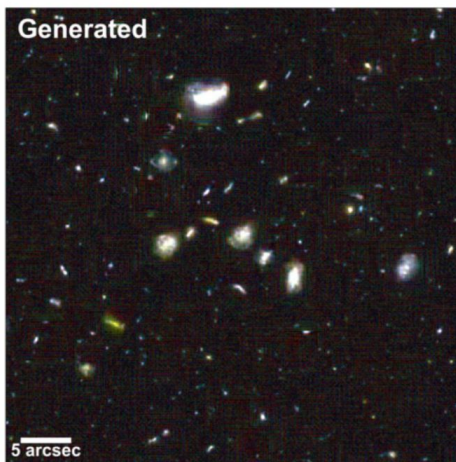
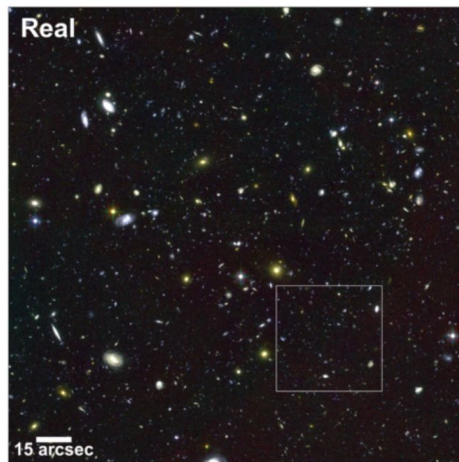
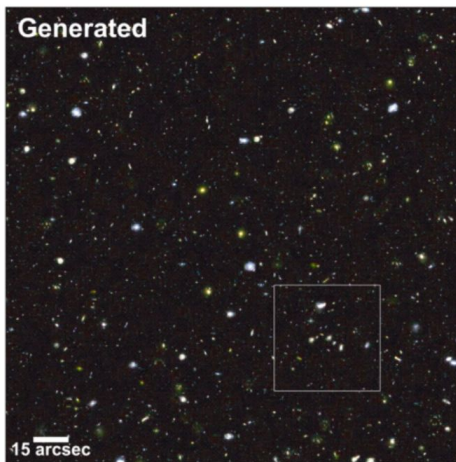


Pressure  
(VAE)



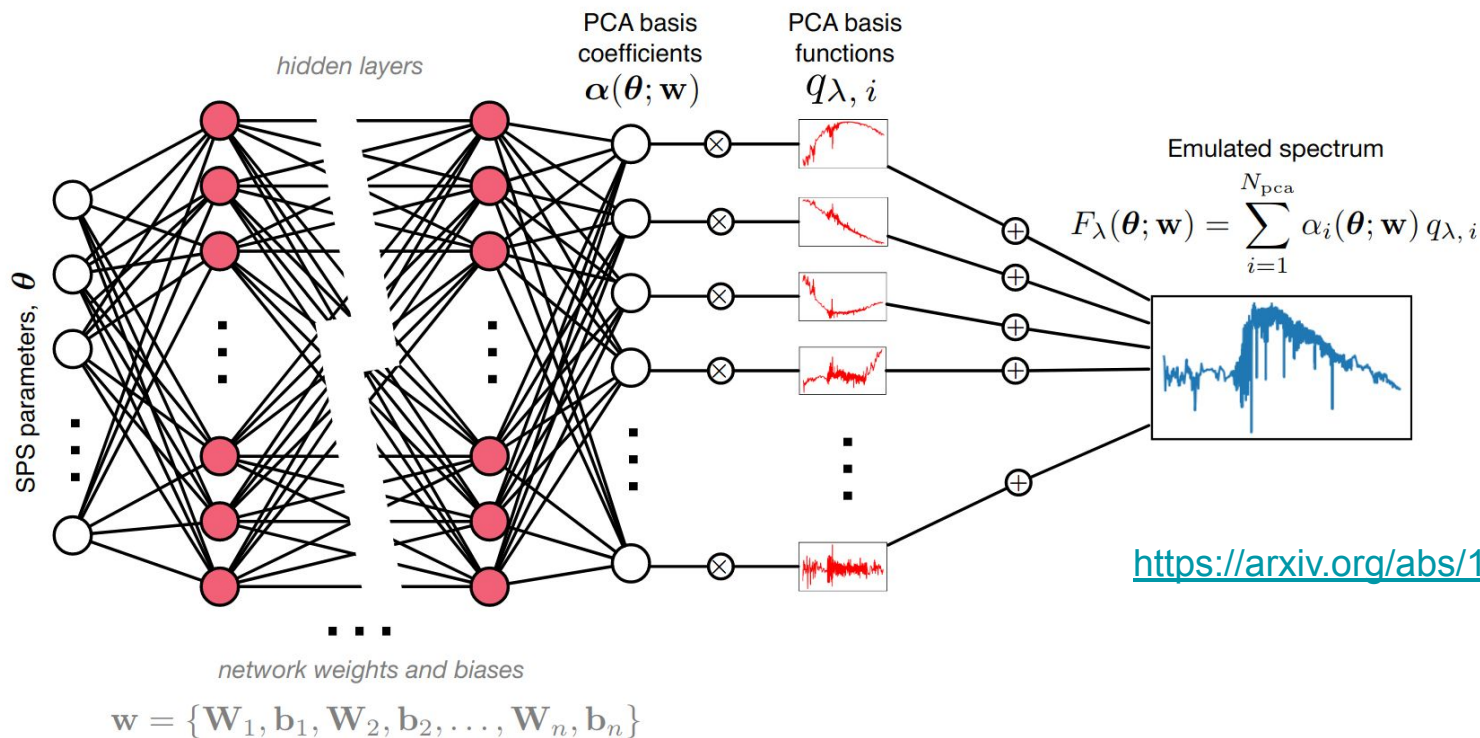
<https://arxiv.org/abs/1903.12173>

# Deep Fields Observations



<https://arxiv.org/abs/1904.10286>

# Stellar population and galaxy spectra



**Figure 1.** Schematic of the PCA neural network emulator set-up. A dense neural network parameterizes the PCA basis coefficients as a function of the SPS model parameters (i.e., taking SPS parameters as input and predicting the basis coefficients). These basis coefficients are then multiplied by their respective PCA basis functions and summed to give the predicted spectrum.

# Creating galaxies given a prior

*Real*



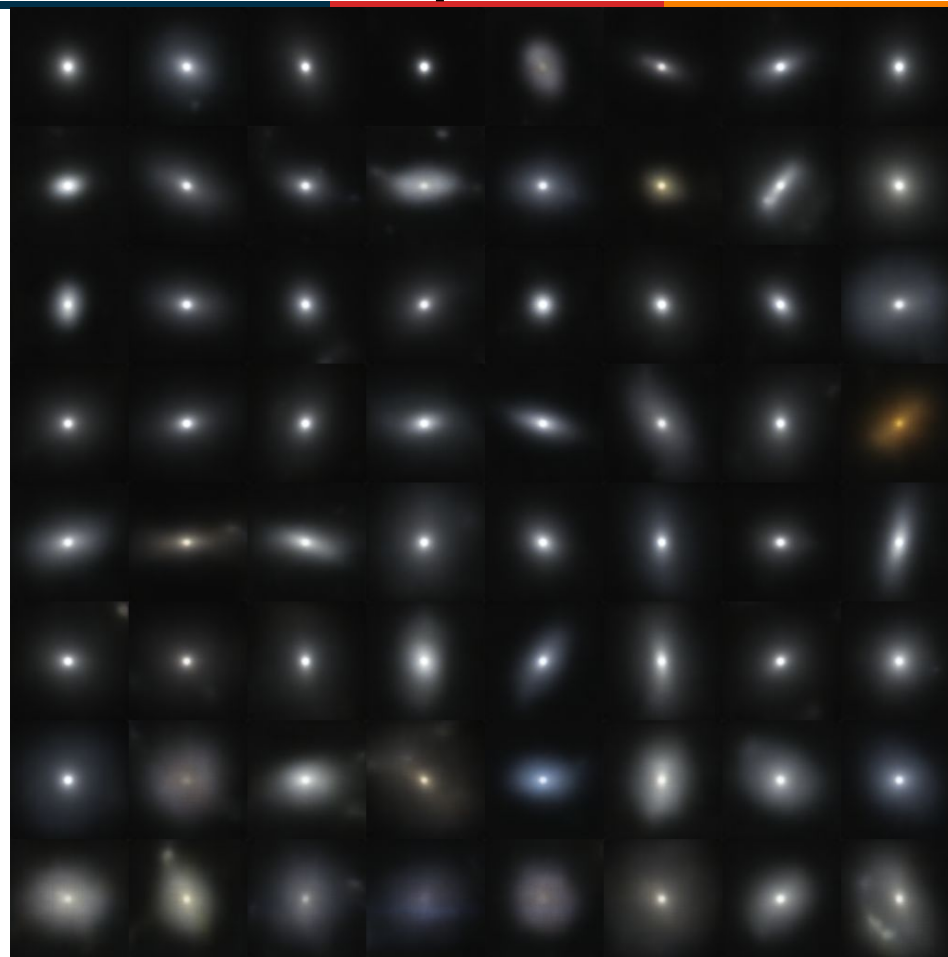
*Ours*



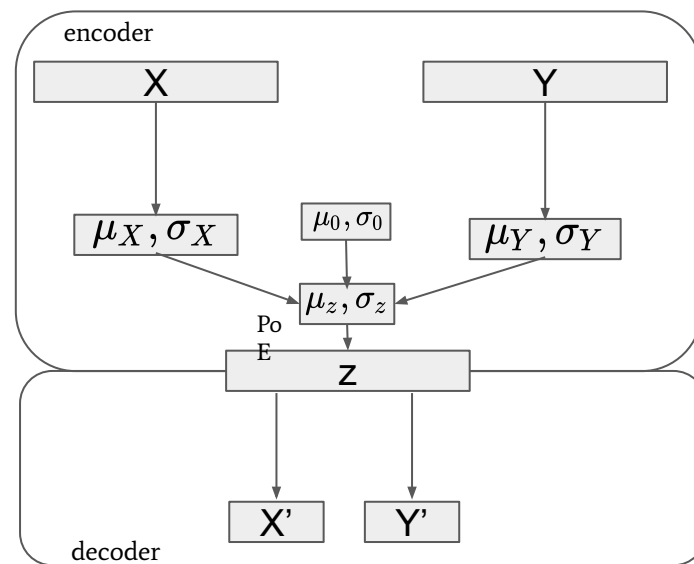
*CAVEP*



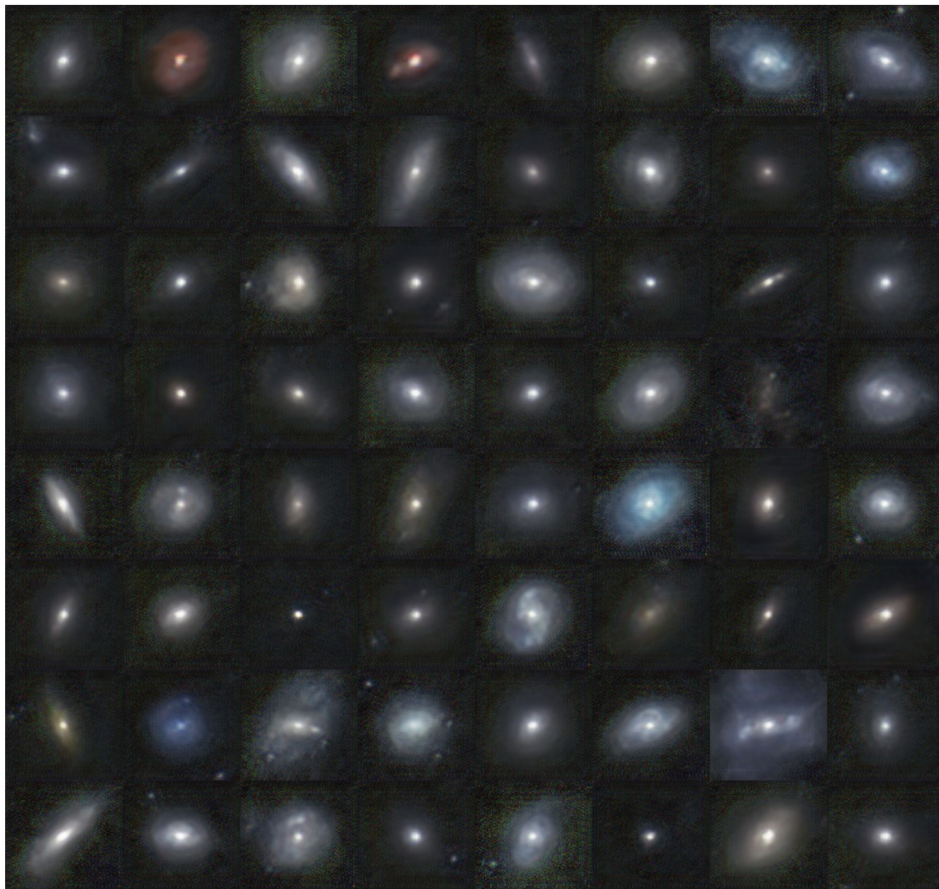
# And more examples...



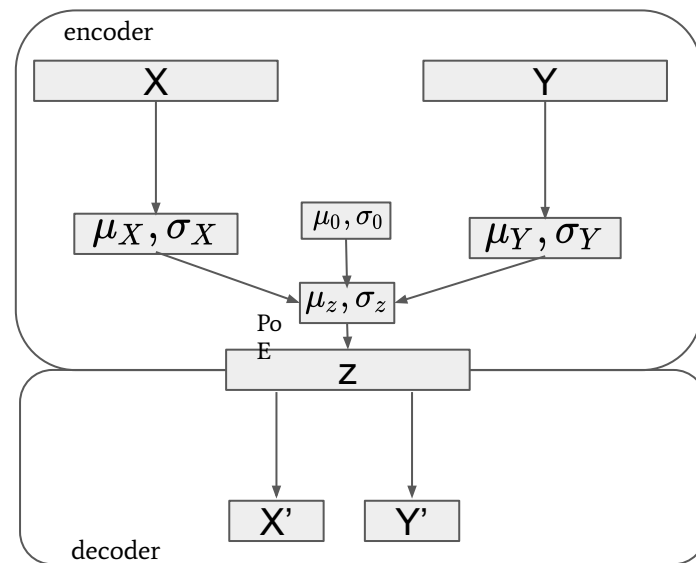
Samples with changing concentration  
(increasing downwards)



# And more examples + super resolution...



Samples with changing concentration  
(increasing downwards)

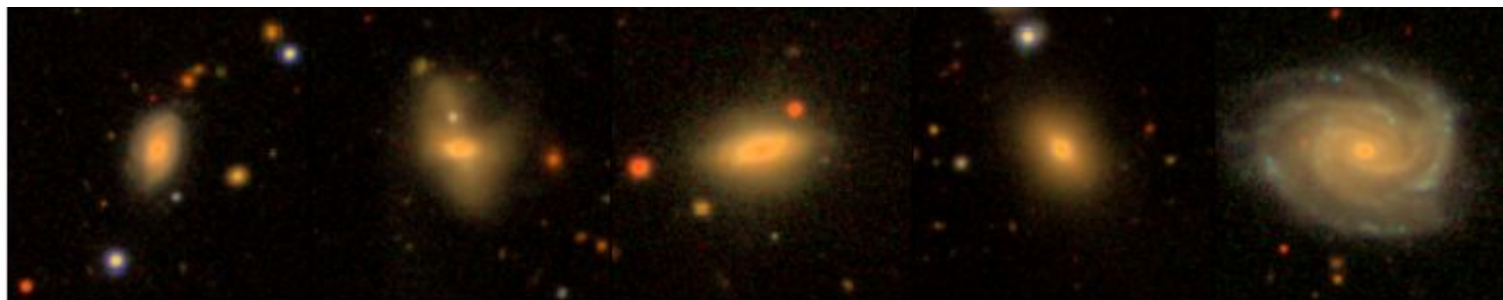


# Optical to Optical to increase S/N

SDSS



DES



Buncher, Sharma,  
Carrasco-Kind, 2021.

# Optical to Optical to increase S/N



Buncher, Sharma,  
Carrasco-Kind, 2021.

# Optical to Optical to increase S/N

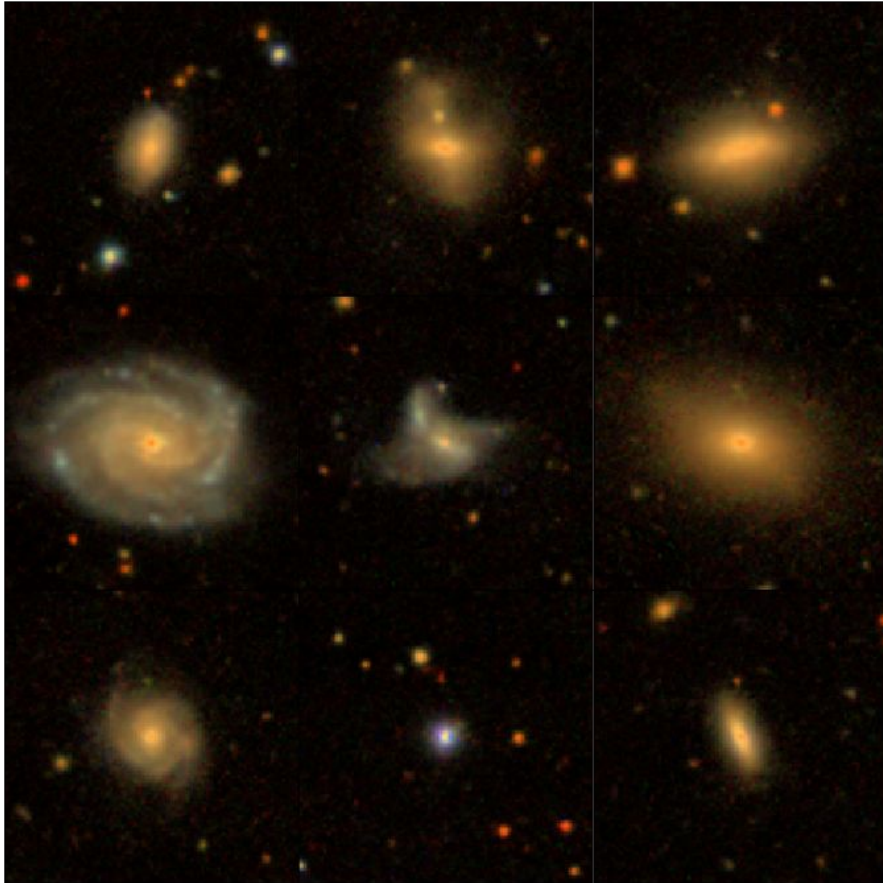


**SDSS Images not in training**

Two different models, using  
CNN+VAE and CycleGAN we  
can increase the magnitude and  
S/N up to 8%

Buncher, Sharma,  
Carrasco-Kind, 2021.

# Optical to Optical to increase S/N



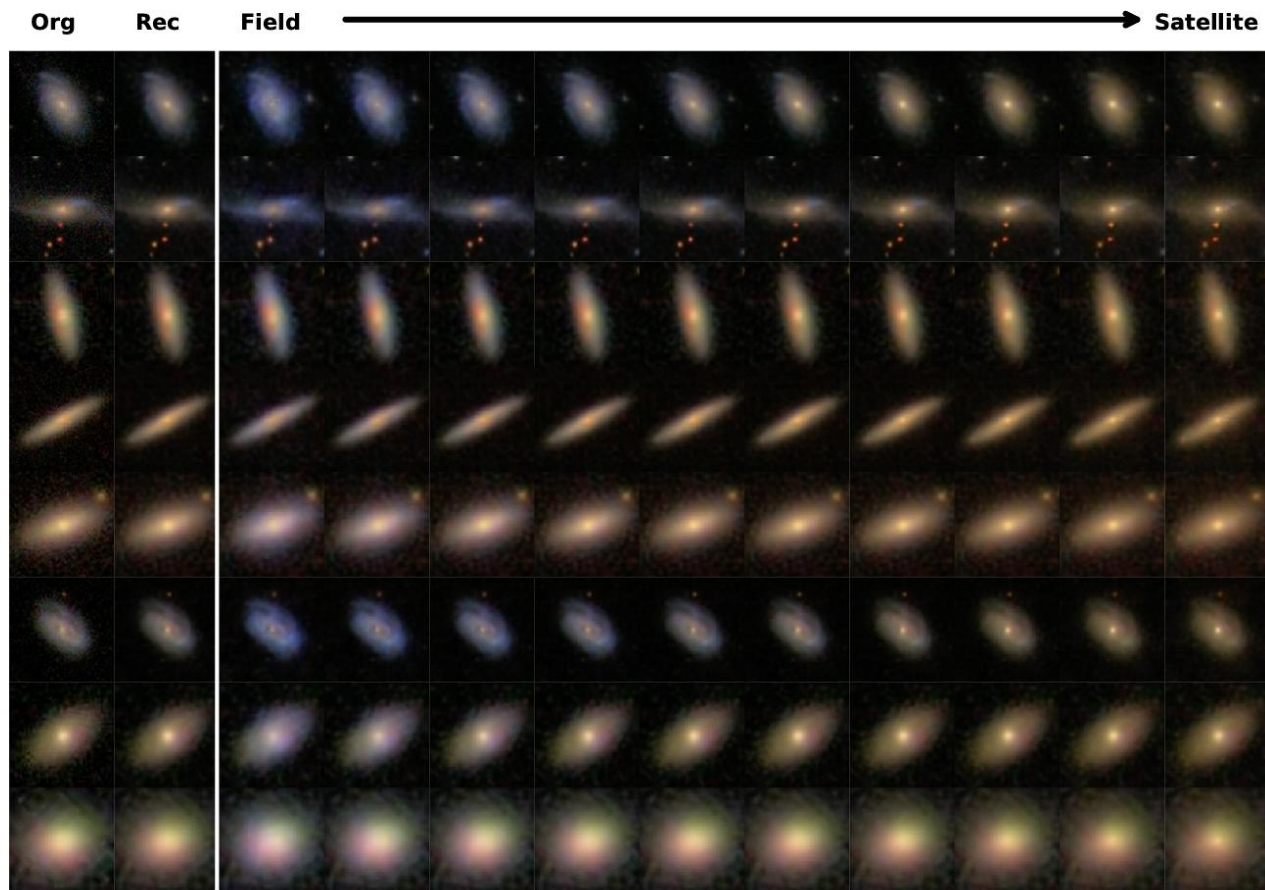
Decoded images by the model

Two different models, using CNN+VAE and CycleGAN we can increase the magnitude and S/N up to 8%

<https://arxiv.org/abs/2011.07124>

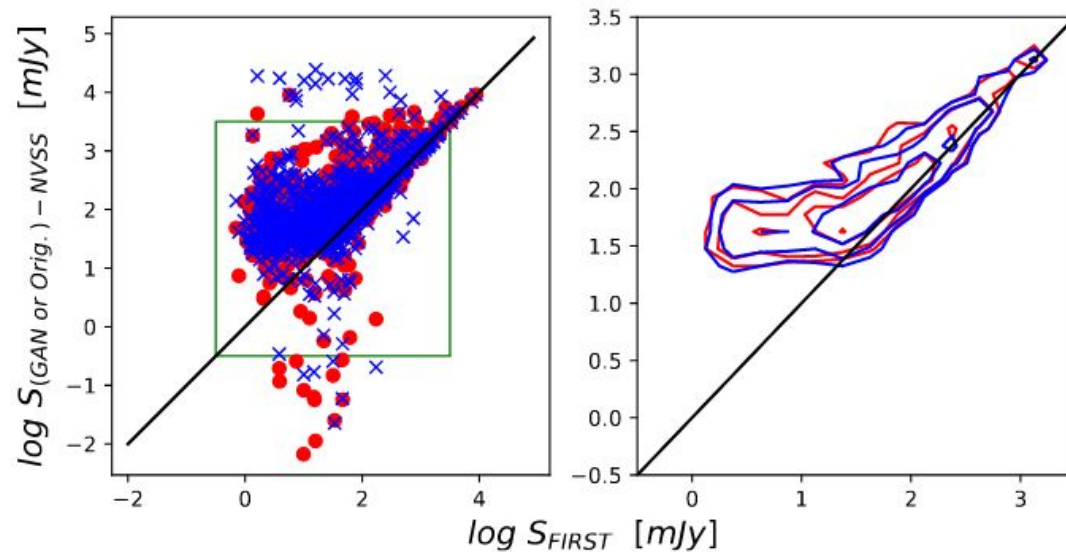
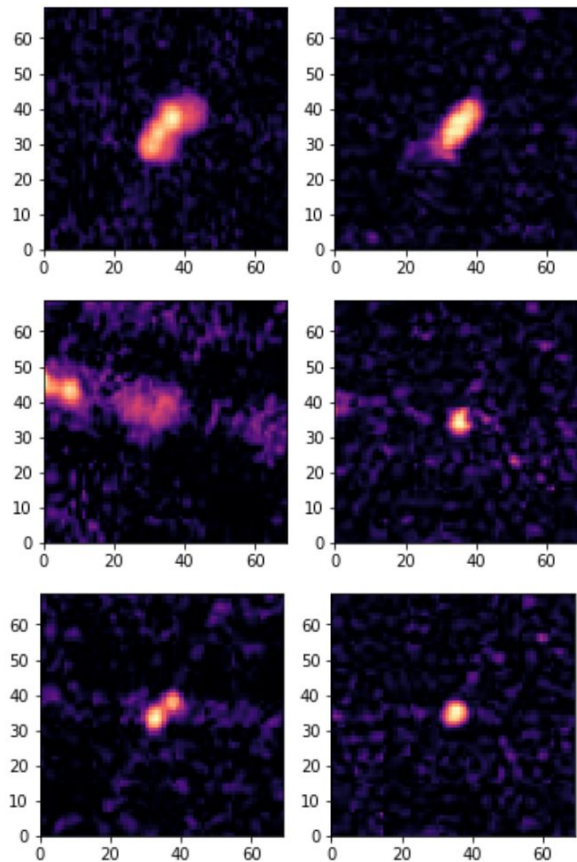
Buncher, Sharma,  
Carrasco-Kind, 2021.

# Galaxy Evolution



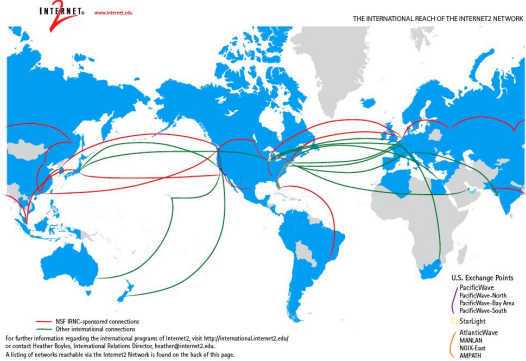
<https://arxiv.org/abs/1812.01114>

# Radio to Radio modeling

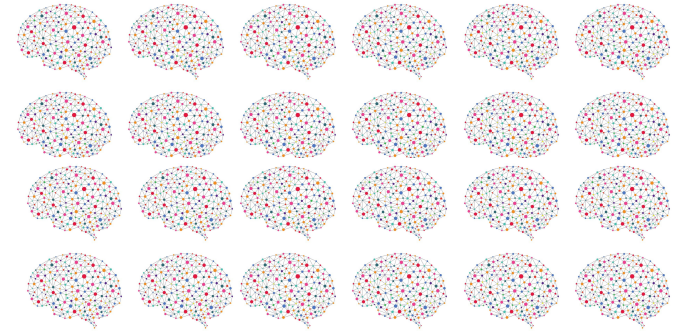
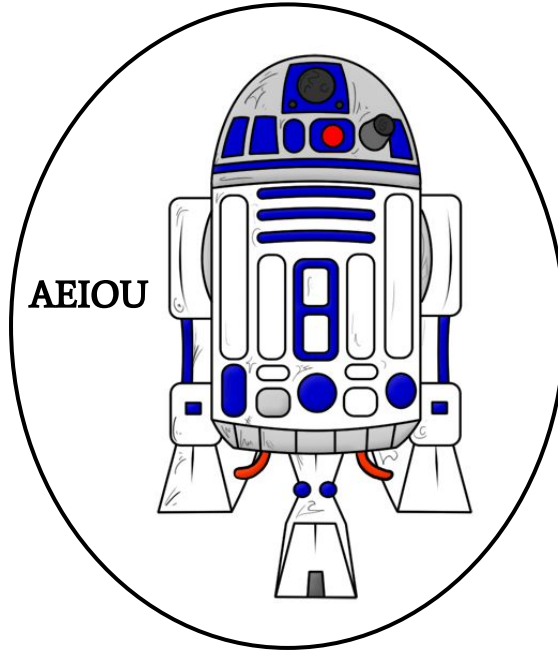


<https://arxiv.org/abs/1906.03874>

# Recap

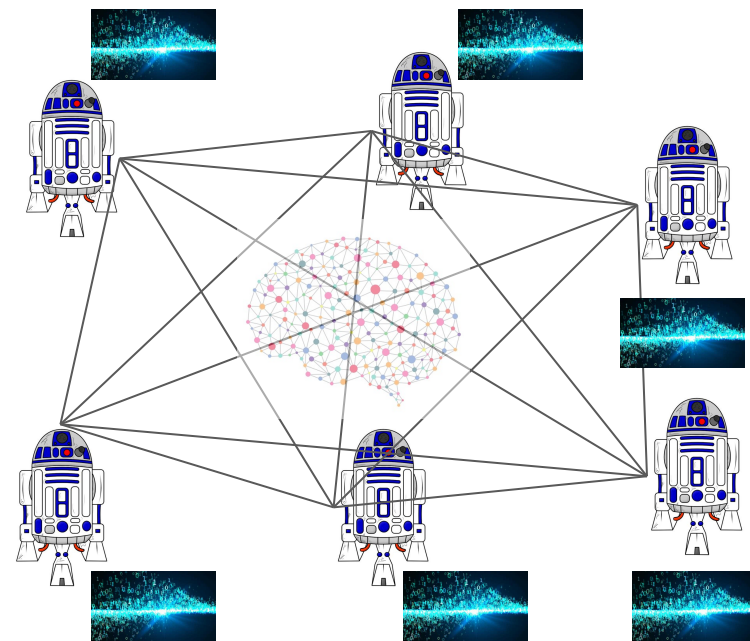


???

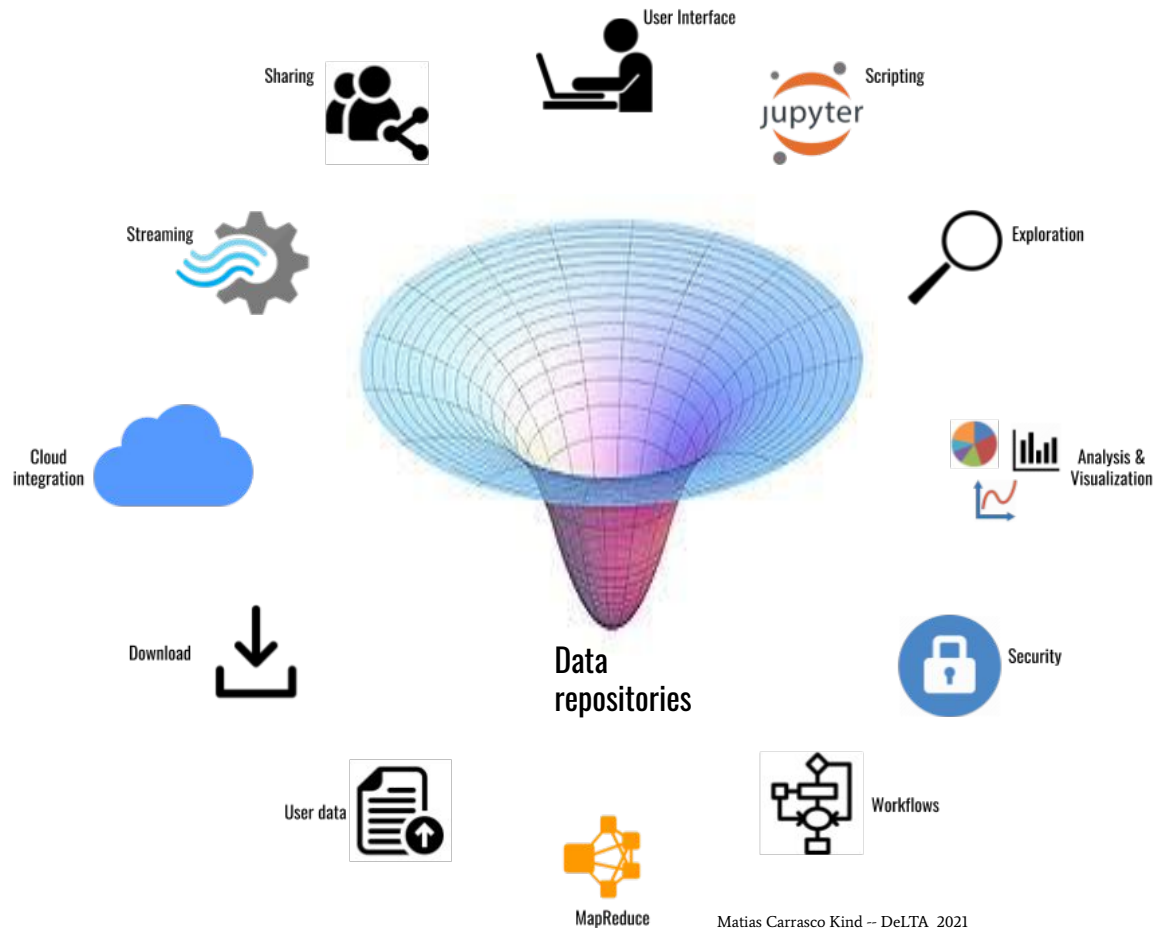


# Challenges

- CyberInfrastructure
- Data distributed (Exabytes) + Data Model
- Replication
- Serving Models (AI assisted)
  - Complexity (I/O)
- Reproducibility vs Explainability vs Interpretability
- Data Management
- Hardware
- Metadata and Bookkeeping
- Cloud computing
- Unified/Standard API (NLP assisted)
- Error modeling and propagation



# Data Gravity

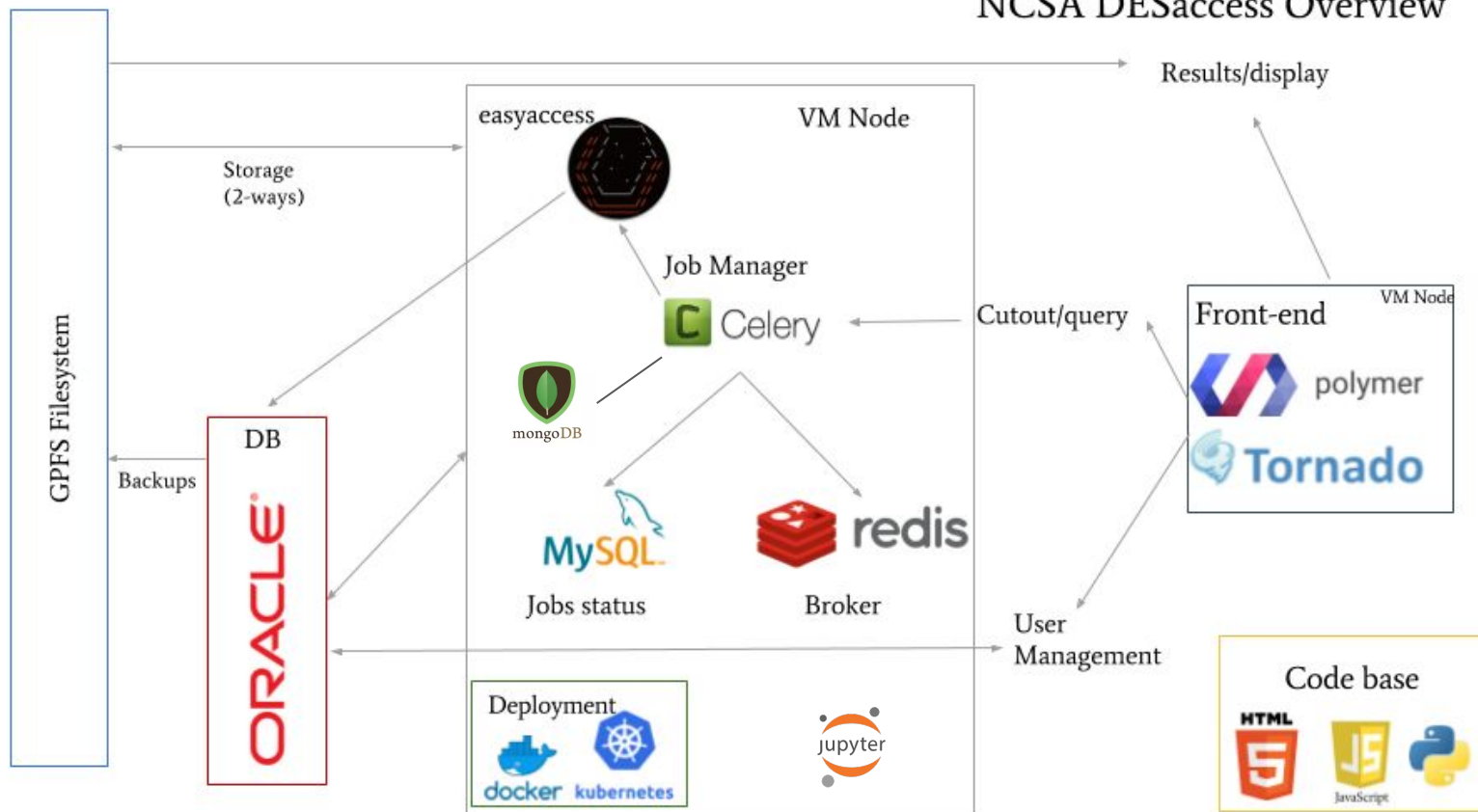


Several meanings around a central data archive, a.k.a “data lake”, **“data gravity”** repository with common components

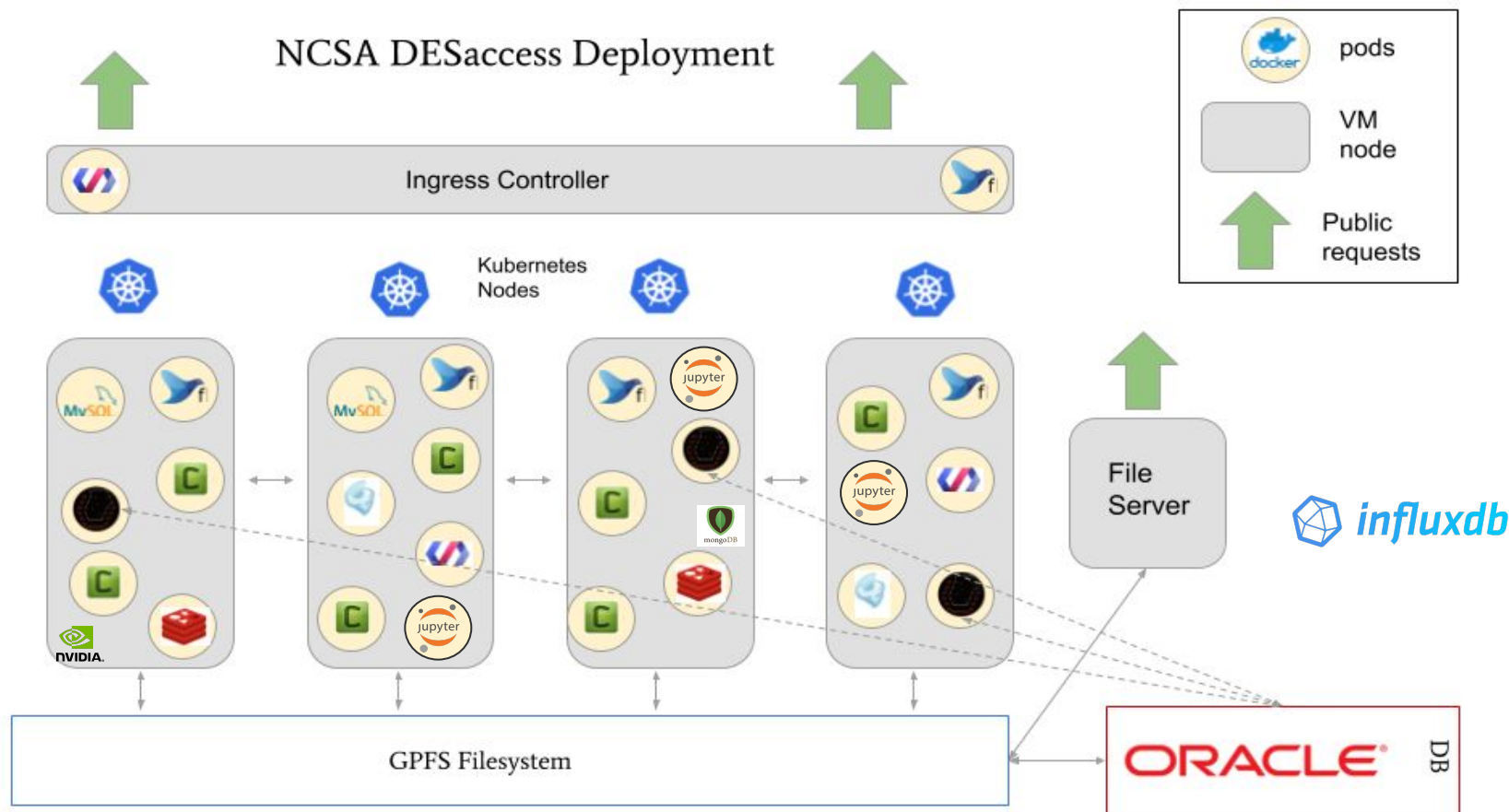
- Storage
- Security
- Retrieving
- Interacting
- Modifying
- Understanding

# NCSA DESaccess: Technology Overview

## NCSA DESaccess Overview



# NCSA DESaccess: Deployment (Hybrid)



# Final Remarks

---

Matias Carrasco Kind -- NCSA  
[mcarras2@illinois.edu](mailto:mcarras2@illinois.edu)  
[github.com/mgckind](https://github.com/mgckind)  
[matias-ck.com](https://matias-ck.com)

[go.ncsa.illinois.edu/delta21](https://go.ncsa.illinois.edu/delta21)

- We live in a era of exponential advances in computing, AI and CI
- We can start seeding these Extremely AI entities which would enable a faster science discovery
- Is a long but realistic path ahead and we can start now
- Can be replicated in other fields

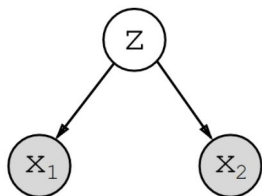
# Multimodal VAE: Training modalities

## Multimodal Generative Models for Scalable Weakly-Supervised Learning

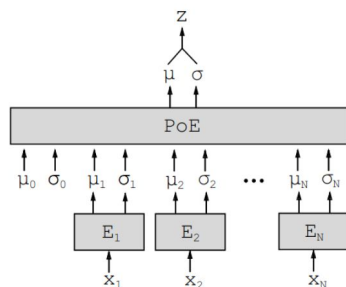
Mike Wu  
Department of Computer Science  
Stanford University  
Stanford, CA 94025  
wumike@stanford.edu

Noah Goodman  
Departments of Computer Science and Psychology  
Stanford University  
Stanford, CA 94025  
ngoodman@stanford.edu

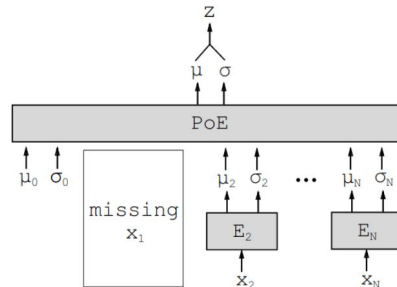
[multimodal-generative-models-for-scalable-weakly-supervised-learning](#)



(a)



(b)



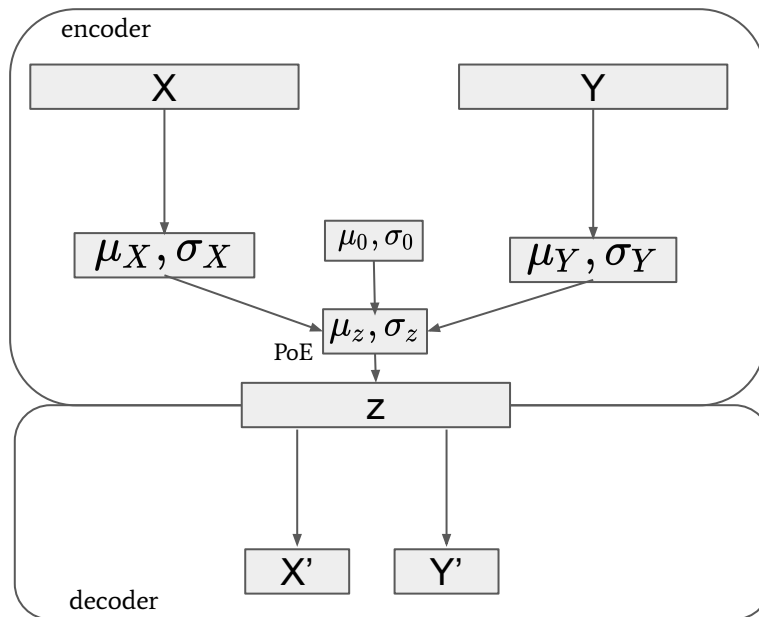
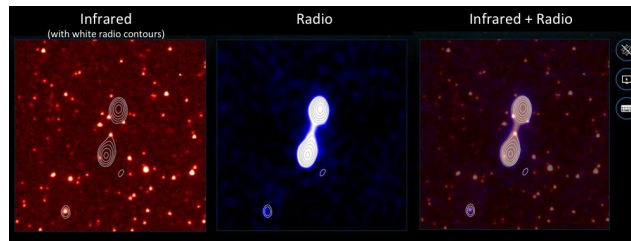
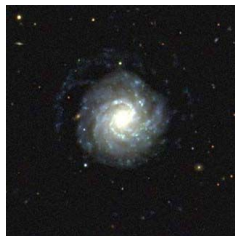
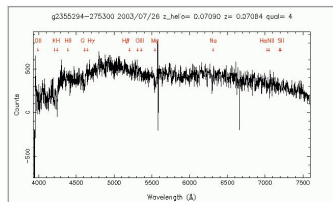
(c)

Learning joint representation of conditionally independent modalities using product of experts.

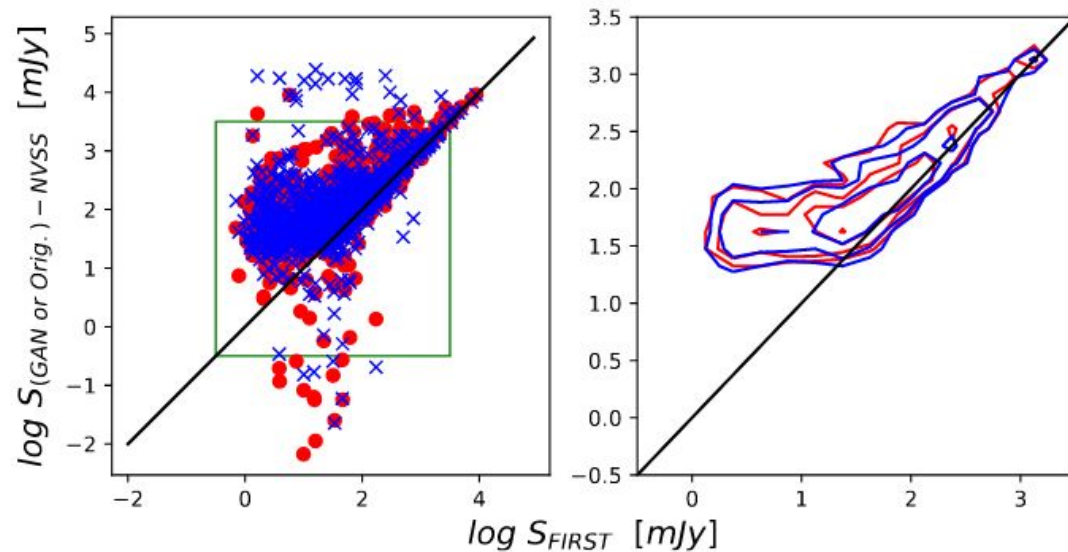
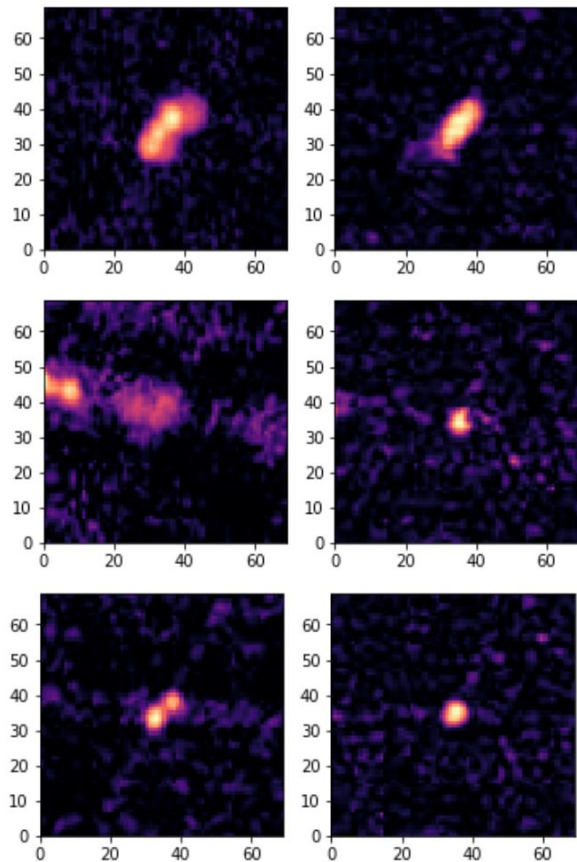
We can:

- Conditional sample with certain attributes
- Sample without any limitations
- Change the attribute of an existing input data
- Similarity search and anomaly detection
- Predict one modality from the others
- Sample and train with missing modalities

# MVAE: It opens interesting options

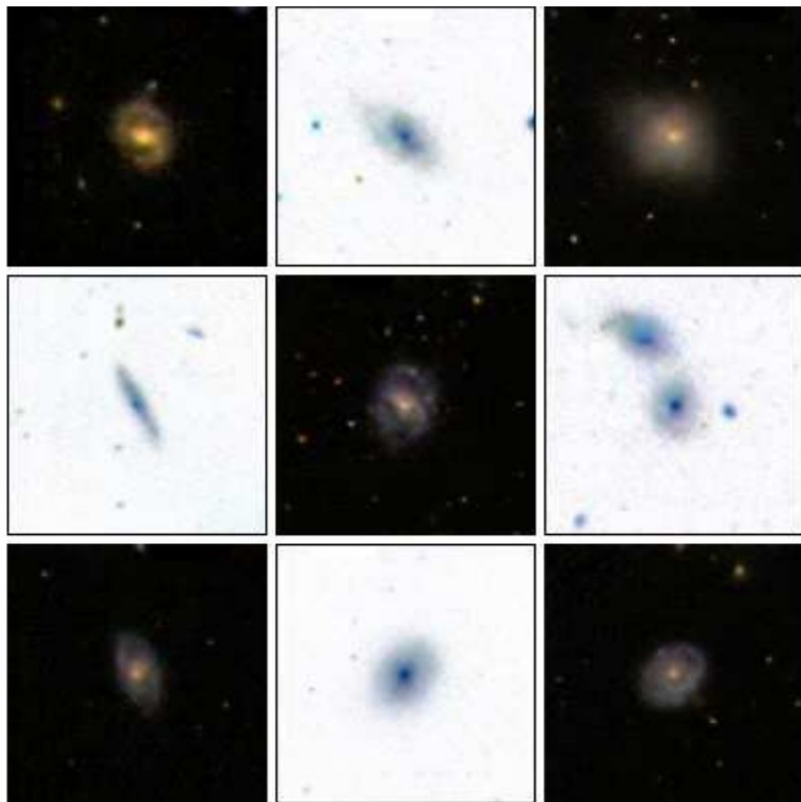


# Radio to Radio modeling

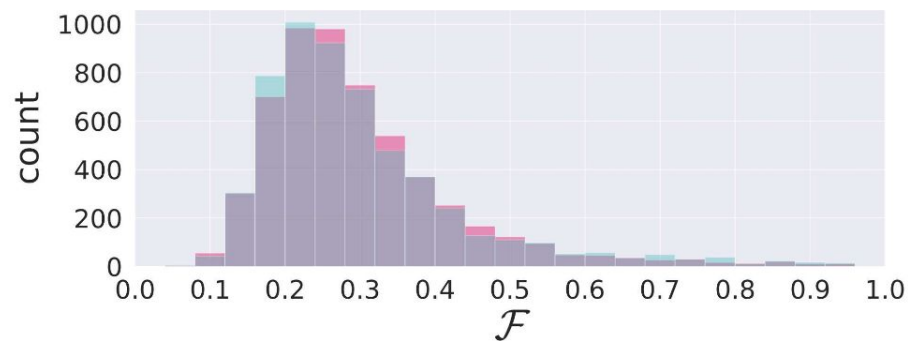


<https://arxiv.org/abs/1906.03874>

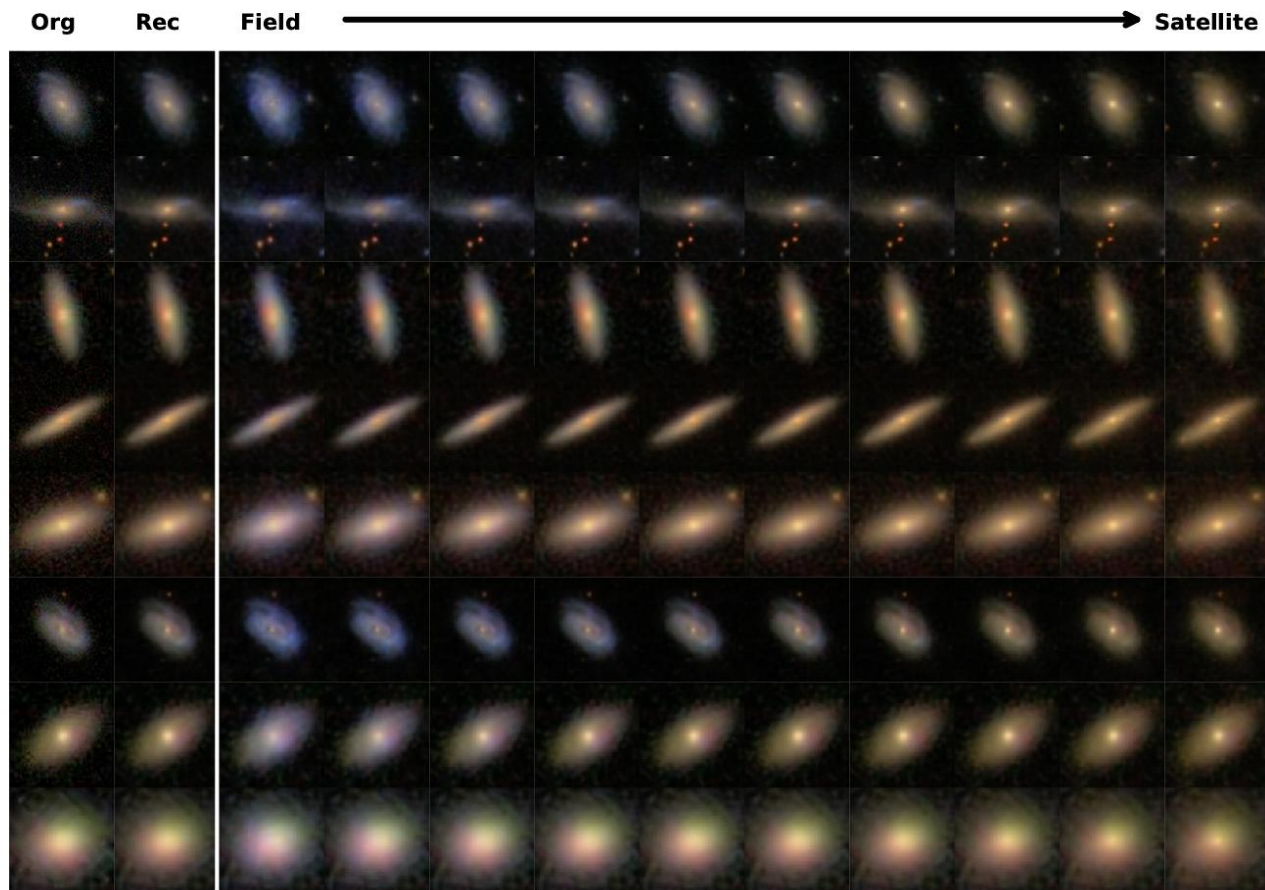
# Synthetic galaxy images



<https://arxiv.org/abs/1811.03081>



# Galaxy Evolution



<https://arxiv.org/abs/1812.01114>