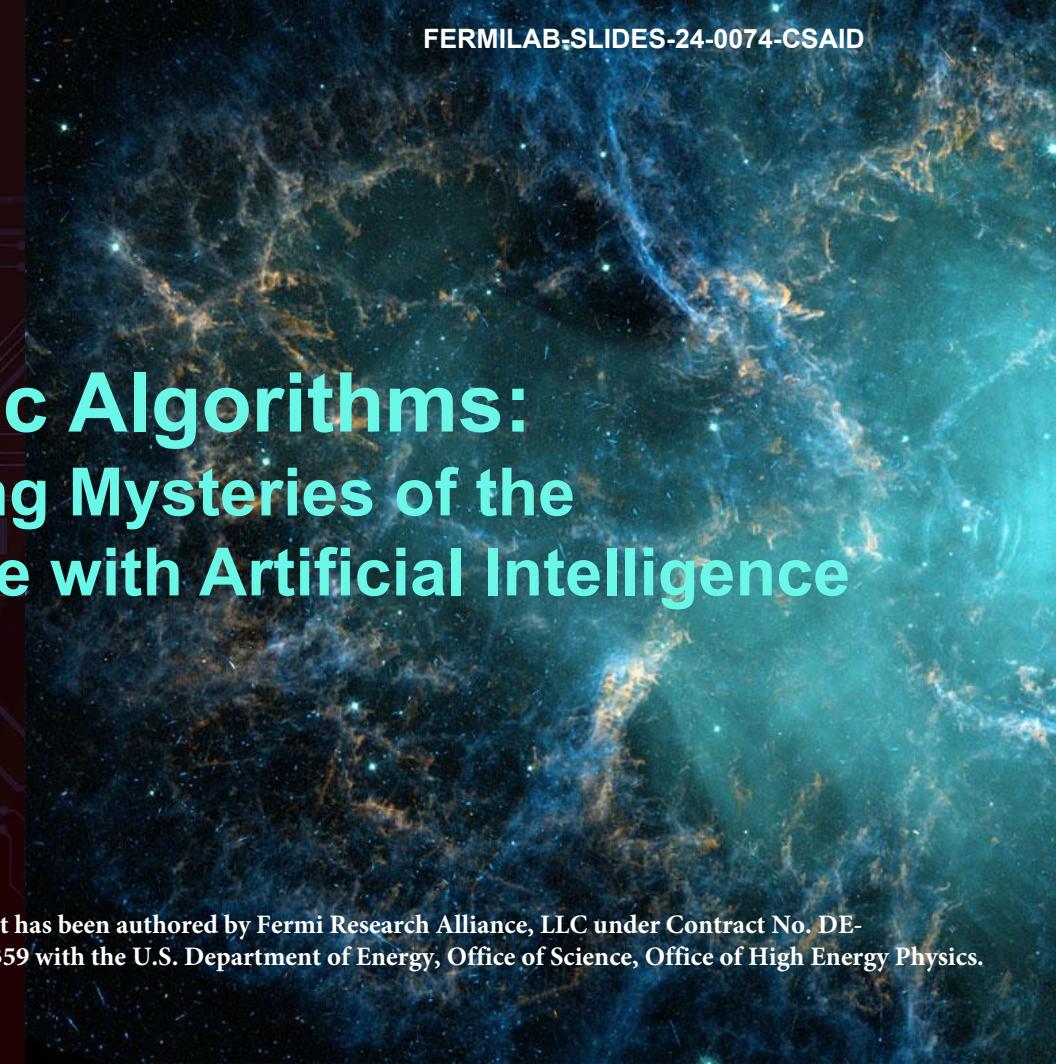




# Cosmic Algorithms: Unveiling Mysteries of the Universe with Artificial Intelligence

Aleksandra Ćiprijanović  
(she/her/hers)

Wilson Fellow Associate Scientist  
Fermilab  
aleksand@fnal.gov

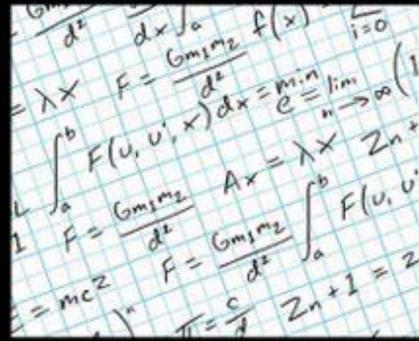


This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

# ASTROPHYSICIST



## what people think I do



### **what friends think I do**



## what Mom thinks I do



## what I think I do

```

def getParameter(parameter, database, view, *args):
    return convertSam_db.getParameter(database, view, *args)

def main():
    print sam_db.count('main', 'blueSample')
    #sam_db.createDBDBfile
    data = np.genfromtxt(DBfile, delimiter=',', comments='#', dtype = [( 'objID', '[S24]', 'float'), ( 'ra', float), ( 'dec', float), ('petroMag_g', float), ('petroMag_r', float), ('petroMag_i', float), ('petroMag_z', float), ('petroMagErr_g', float), ('petroMagErr_r', float), ('petroMagErr_i', float), ('petroMagErr_z', float), ('extinction_g', float), ('extinction_r', float), ('extinction_i', float), ('extinction_z', float), ('isgal', float), ('isQSO', float), ('lumExp_g', float), ('lumExp_r', float), ('lumDev_r', float), ('devRad_r', float)], names=True) #
    missing_val = np.where(np.isnan(data))
    print missing_val
    print plotData
    plotScatter(plotData, 'red', 'black', 'r_vs_z', 'Apparent magnitude vs. redshift', 'z', 'a')
    exit()

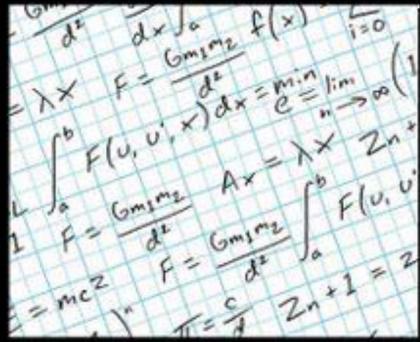
```

## what I really do

# ASTROPHYSICIST



what people think I do



what friends think I do



what Mom thinks I do

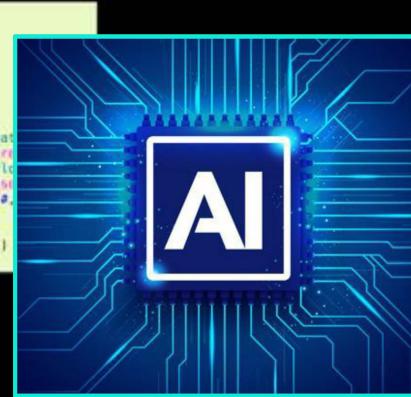


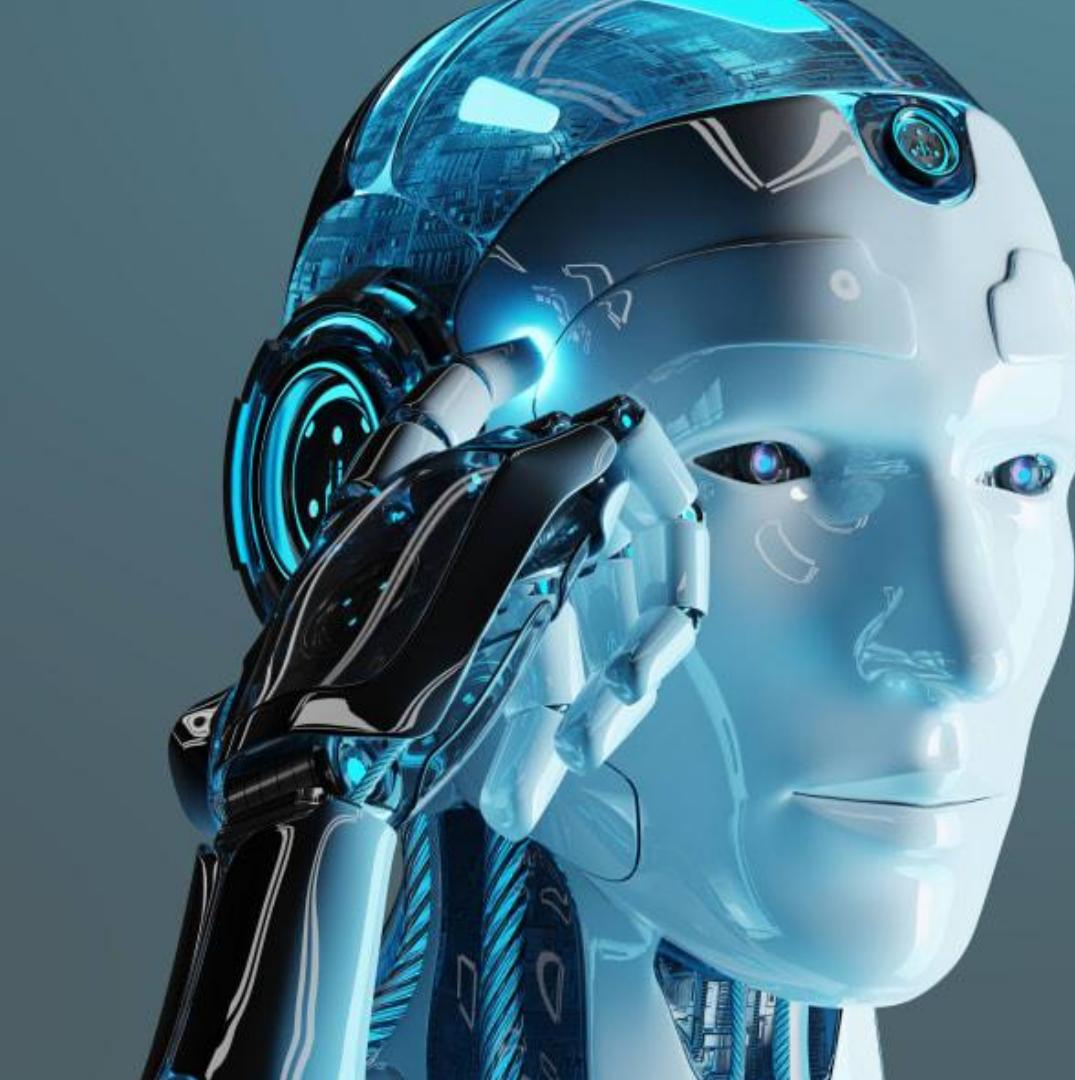
what I think I do

```
def getParameter(parameter, database, view, *args):
    return convertisam_db.getFromDB(parameter, database, view, *args)

def main():
    print sam_db.count('main', 'blueSample')
    sam_db.createDBfile()
    data = np.genfromtxt(DBfile, delimiter=',', comments='%', dtype=[('objID', 'i524'), ('ra', float), ('petroMag_u', float), ('petroMag_g', float), ('petroMag_r', float), ('petroMag_i', float), ('petroMag_z', float), ('petroMagErr_u', float), ('petroMagErr_g', float), ('petroMagErr_r', float), ('petroMagErr_i', float), ('petroMagErr_z', float), ('extinction_r', float), ('extinction_z', float), ('isgal_r', float), ('isgal_z', float), ('latExp_r', float), ('expmag_r', float), ('lnLBr_r', float), ('devMag_r', float)], names=True)
    plotData = [data[:,1], data[:,11]]
    plotData[1] = np.log10(plotData[1])
    print plotData
    plotScatter(plotData, 'red', 'black', 'r vs z', 'Apparent Magnitude vs. redshift', 'z', 'r')
    exit()
```

what I really do





# What is Artificial Intelligence (AI)?



# What is Artificial Intelligence (AI)?



OpenAI

Research ▾

Product ▾

Safety

Company ▾

## Introducing ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

[Try ChatGPT ↗](#)

[Read about ChatGPT Plus](#)





Can you help me prepare a lecture about AI in Astrophysics?



Can you help me prepare a lecture about AI in Astrophysics?



Certainly, I'd be happy to help you prepare a lecture about AI in astrophysics. Here's an outline of the lecture that you can use as a starting point:





Can you help me prepare a lecture about AI in Astrophysics?



Certainly, I'd be happy to help you prepare a lecture about AI in astrophysics. Here's an outline of the lecture that you can use as a starting point:



I. Introduction

II. How AI is used in Astrophysics

III. Case Studies

IV. Challenges and Opportunities

V. Conclusion



Can you help me prepare a lecture about AI in Astrophysics?



Certainly, I'd be happy to help you prepare a lecture about AI in astrophysics. Here's an outline of the lecture that you can use as a starting point:



I. Introduction

II. How AI is used in Astrophysics

III. Case Studies

IV. Challenges and Opportunities

V. Conclusion

You can add more details to each section and include additional case studies or examples as appropriate. You may also want to include visuals such as images or animations to help illustrate key points. Good luck with your lecture!

# Talk Outline

Introduction

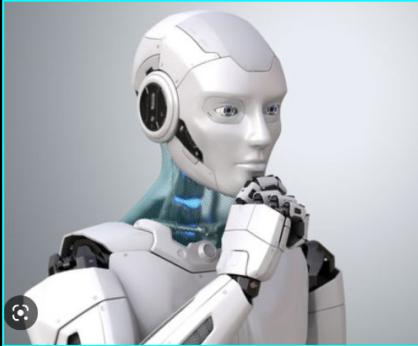
How is AI used in Astrophysics?

Case Studies

Challenges and Opportunities

Conclusion

So, if not



what is AI?

So, if not

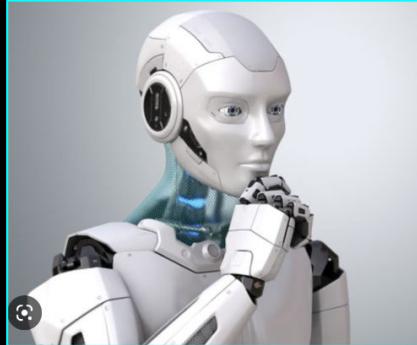


what is AI?

## Artificial Intelligence

Machines/software that  
**performs task that  
usually require human  
intelligence:**  
learning, perception,  
decision-making....

So, if not



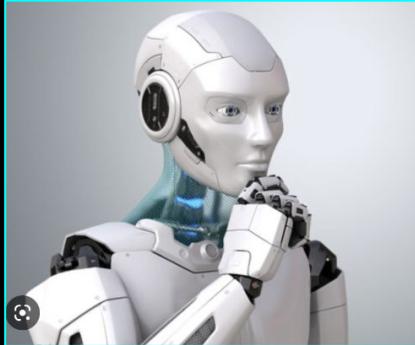
what is AI?

Artificial Intelligence

Machine Learning

Use data to **learn without explicit rules** and improve over time.

So, if not



what is AI?

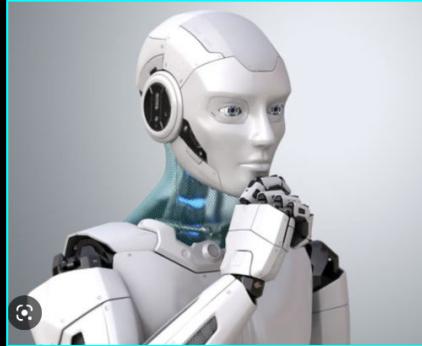
Artificial Intelligence

Machine Learning

Deep Learning

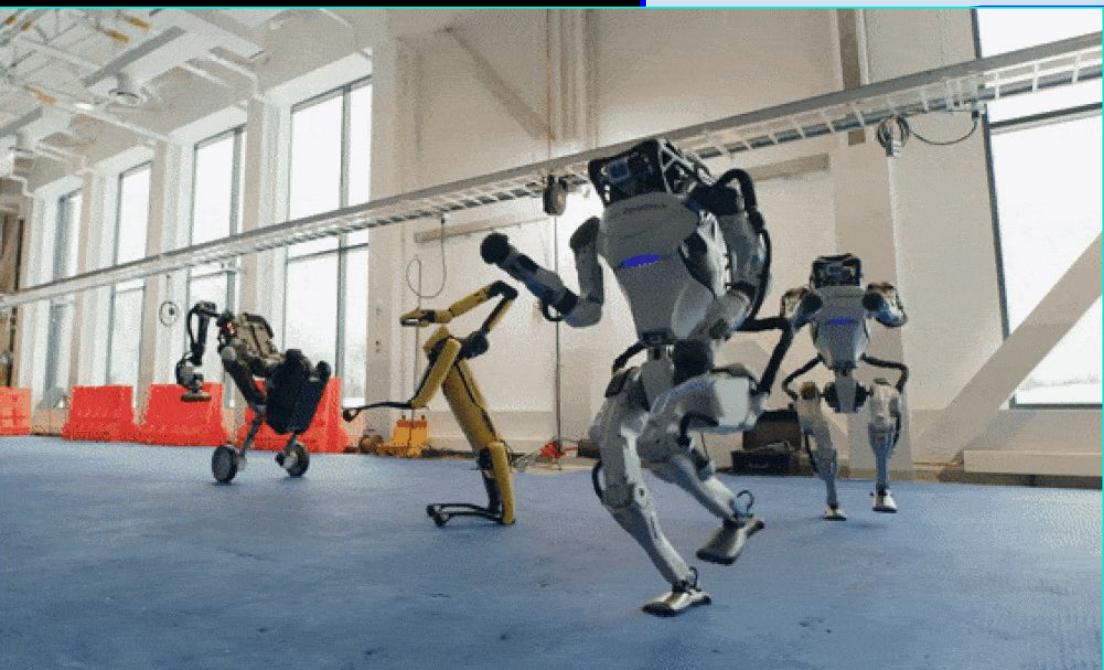
Learn **complex tasks using deep neural networks** - image and video analysis, speech recognition, language translation, autonomous vehicles, robotics...

So, if not



what is AI?

Artificial Intelligence



Machine Learning

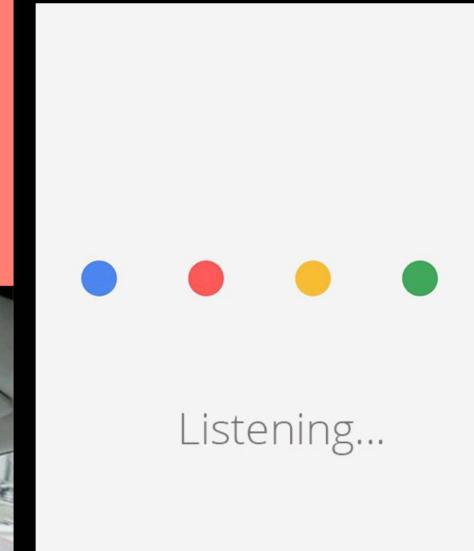
Deep Learning

Learn **complex tasks using deep neural networks** - image and video analysis, speech recognition, language translation, autonomous vehicles, robotics...









# AI methods are inspired by neurons in our brains.

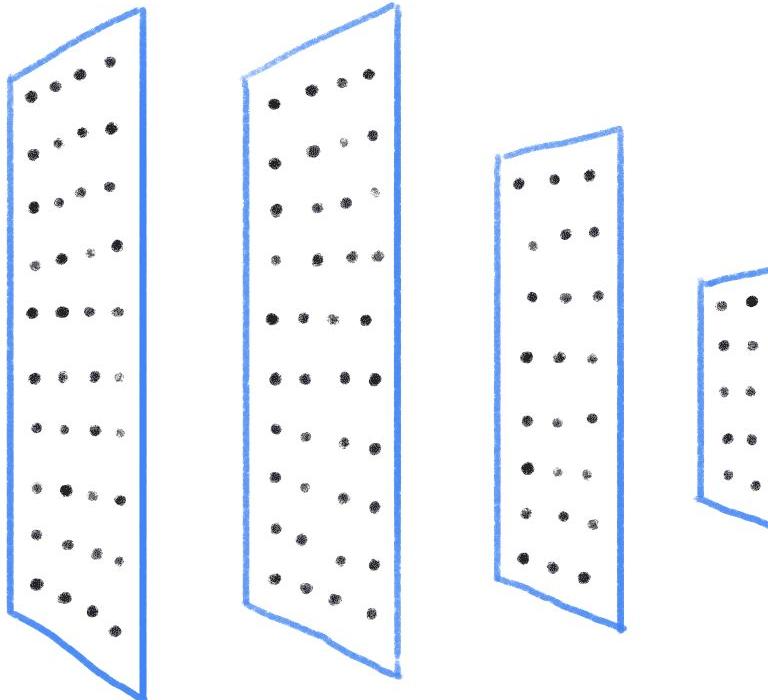
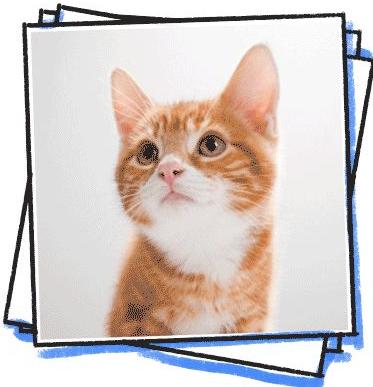


Models can have **millions of trainable parameters**.  
Trained with **thousands of example images**.

CAT

(LABELED)  
PHOTOS

DOG



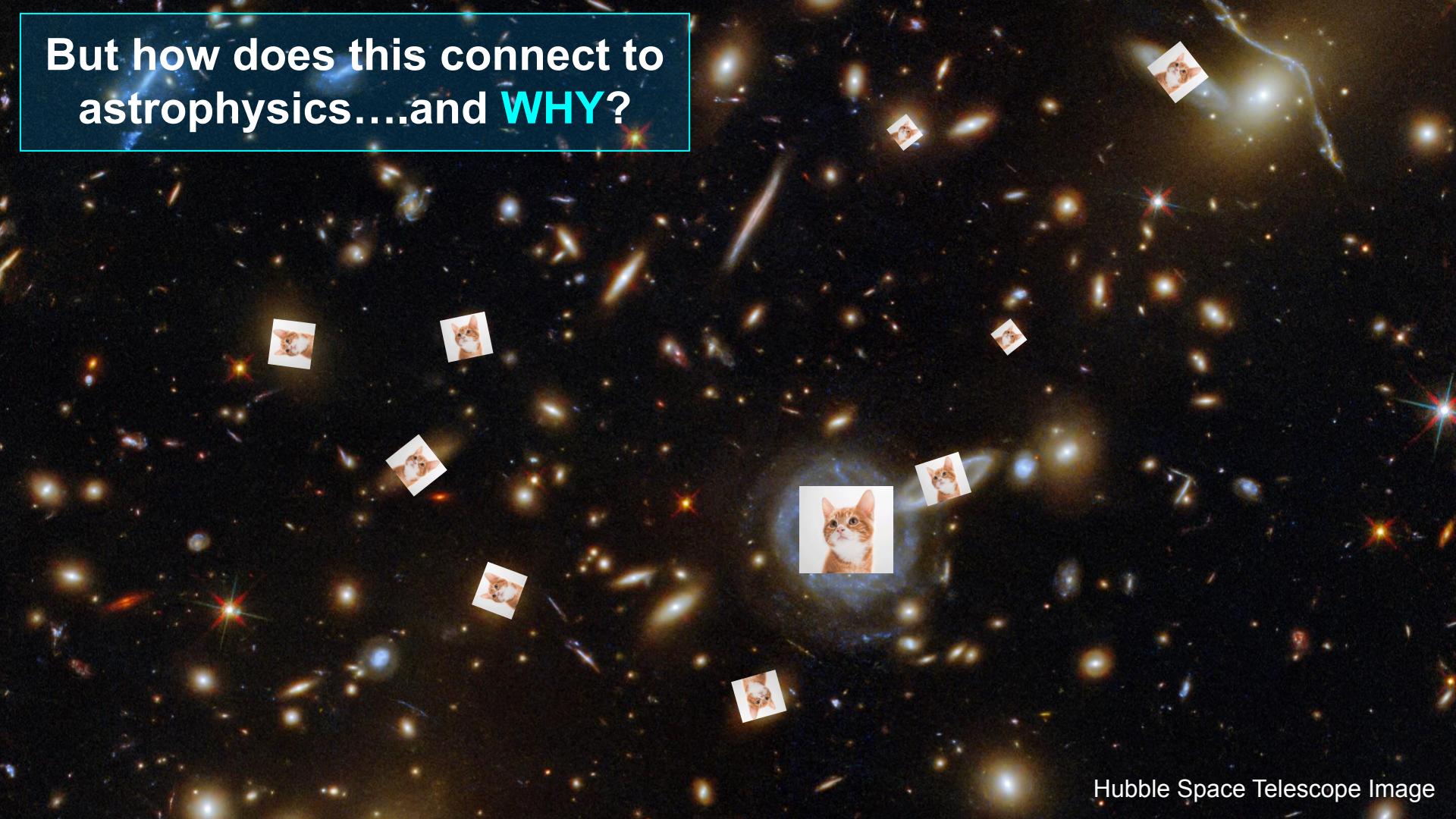
OUTPUT

But how does this connect to astrophysics....and **WHY?**



Hubble Space Telescope Image

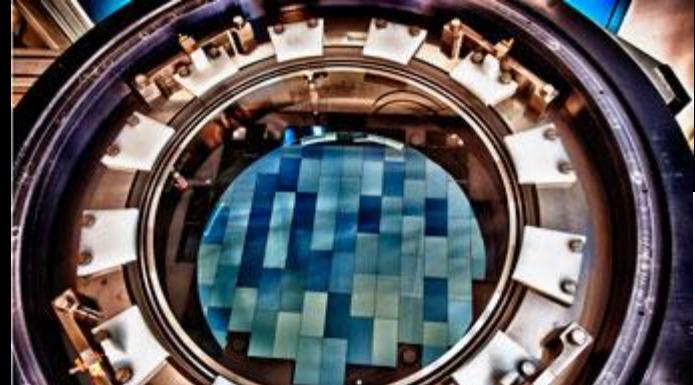
But how does this connect to  
astrophysics....and **WHY?**



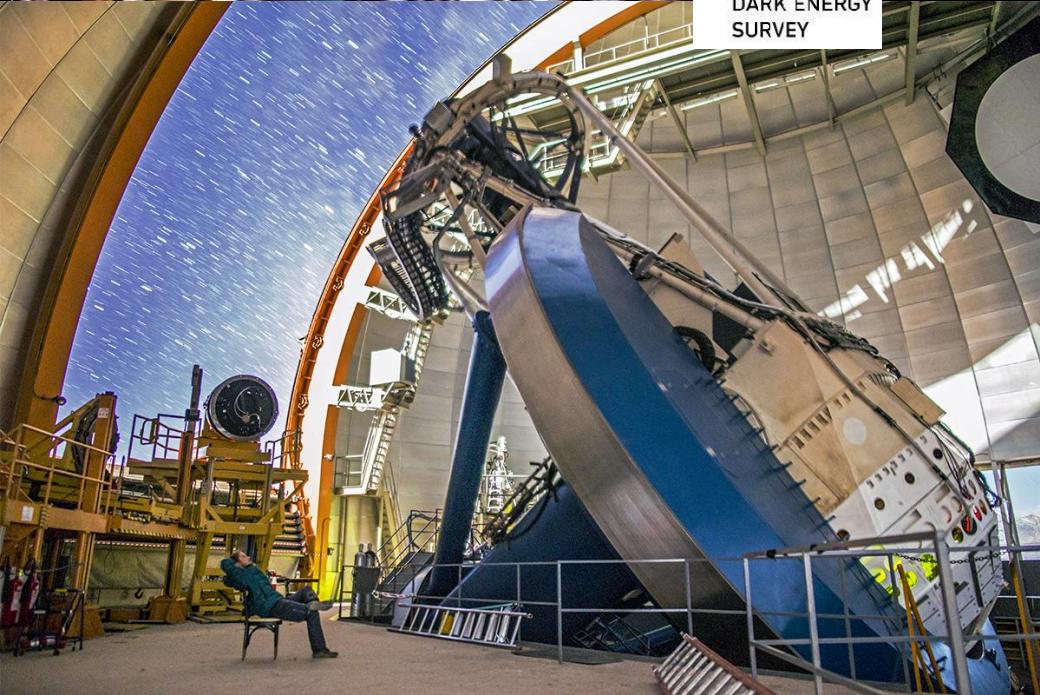
Hubble Space Telescope Image



iPhone camera



Our Camera!!!

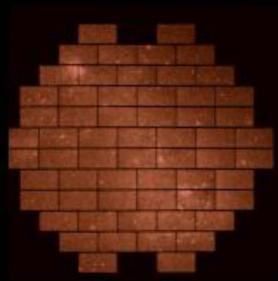


iPhone camera

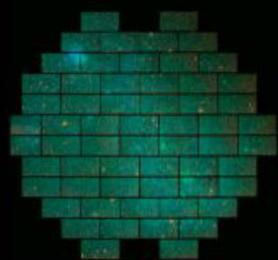
Victor M. Blanco 4-meter Telescope at the **Cerro Tololo Inter-American Observatory (CTIO)** in the Chilean Andes

- iPhone  
Camera

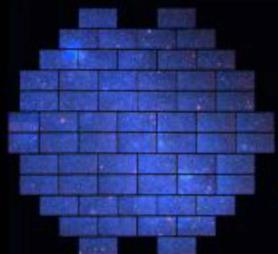




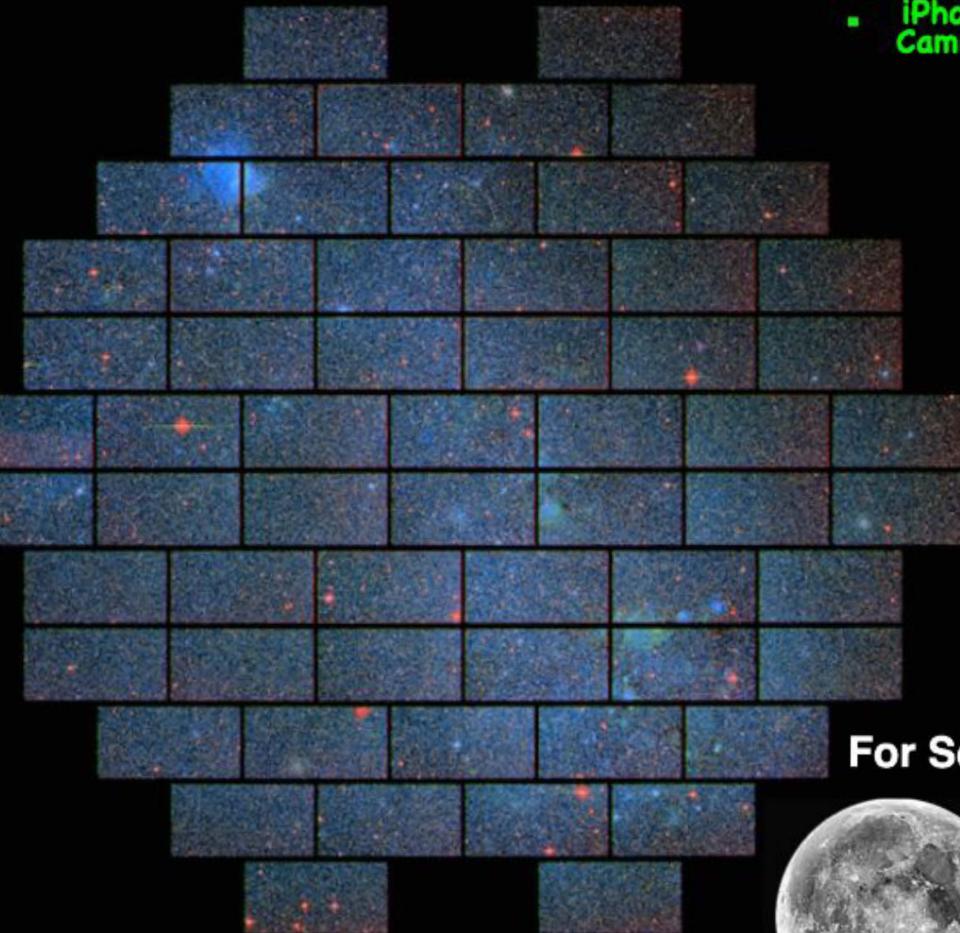
+



+



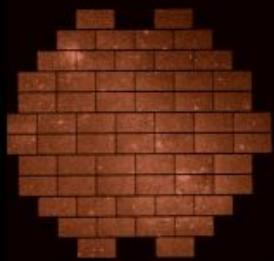
=



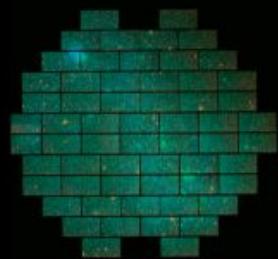
+ 2 more filters

▪ iPhone  
Camera

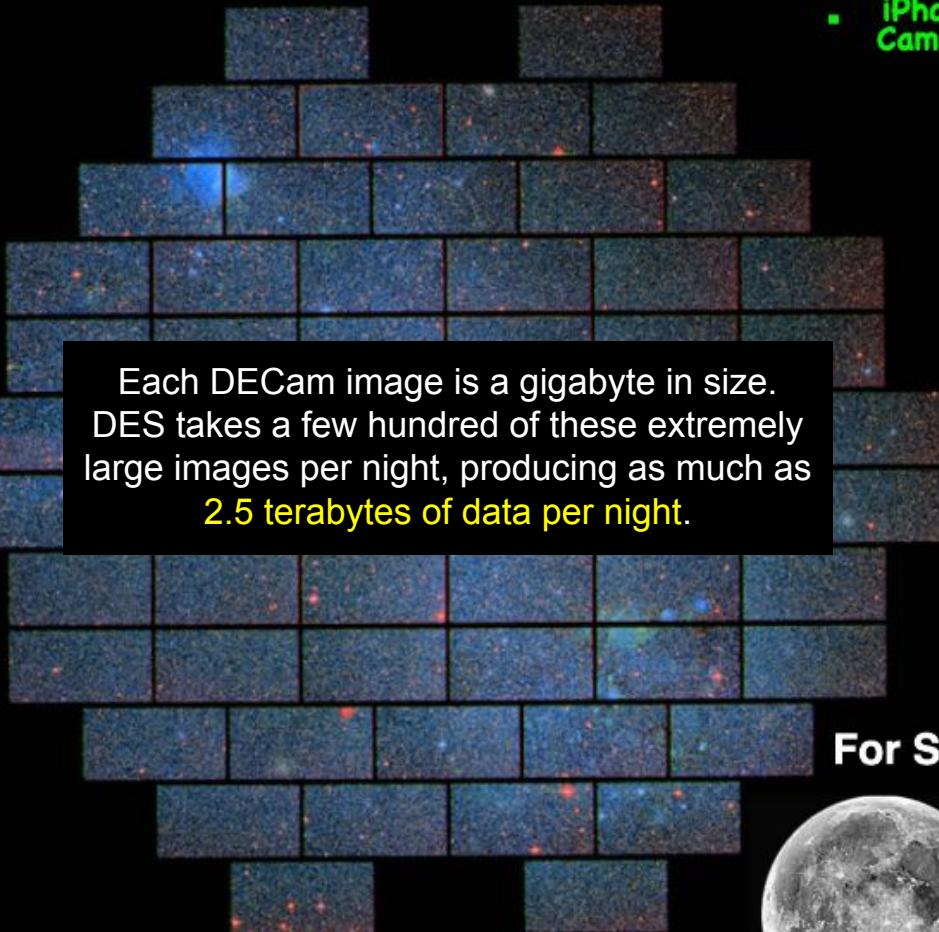




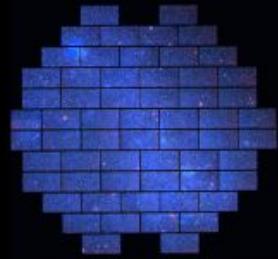
+



=



Each DECam image is a gigabyte in size.  
DES takes a few hundred of these extremely  
large images per night, producing as much as  
2.5 terabytes of data per night.

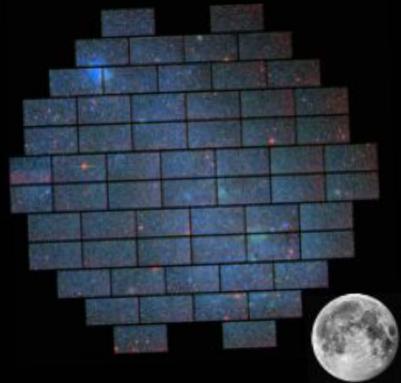


+

+ 2 more filters

For Scale



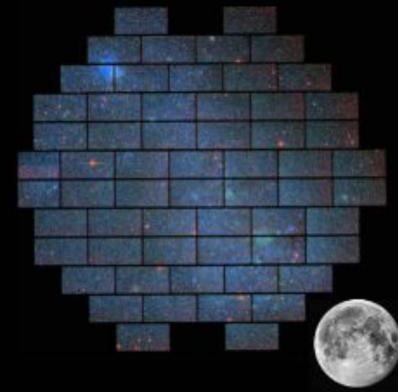
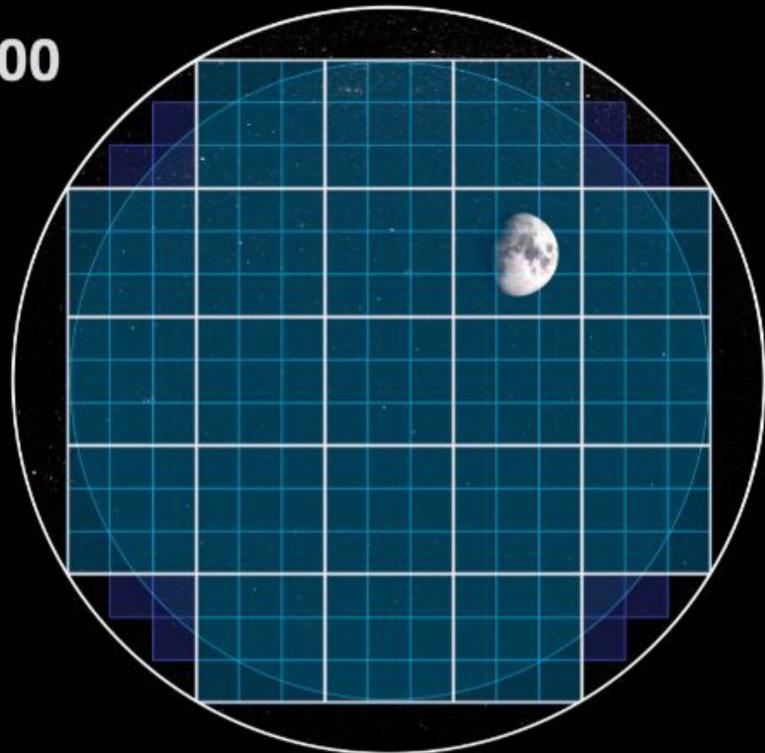


**DECam**  
**570,000,000**  
**pixels**

■  
**HDTV**  
**2,000,000**  
**pixels**

# LSSTCam

**3,200,000,000  
pixels**



■  
**HDTV  
2,000,000  
pixels**

Need 1,500 HDTV screens to view a **single image** taken with LSSTCam

Enough 4k TVs to cover half a basketball court

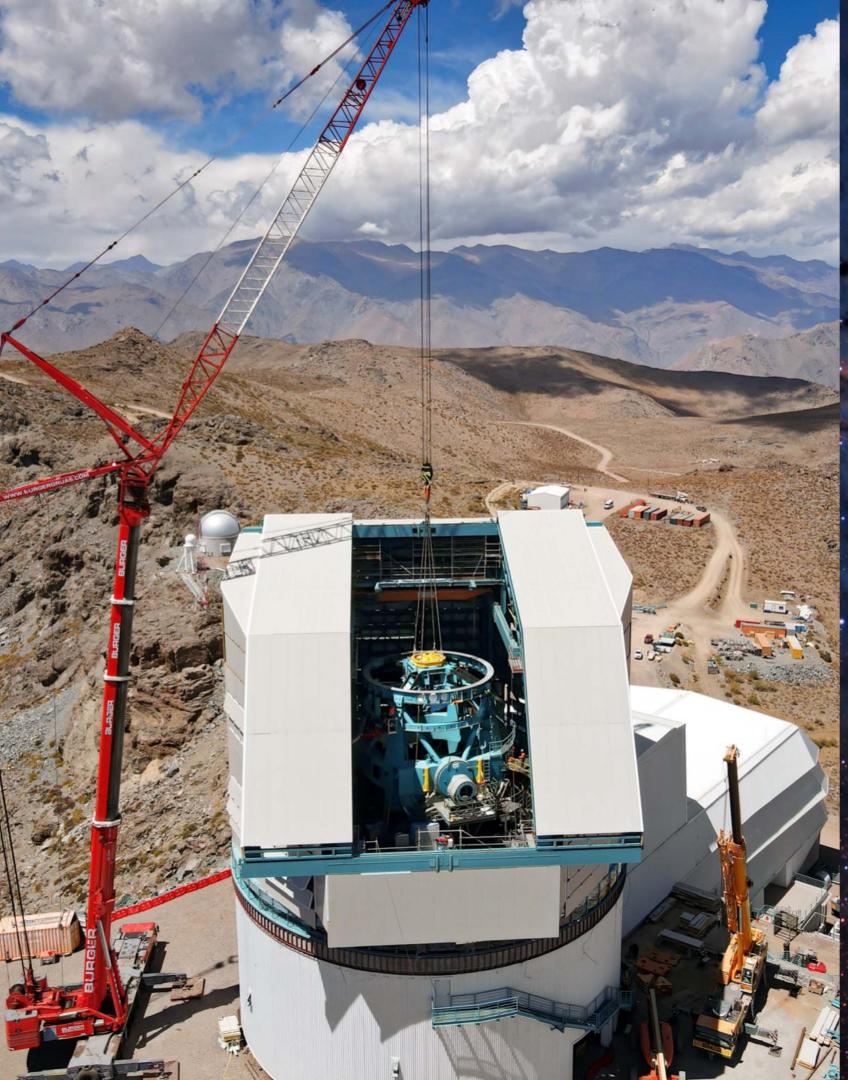


**El Peñón peak of Cerro Pachón in northern Chile**, alongside the existing Gemini South and Southern Astrophysical Research Telescope.



**El Peñón peak of Cerro Pachón in northern Chile**, alongside the existing Gemini South and Southern Astrophysical Research Telescope.





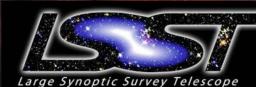
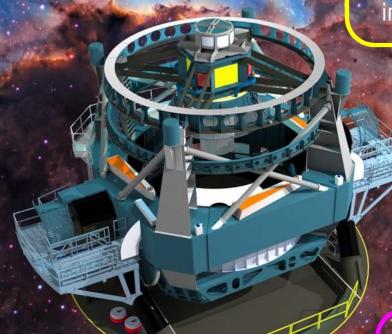
# LSST Project in Numbers

8,4 meters  
Primary mirror diameter

3 200 Megapixels  
Resolution of the Telescope Camera

1.23 F/D  
Telescope aperture

800 times  
Number of times a same object will be captured



15 TB  
Amount of data collected every night

3 Nights  
Time needed for an all-sky imaging

15 seconds  
Exposure time needed to capture an image

37 Billion  
Number of celestial objects detected after 10 years

Credit Image: Rolf Wahl Olsen  
[www.rolfolsenastrophotography.com](http://www.rolfolsenastrophotography.com)



Source: [www.astrospace-page.blogspot.com](http://www.astrospace-page.blogspot.com)

# But that is not all ..... Not even close!







How do we...

- ... process the data?
- ... search through the data?
- ... learn from the data?
- ... find unusual objects?
- ... send alerts quickly?
- ... combine/compare datasets?
- ...
- ...
- ...



How do we...

- ... process the data?
- ... search through the data?
- ... learn from the data?
- ... find unusual objects?
- ... send alerts quickly?
- ... combine/compare datasets?



# Talk Outline

Introduction

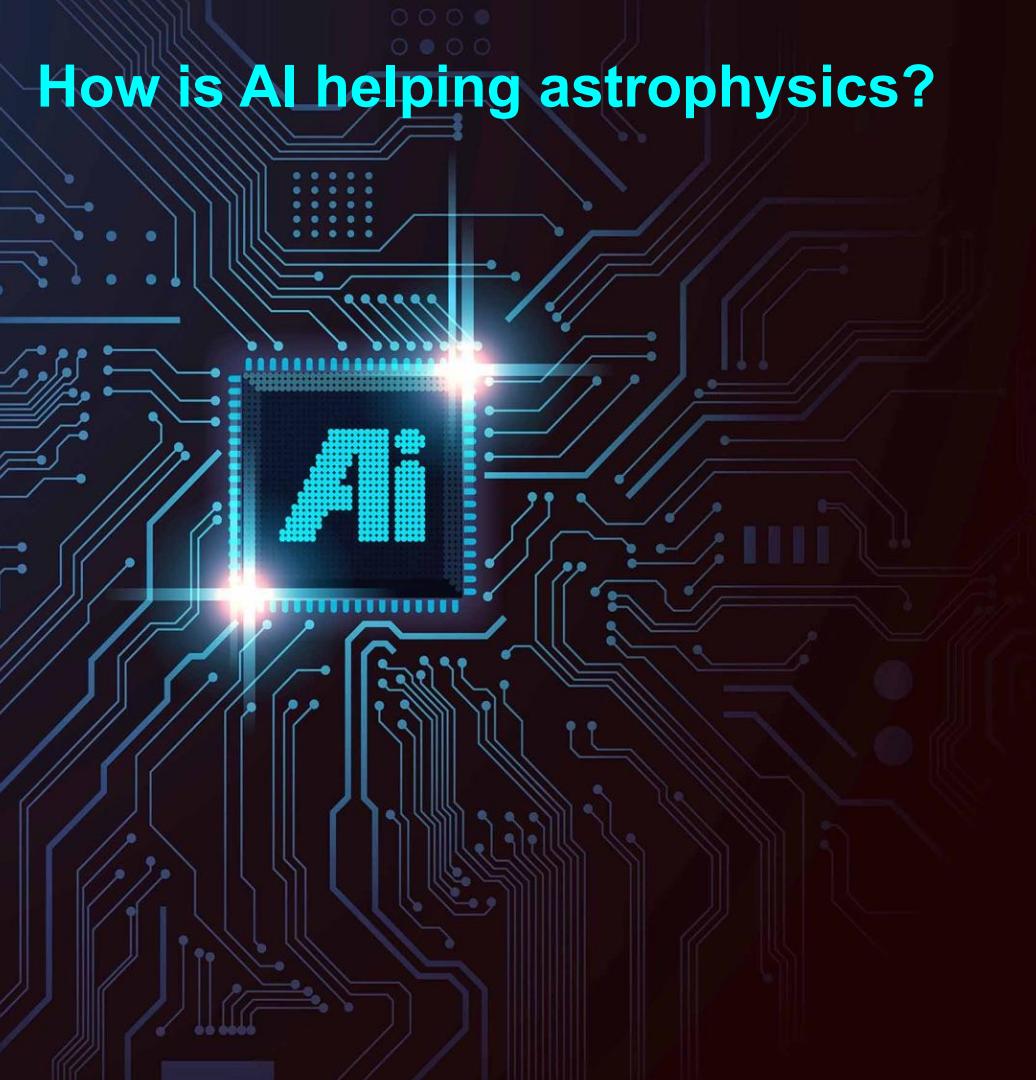
How is AI used in Astrophysics?

Case Studies

Challenges and Opportunities

Conclusion

# How is AI helping astrophysics?



# How is AI helping astrophysics?

Image/data processing and analysis

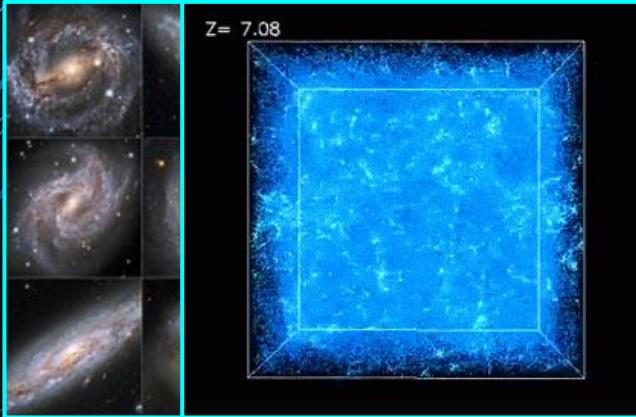


AI

# How is AI helping astrophysics?

Image/data processing and analysis

Speed up simulations



# How is AI helping astrophysics?



Image/data processing and analysis

Speed up simulations

Scheduling and operations

# How is AI helping astrophysics?



Hubble: supernova in the spiral galaxy NGC 2525

Image/data processing and analysis

Speed up simulations

Scheduling and operations

Alert systems and real-time analysis

# Talk Outline

Introduction

How is AI used in Astrophysics?

Case Studies

Challenges and Opportunities

Conclusion

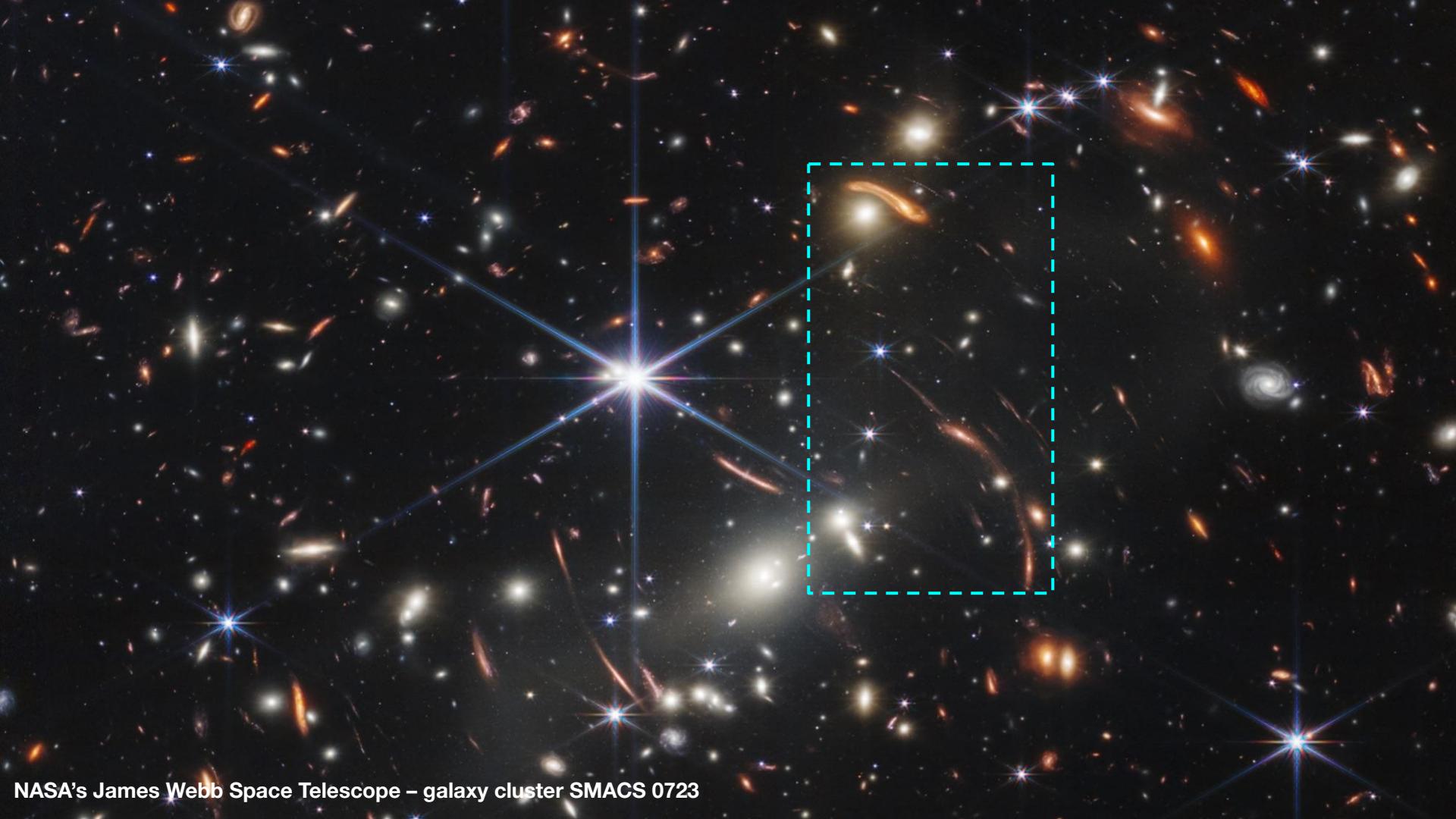
# Example 1



NASA's James Webb Space Telescope – galaxy cluster SMACS 0723



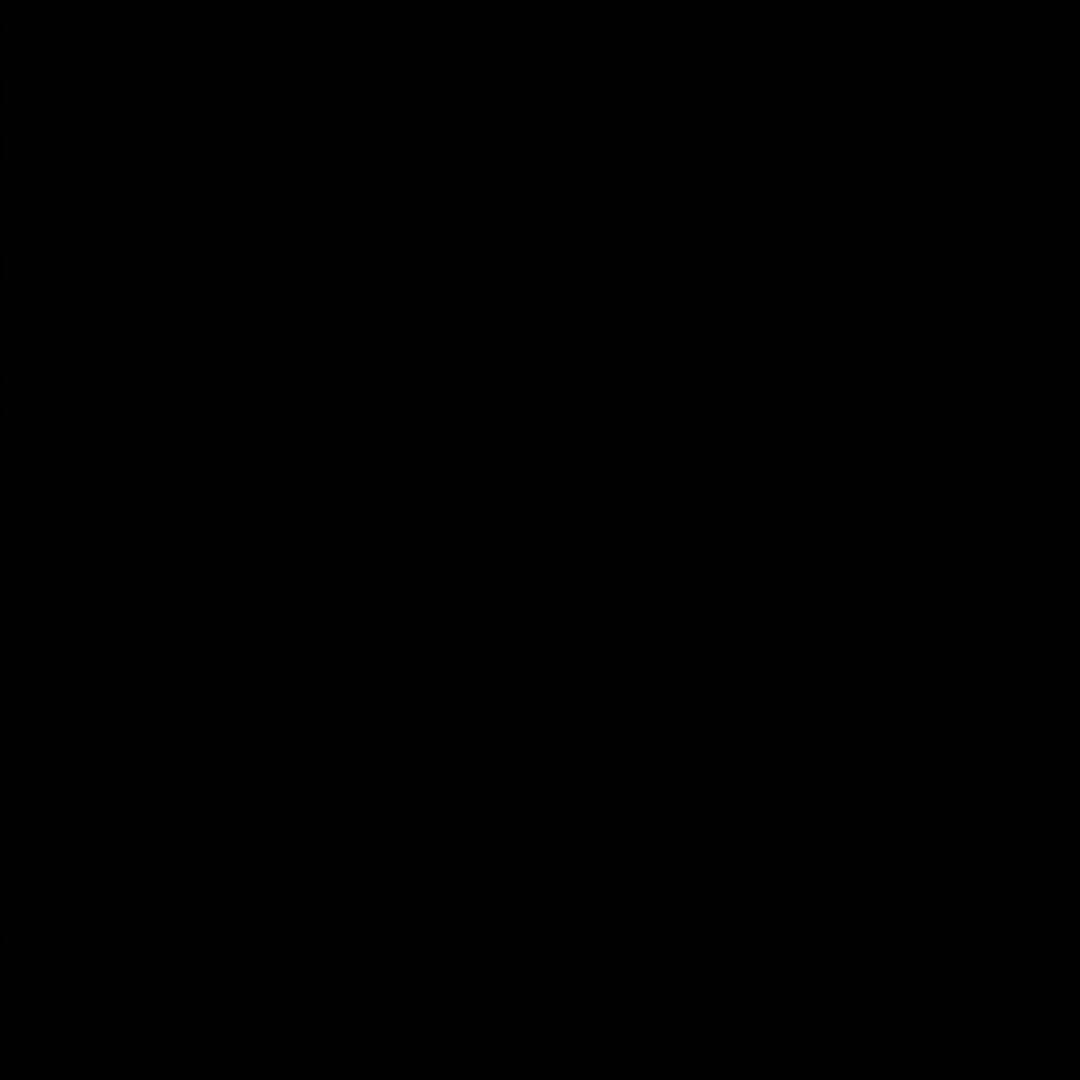
NASA's James Webb Space Telescope – galaxy cluster SMACS 0723

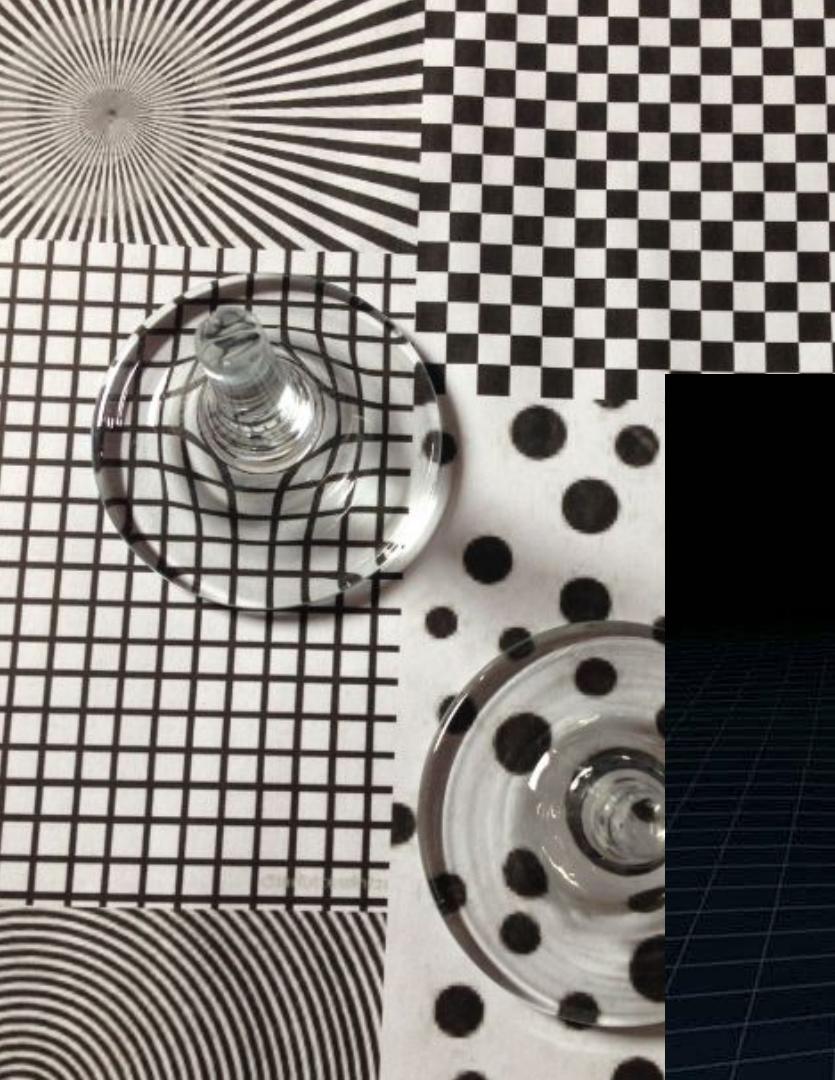


NASA's James Webb Space Telescope – galaxy cluster SMACS 0723



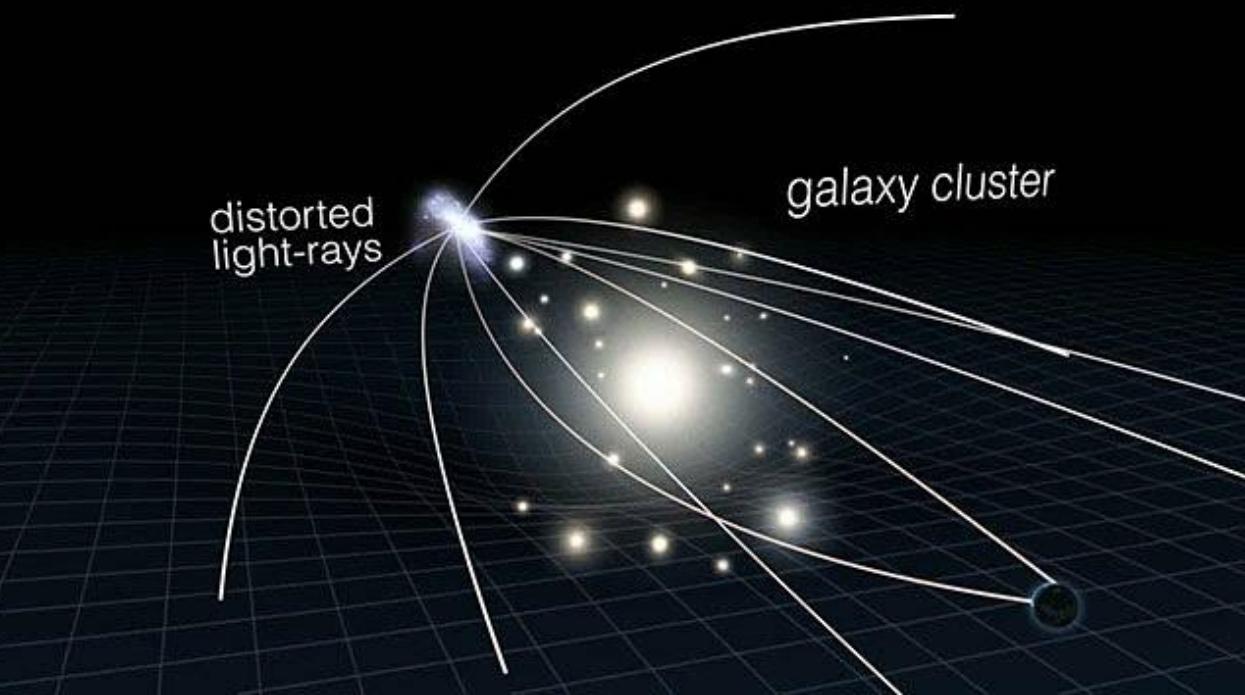
Galaxy images are  
distorted!



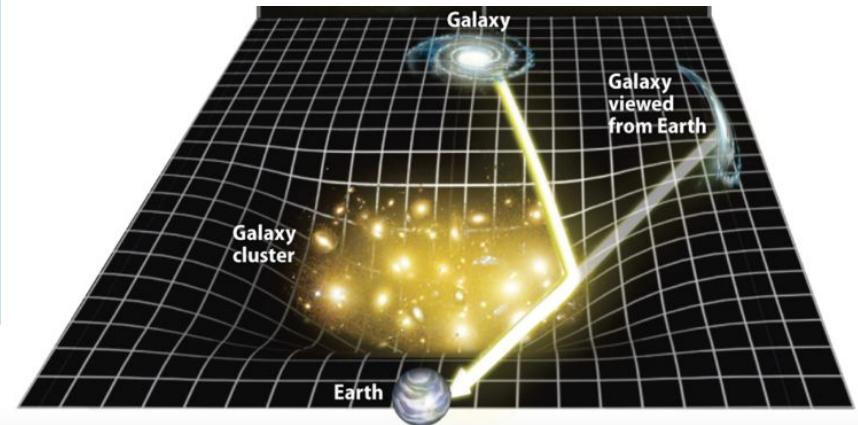


# Gravitational Lensing

A massive celestial body (galaxy cluster) causes curvature of spacetime so that the path of light is visibly bent, as if by a lens.

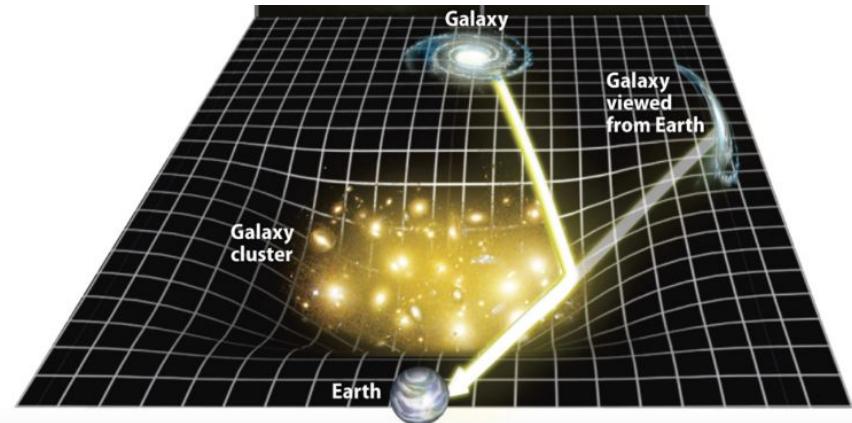


Compared to regular galaxies,  
lensed galaxies are quite rare!



Compared to regular galaxies, lensed galaxies are quite rare!

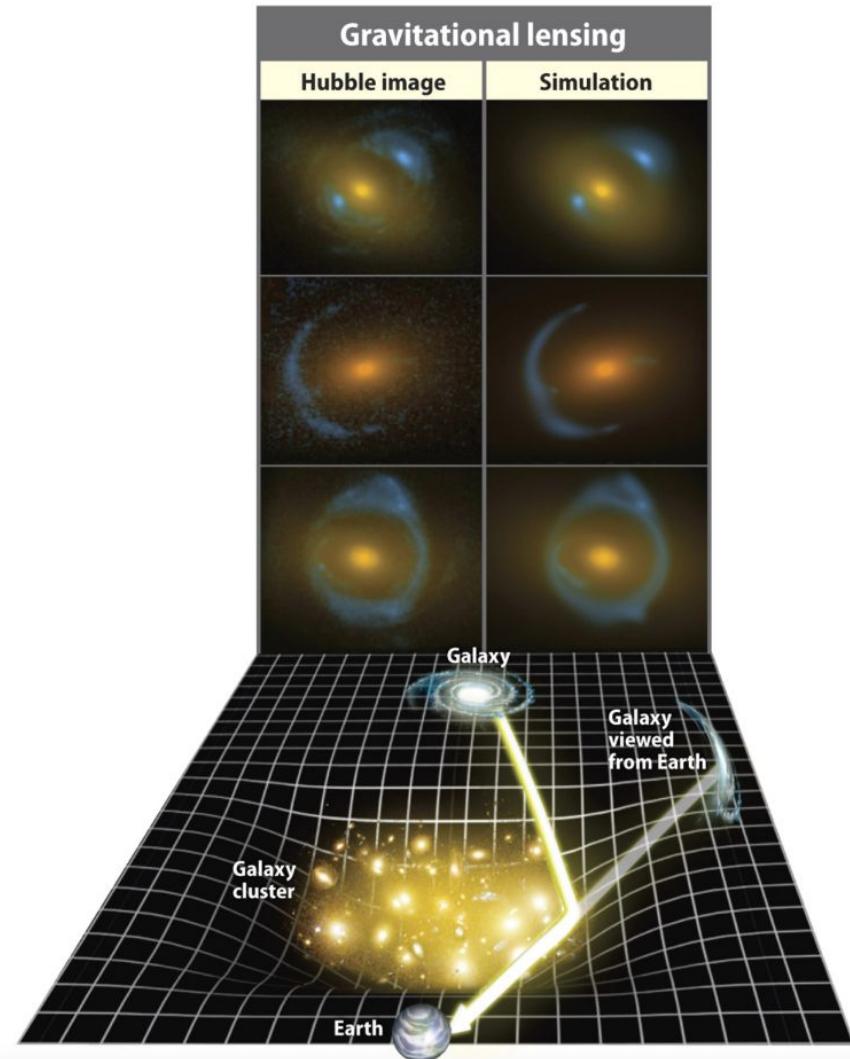
We can learn a lot about the evolution of our Universe and it's content!



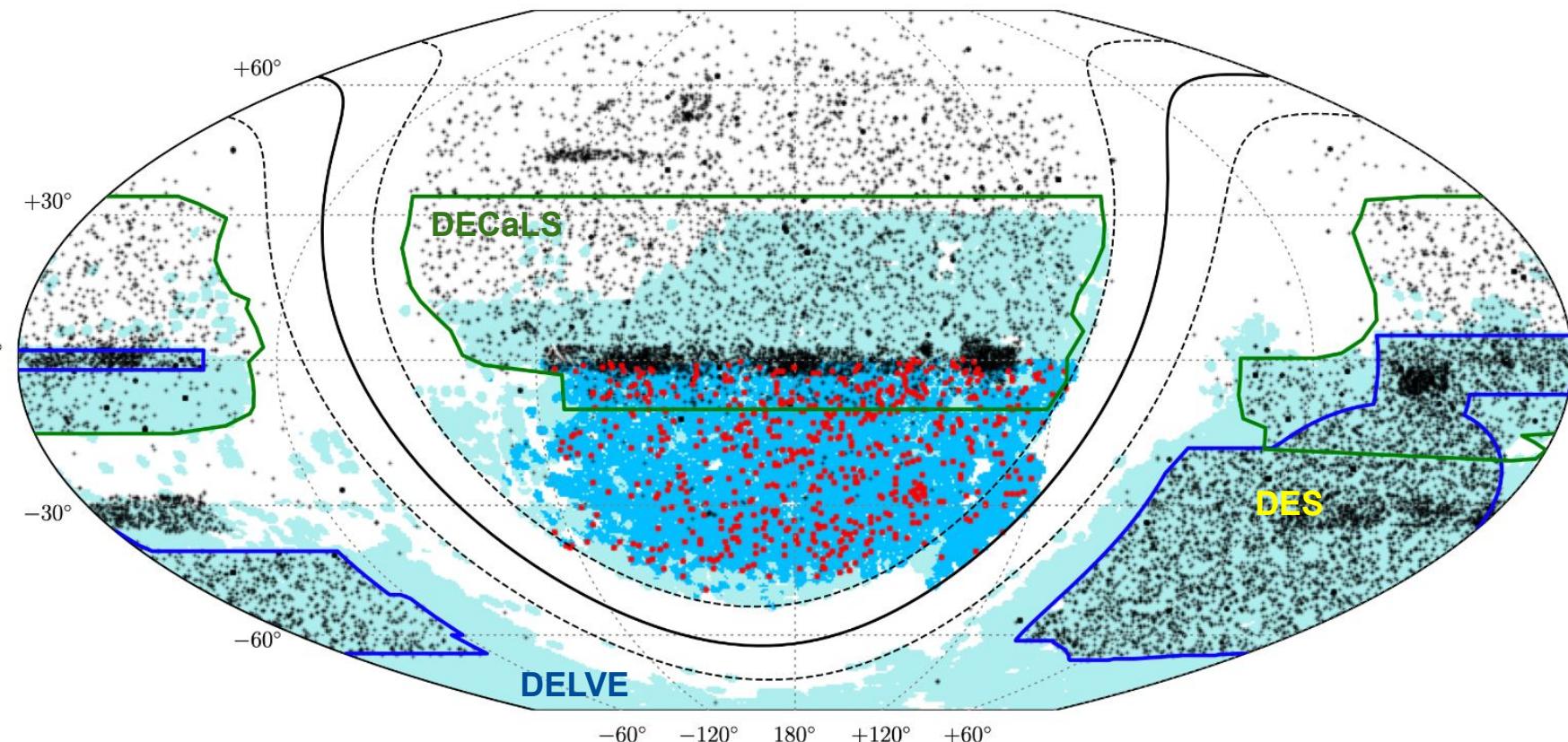
Compared to regular galaxies, lensed galaxies are quite rare!

We can learn a lot about the evolution of our Universe and it's content!

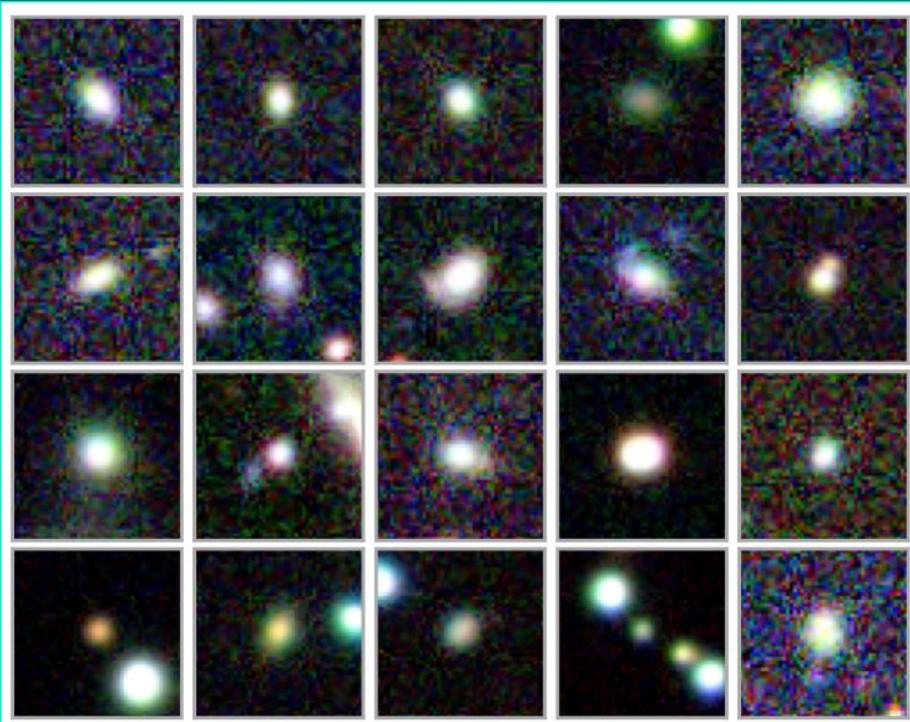
**We can also easily simulate gravitational lenses!**



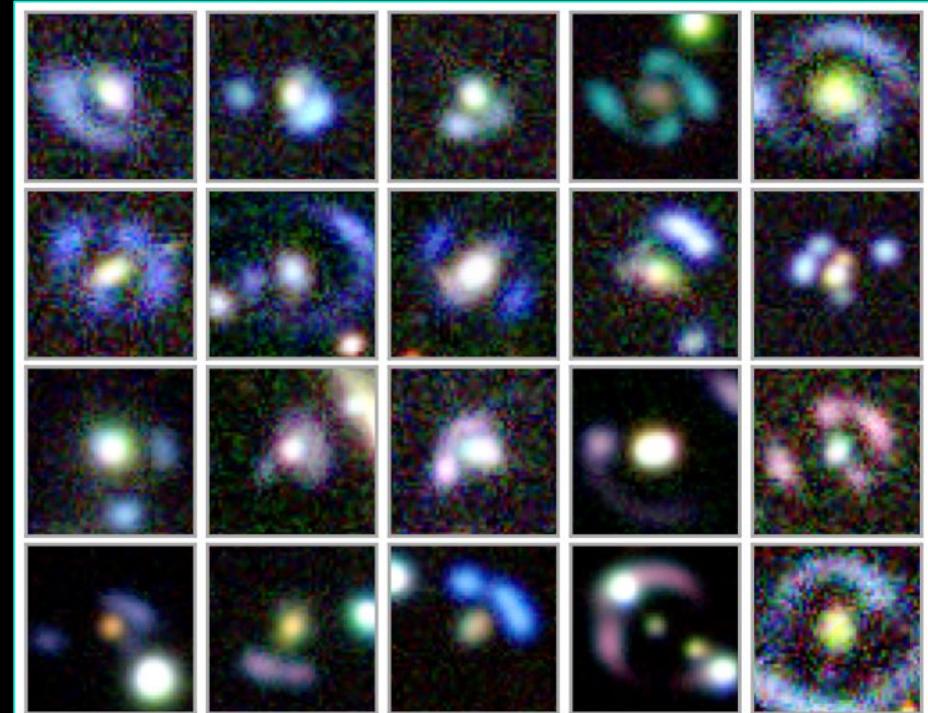
# Master Lens Database augmented by the results from DECam Local Volume Exploration Survey (DELVE)



Non-lens

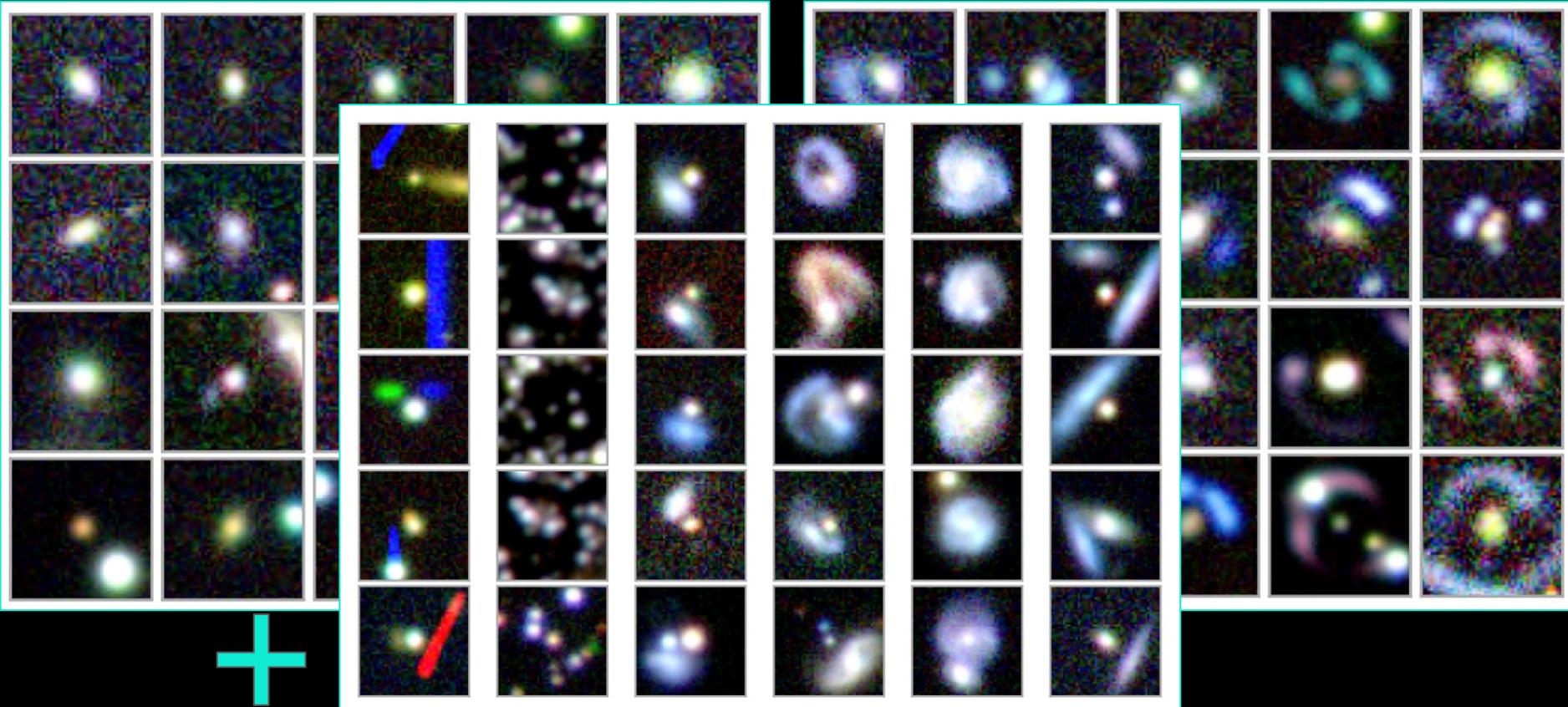


Lens

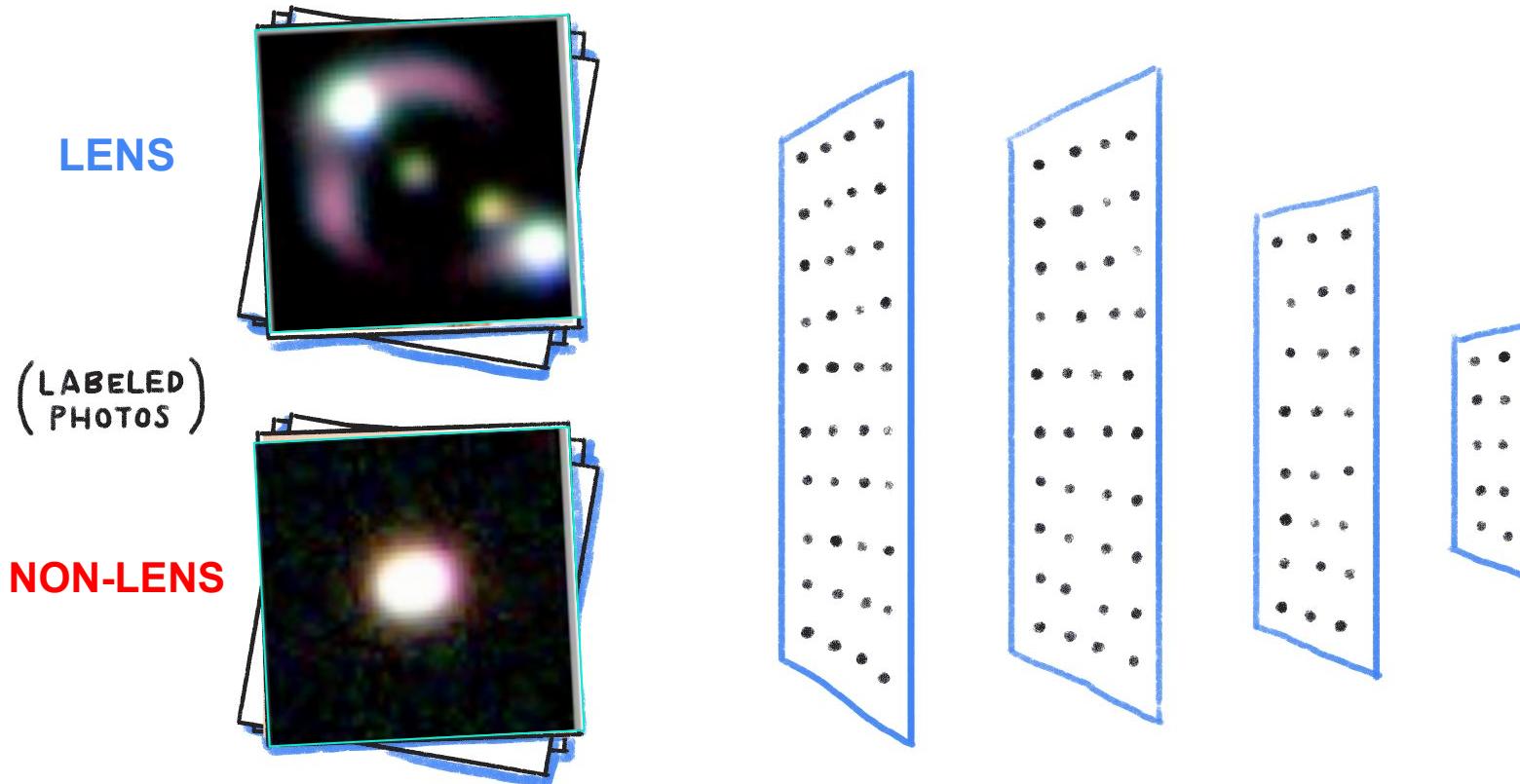


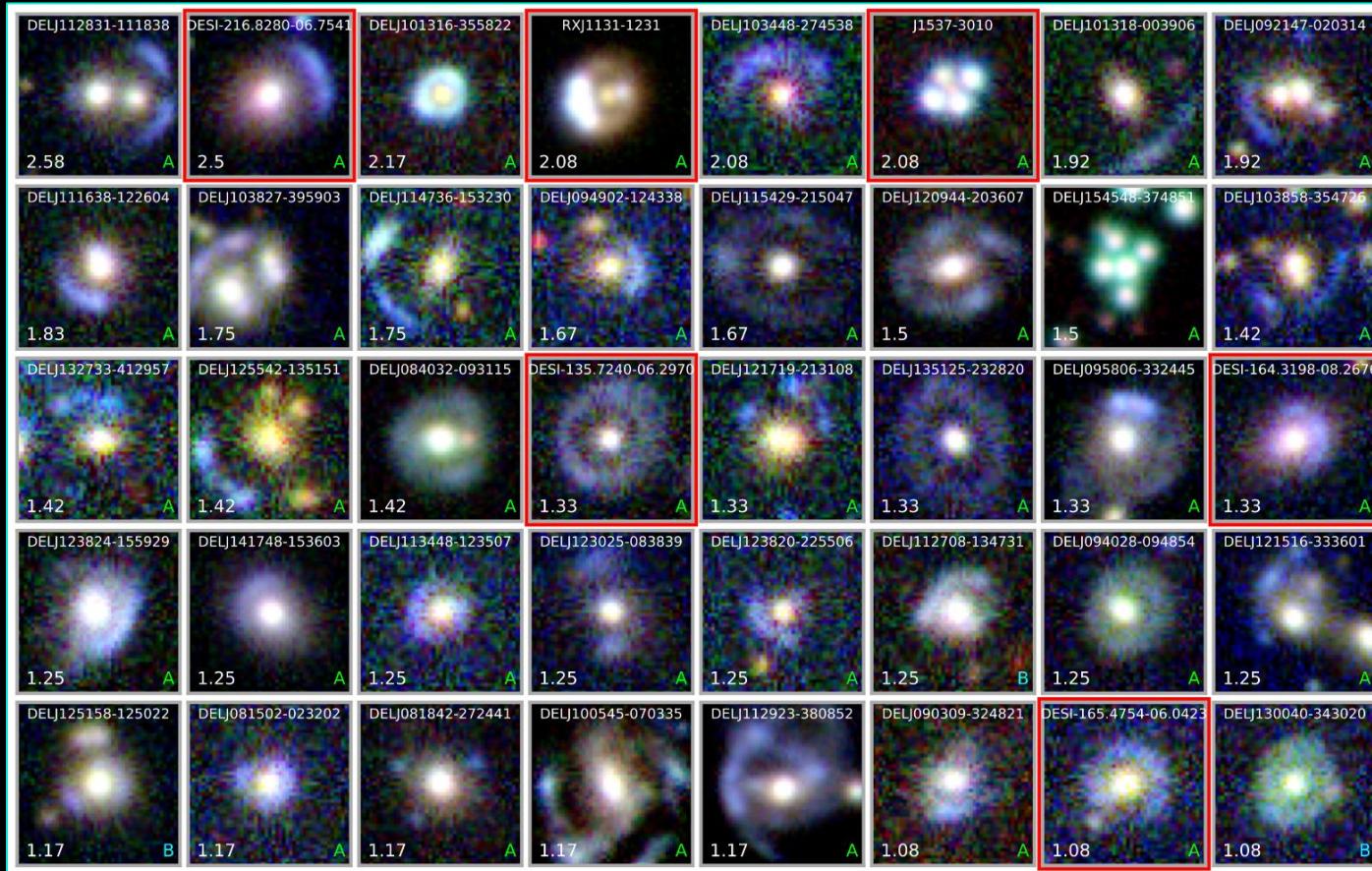
Non-lens

Lens



Model is trained with examples of lensed and non-lensed galaxies to learn to distinguish between them.





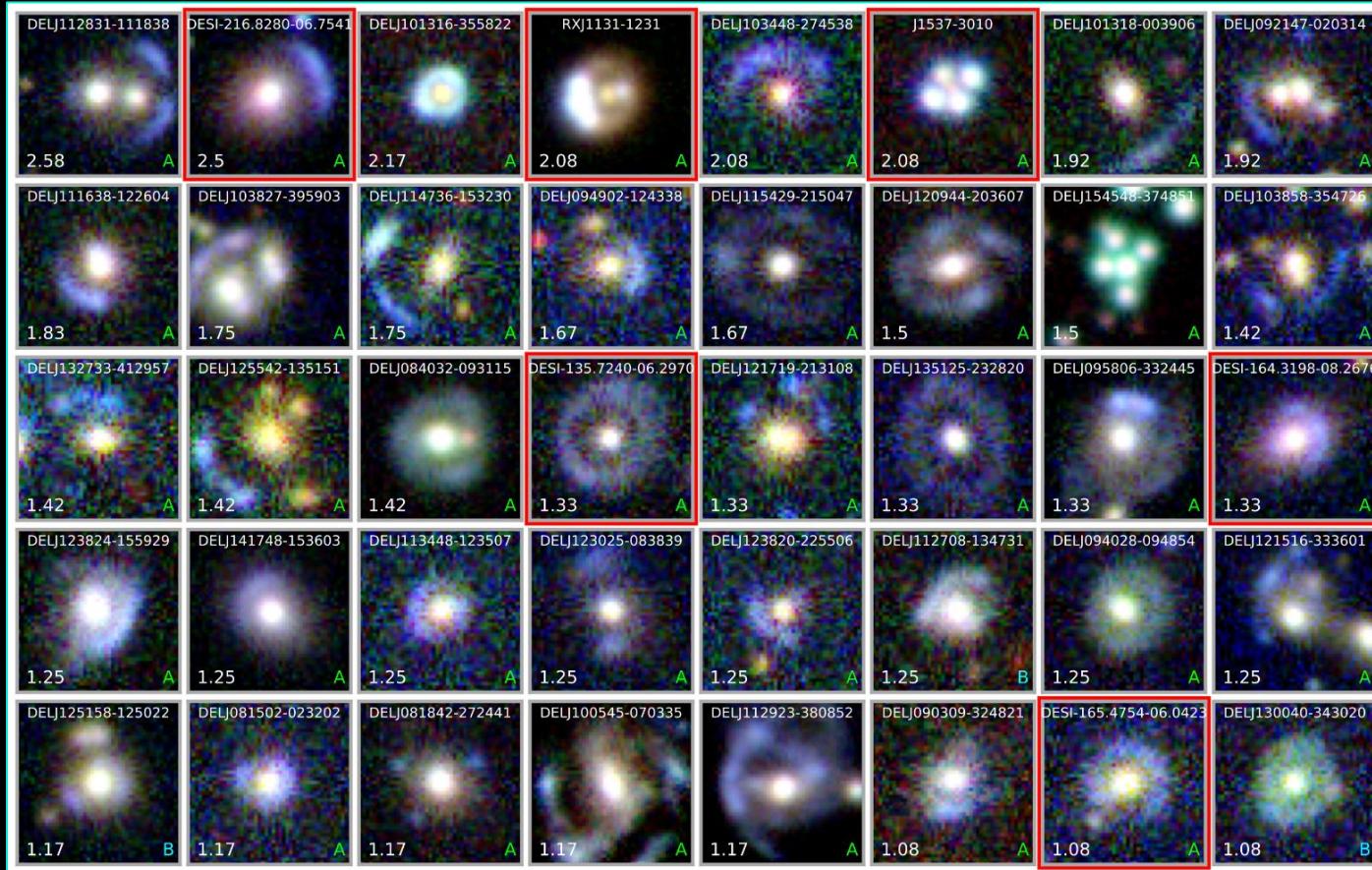
CNN finds 50,000 candidates



Two rounds of visual inspection



617 final candidates



CNN finds 50,000 candidates



Two rounds of visual inspection



617 final candidates

Hubble ~100 lenses, but LSST, Roman, Euclid will find ~10,000 !

# Why is this important?



Probe galaxy structure

Probe distribution of matter

Enable observations of the distant universe

Cosmology and how the universe evolves

# Why is this important?



Probe galaxy structure

Probe distribution of matter

Enable observations of the distant universe

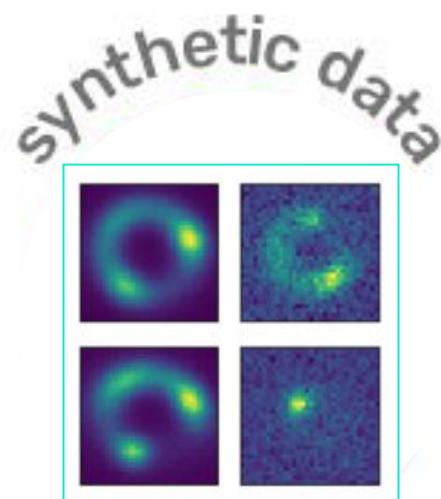
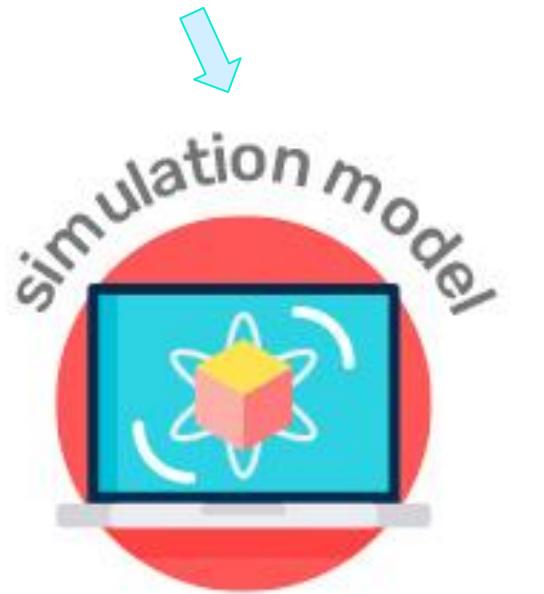
Cosmology and how the universe evolves

## How?

By inferring properties of the lens system.

# Simulation is easy....inference is hard!

Galaxy source and lens masses, shapes, positions, gravity, observational effects...



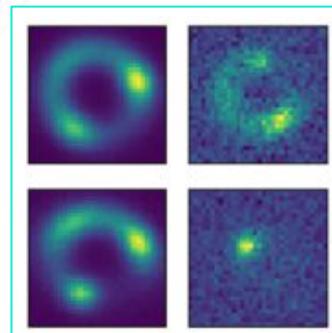
# Simulation is easy....inference is hard!

Galaxy source and lens masses, shapes, positions, gravity, observational effects...

*simulation model*



*synthetic data*



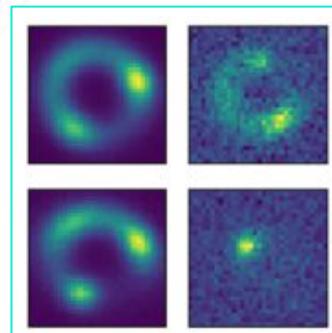
# Simulation is easy....inference is hard!

Galaxy source and lens masses, shapes, positions, gravity, observational effects...

*simulation model*



*synthetic data*



# Simulation is easy....inference is hard!

Galaxy source and lens masses, shapes, positions, gravity, observational effects...

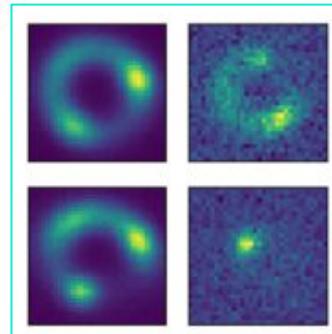
Detector properties, light propagation, galaxy properties, cosmology....

$$p(x|\theta) = \int \int \int p(x \mid \text{detect}) \ p(\text{detect} \mid \text{light}) \ p(\text{light} \mid \text{cosmo}) \ p(\text{light} \mid \theta) \ d\text{detect} \ d\text{light} \ d\text{cosmo}$$

simulation model



synthetic data



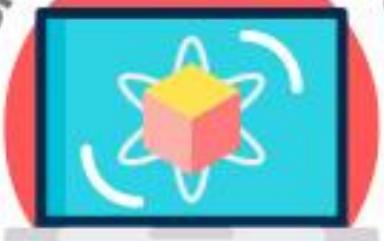
# Simulation is easy....inference is hard!

Galaxy source and lens masses, shapes, positions, gravity, observational effects...

Detector properties, light propagation, galaxy properties, cosmology....

$$p(x|\theta) = \iiint p(x \mid \text{detect}) \ p(\text{detect} \mid \text{light}) \ p(\text{light} \mid \text{cosmology})$$

simulation model



# But with AI...

It's not necessary to explicitly calculate the integral!



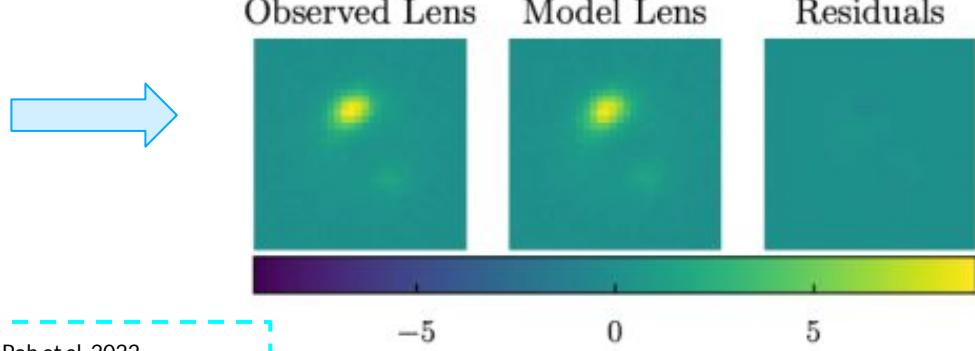
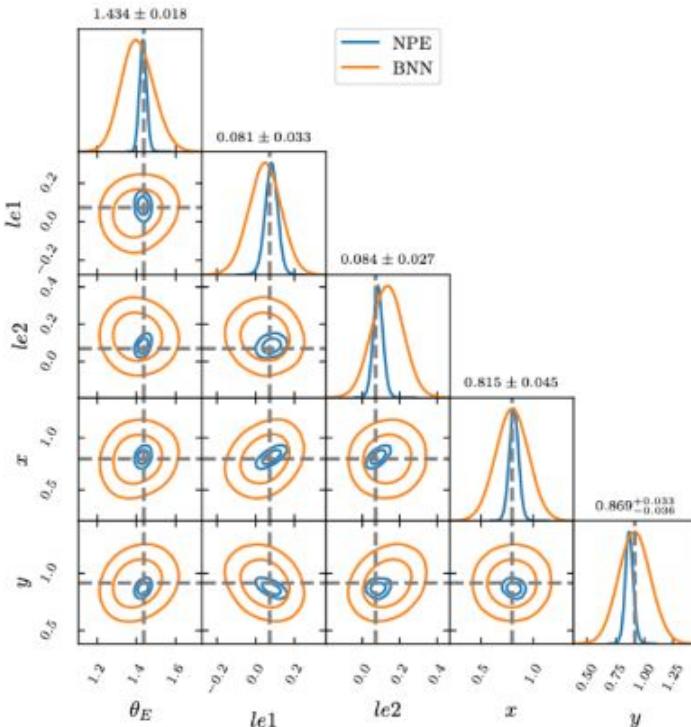
# But with AI...

It's not necessary to explicitly calculate the integral!



# But with AI...

## It's not necessary to explicitly calculate the integral!



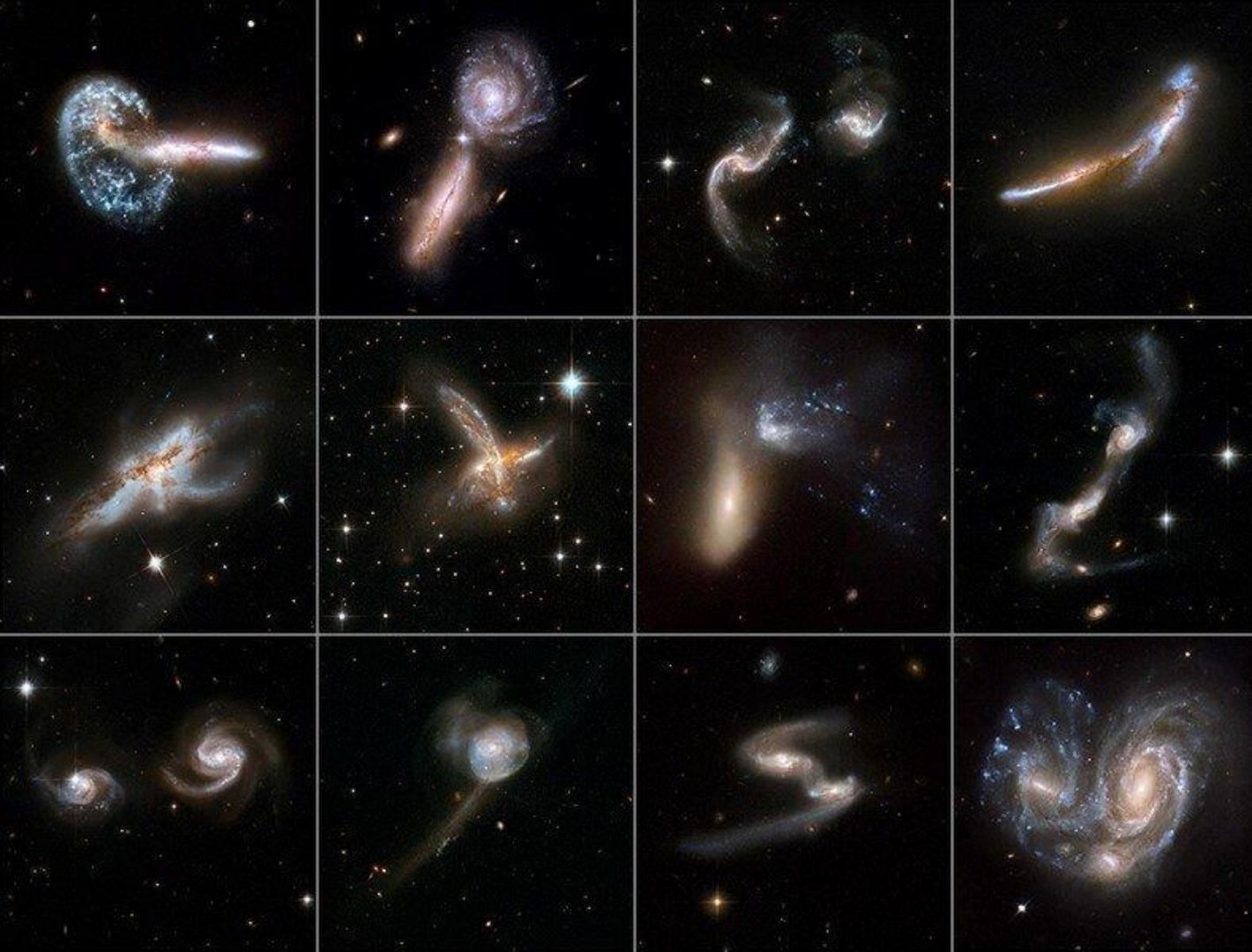
Poh et al. 2022.  
Poh et al. 2024. in prep.



How about  
these  
weird-looking  
galaxies?

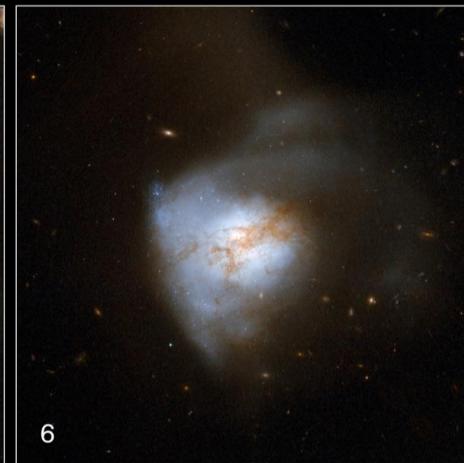
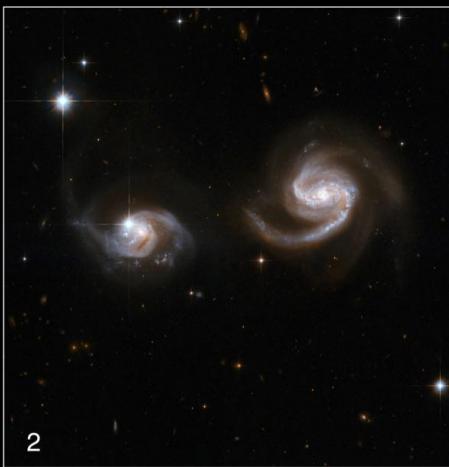
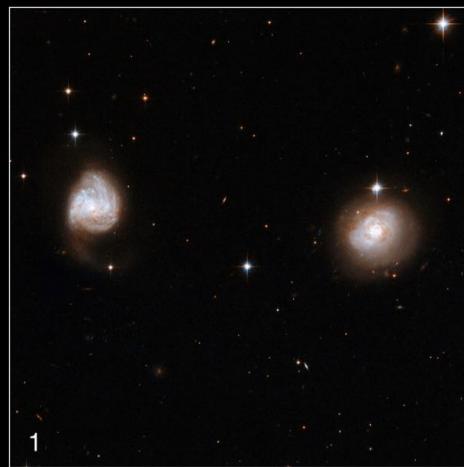
## Example 2

Merging galaxies:  
a process that lasts  
for millions of years!

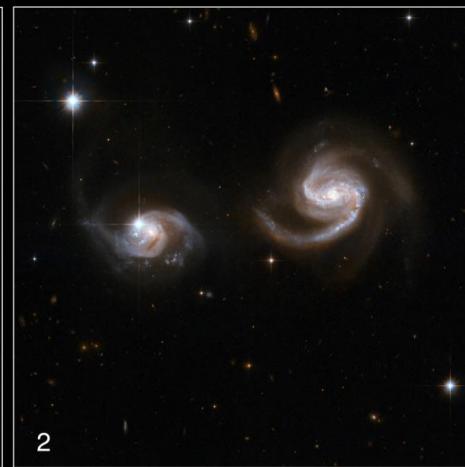
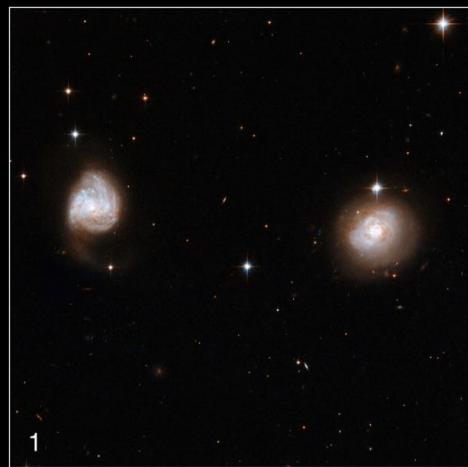


How about  
these  
weird-looking  
galaxies?

Merging galaxies:  
a process that lasts  
for millions of years!



If we want to study the process we need a lot of galaxy pairs in different merger stages!



If we want to study the process we need a lot of galaxy pairs in different merger stages!

Unfortunately we cannot easily use observations for that...

Are galaxies really merging or just visually overlapping?

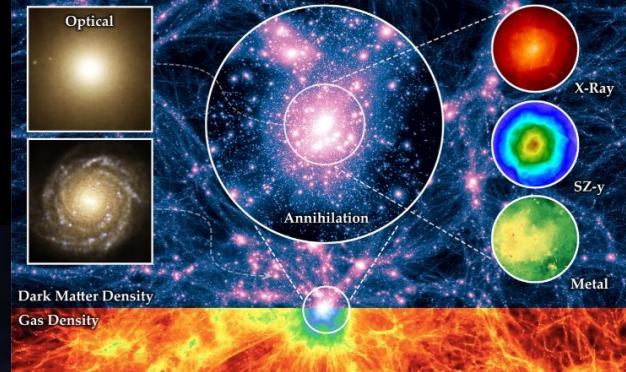


**We need simulations!**

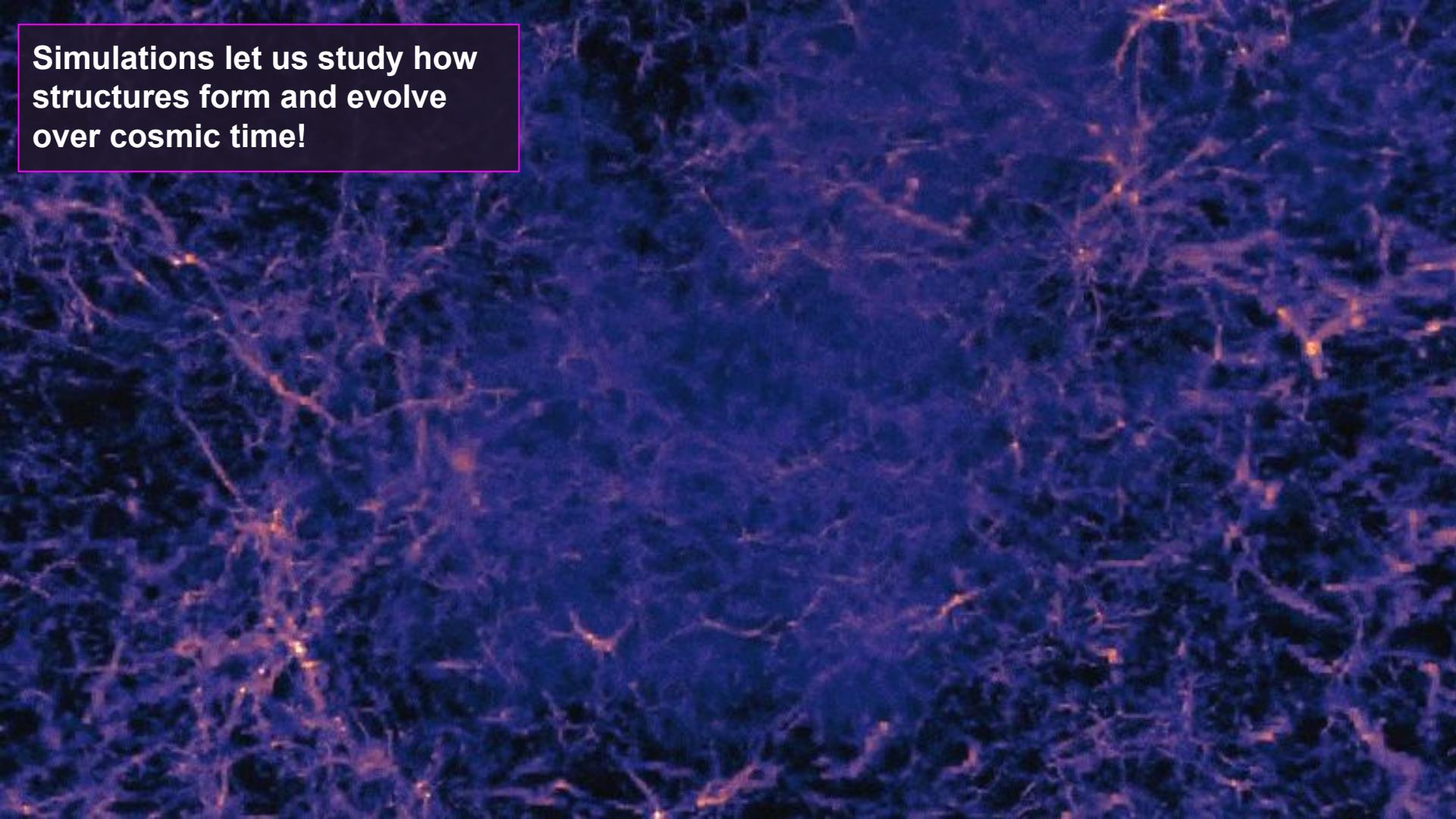
# We need even bigger simulations with many galaxies!

## The Illustris Simulation

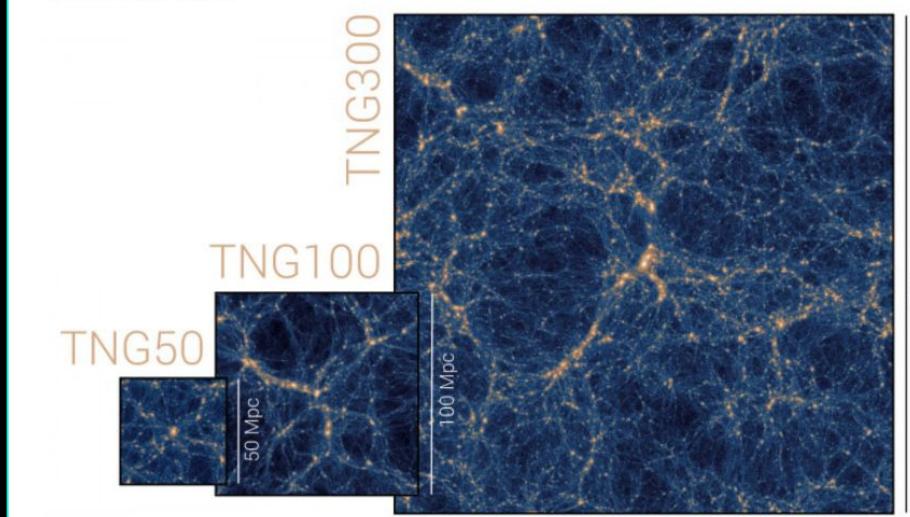
M. Vogelsberger S. Genel V. Springel P. Torrey D. Sijacki D. Xu G. Snyder S. Bird D. Nelson L. Hernquist



**Simulations let us study how  
structures form and evolve  
over cosmic time!**



1pc = 20 trillion miles

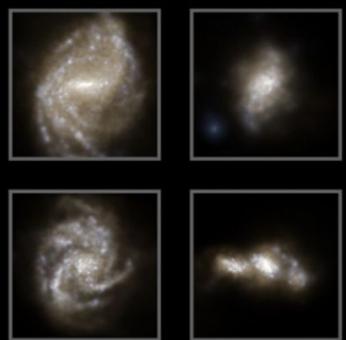




ellipticals

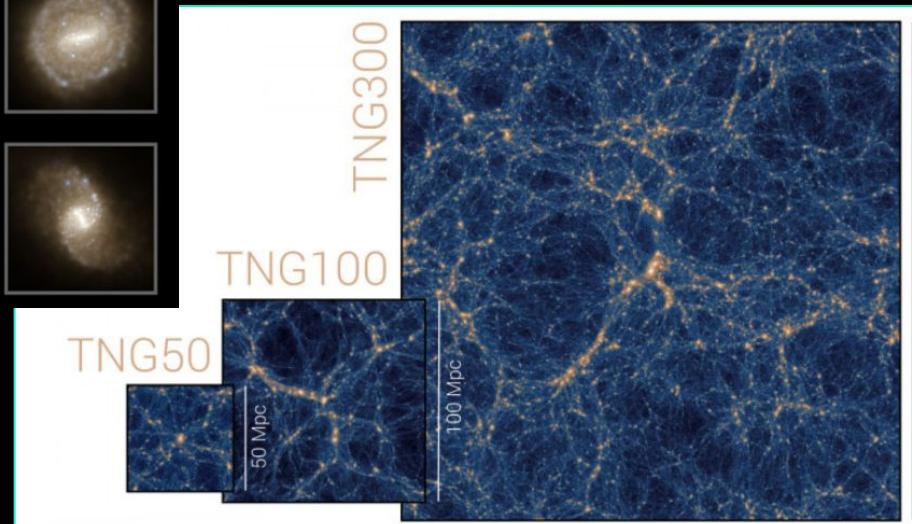


disk galaxies

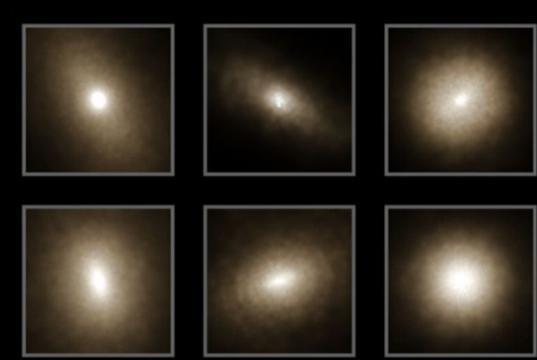


irregular

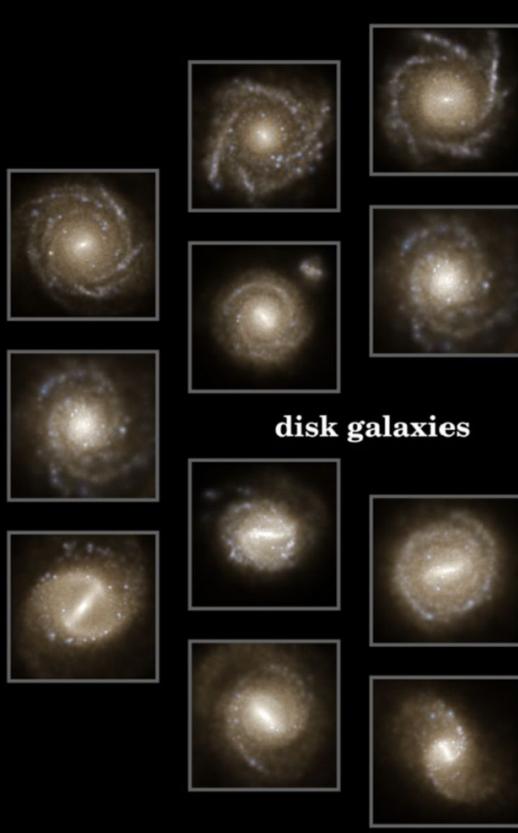
1pc = 20 trillion miles



300 Mpc



ellipticals

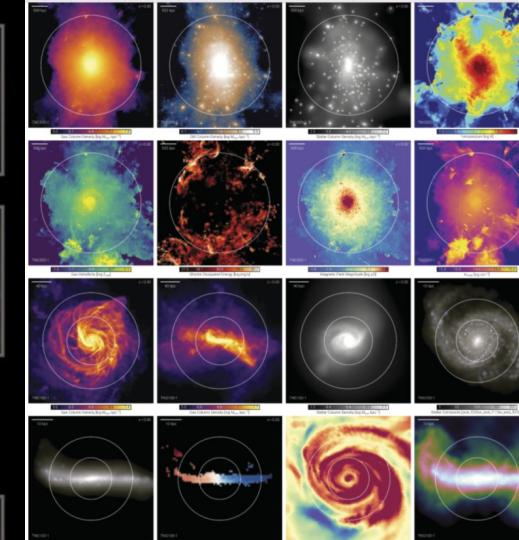


disk galaxies



irregular

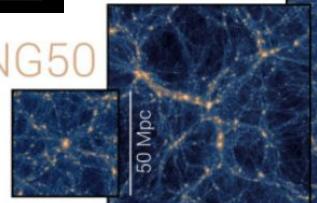
1pc = 20 trillion miles



TNG300

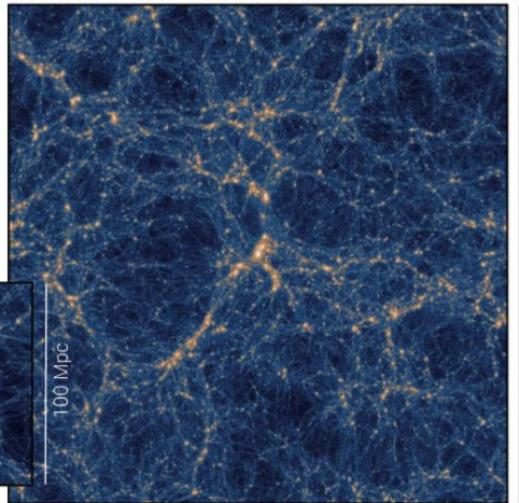
TNG100

TNG50



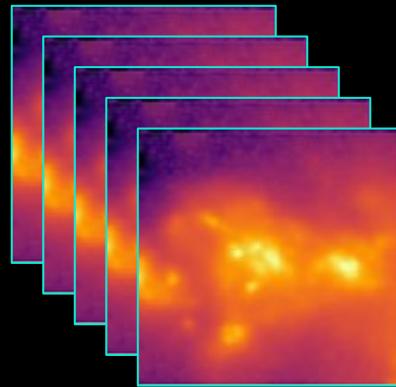
50 Mpc

100 Mpc



300 Mpc

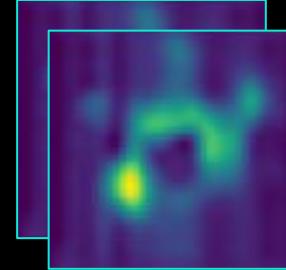
# Let's train the model:



SIMULATED  
IMAGES  
+  
LABELS

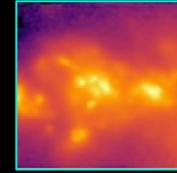
Regular Training

FIND AND REFINE  
FEATURES

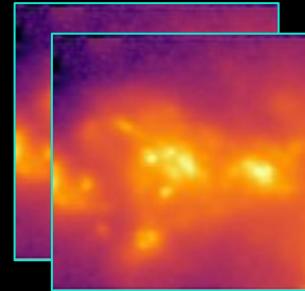


find useful  
features

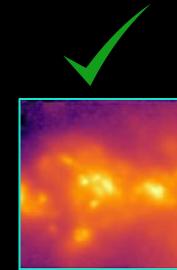
CLASSIFY



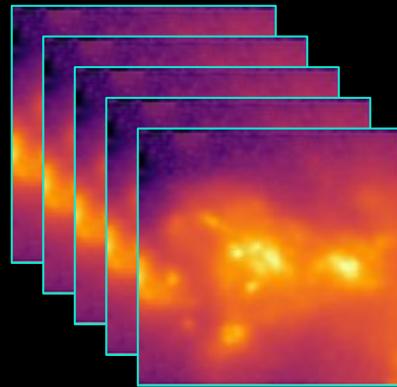
Testing the model



CLASSIFY



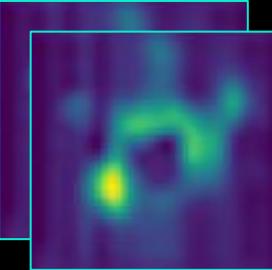
# Let's train the model:



SIMULATED  
IMAGES  
+  
LABELS

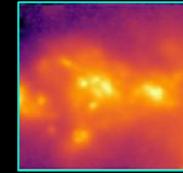
Regular Training

FIND AND REFINE  
FEATURES

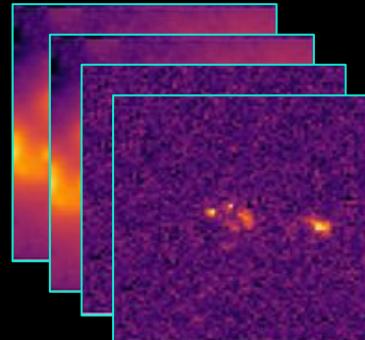


find useful  
features

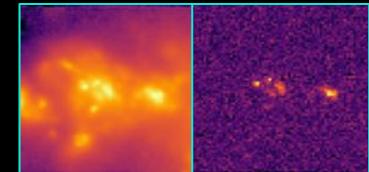
CLASSIFY



Testing the model



CLASSIFY



# Combining Datasets

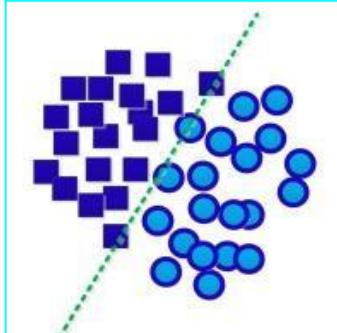
Why does this happen?

# Combining Datasets

Why does this happen?

Train the model  
on source  
dataset and find  
the decision  
boundary.

Source Domain

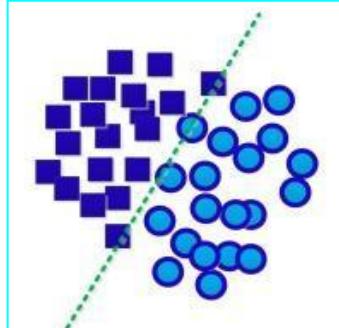


# Combining Datasets

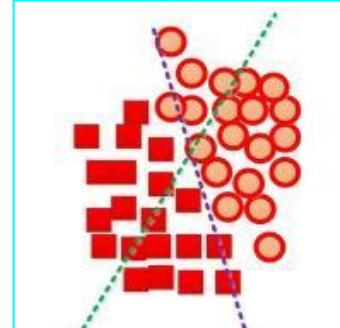
Why does this happen?

New domain is shifted,  
learned decision boundary doesn't work.

Source Domain



Target Domain

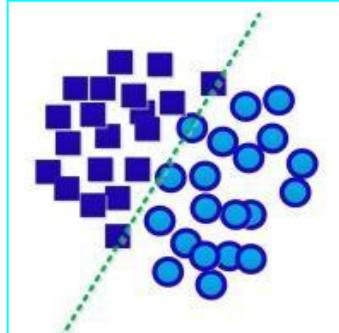


# Combining Datasets

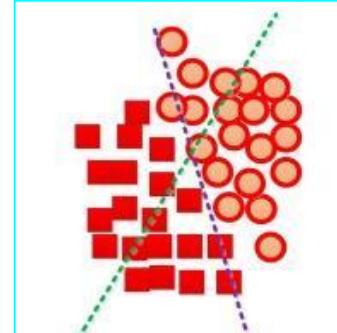
Why does this happen?

We need to align the data during training!

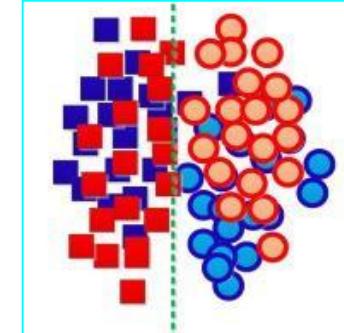
Source Domain



Target Domain



Domain Alignment



# Solving the dataset shift problem!

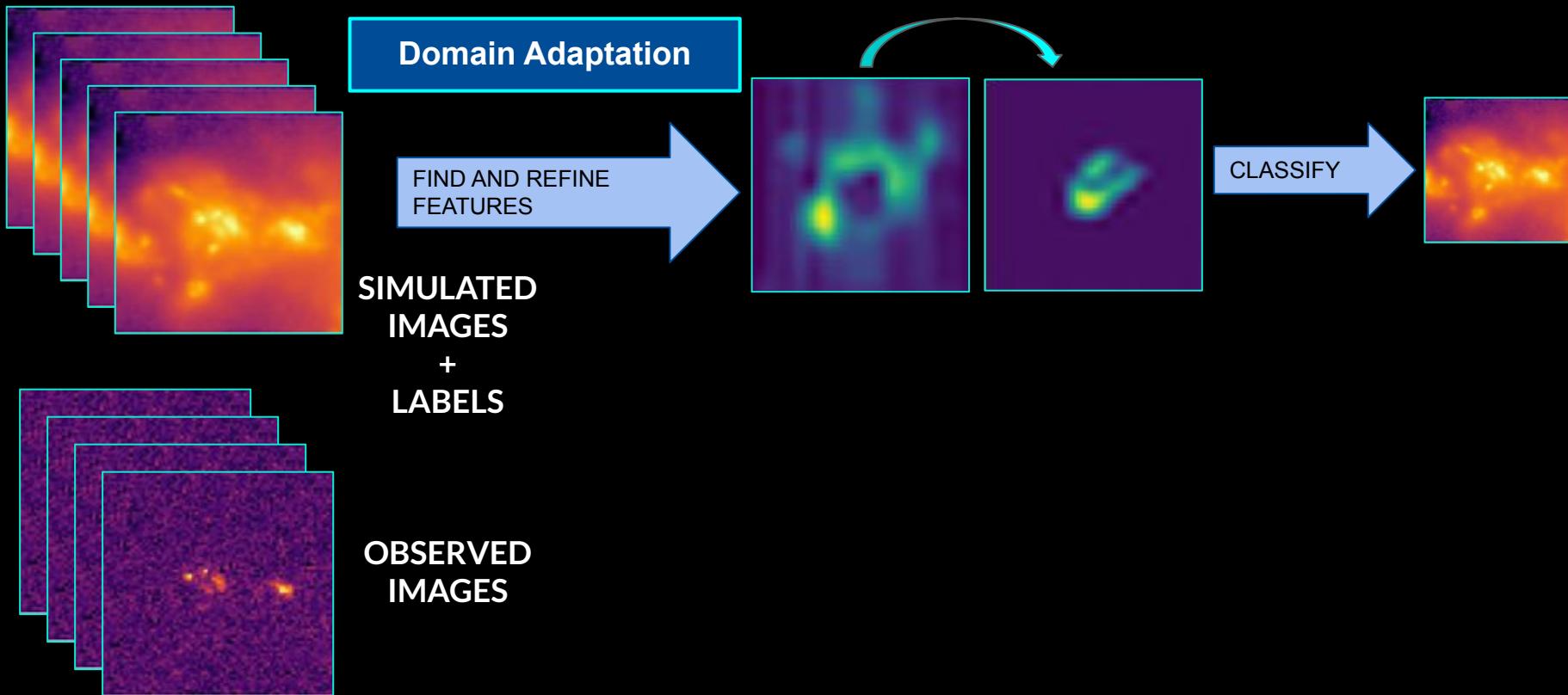
Train together and learn to use domain-invariant features.

Source - day

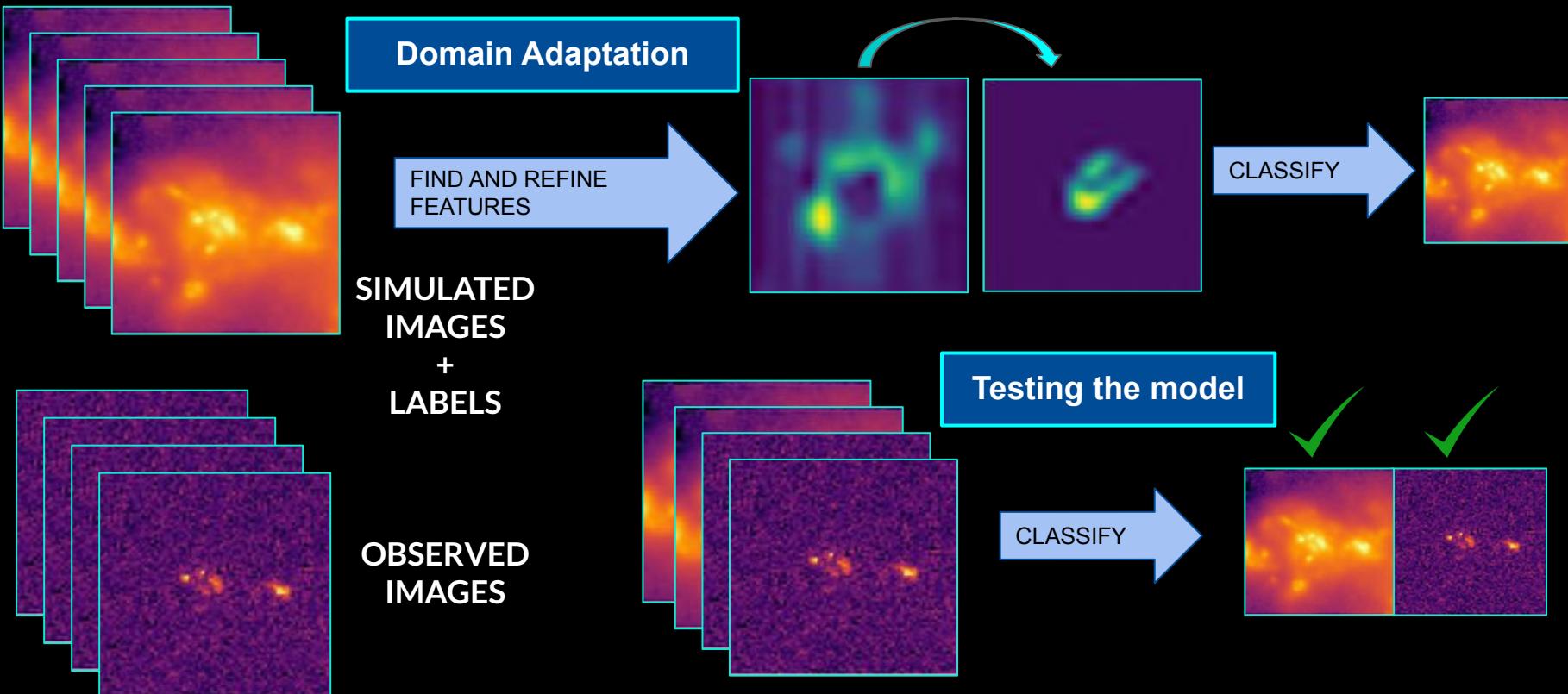
Target - night, rain



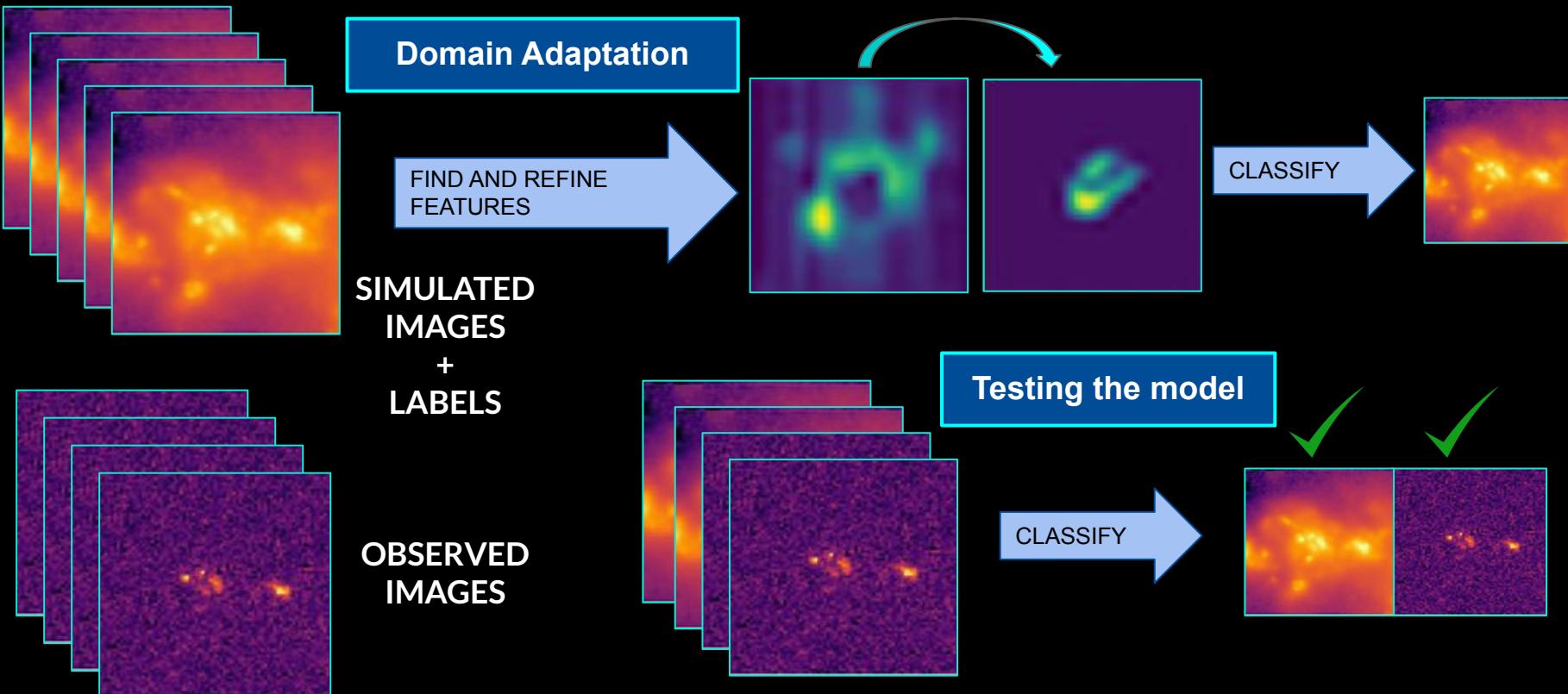
# Combining Datasets



# Combining Datasets



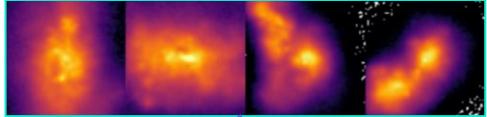
# Combining Datasets



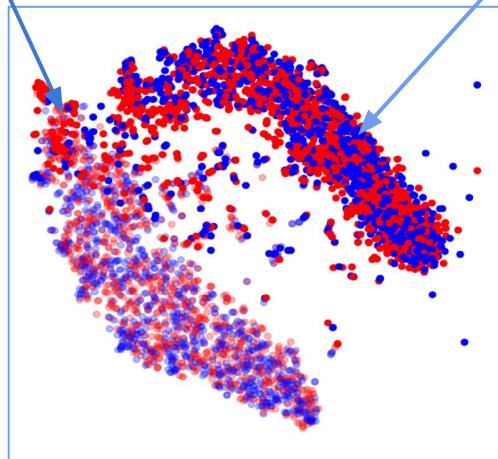
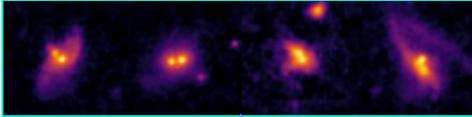
Enable merger searches in both clean and real noisy images!

# Combining Datasets

Source - Illustris



Target - SDSS observations

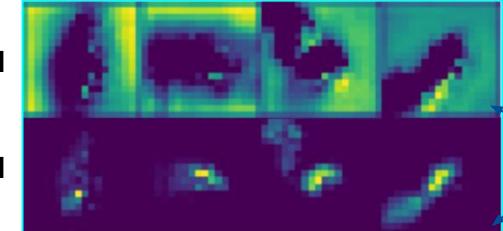
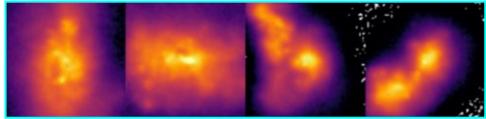


This is how the network sees the data.  
2D representation of network's latent space.

Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

# Combining Datasets

Source - Illustris



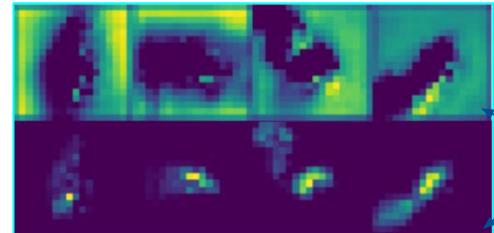
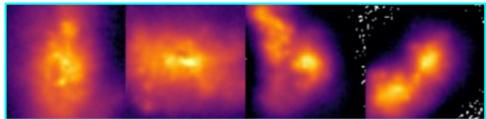
Important regions are highlighted!

Regular Training

Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

# Combining Datasets

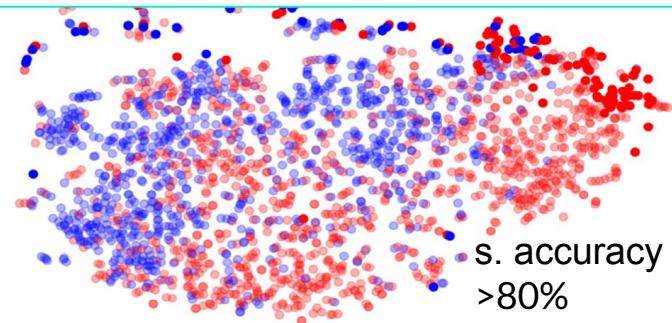
Source - Illustris



Important regions are highlighted!

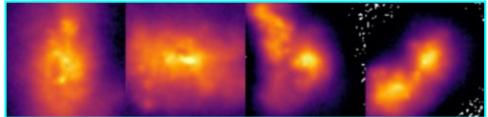
Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

Regular Training

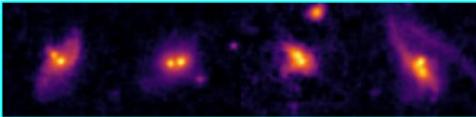


# Combining Datasets

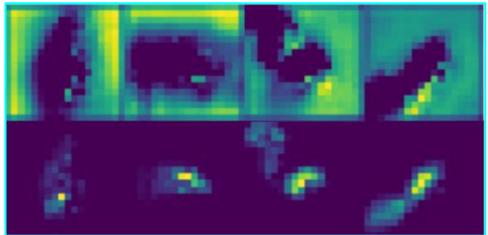
Source - Illustris



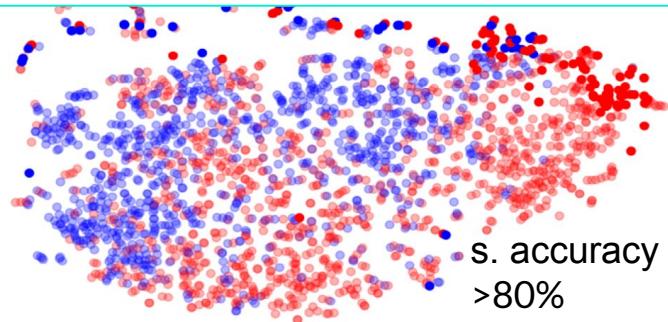
Target - SDSS observations



M  
NM



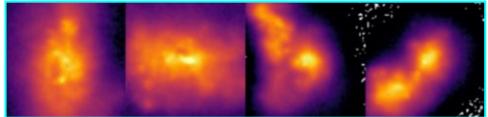
Regular Training



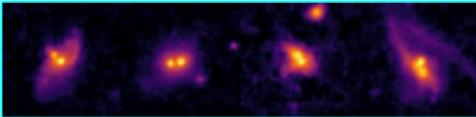
Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

# Combining Datasets

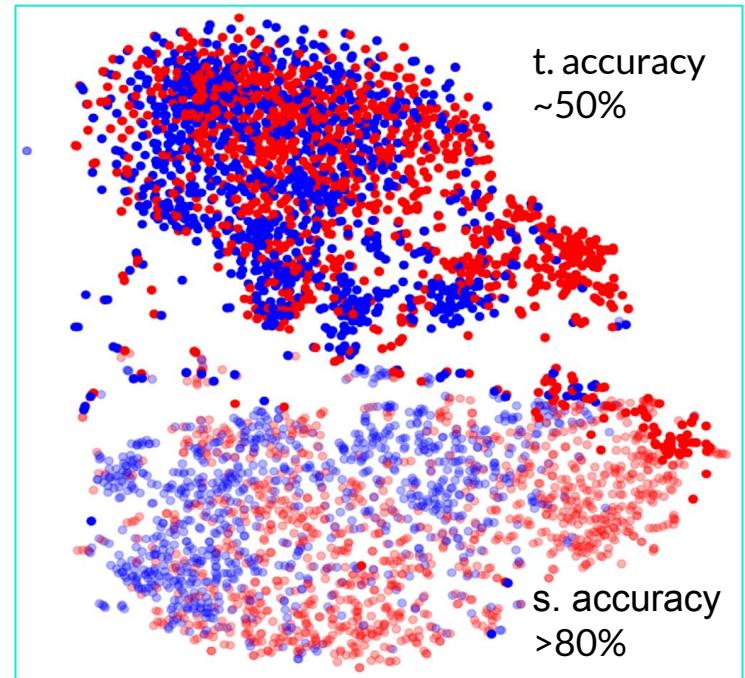
Source - Illustris



Target - SDSS observations



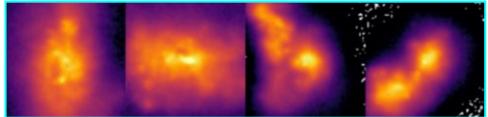
M  
NM



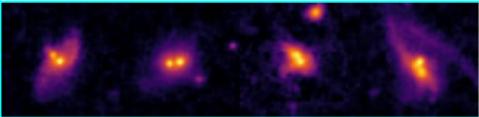
Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

# Combining Datasets

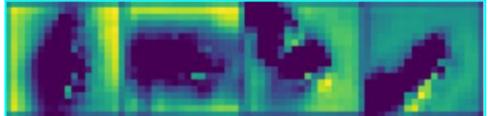
Source - Illustris



Target - SDSS observations



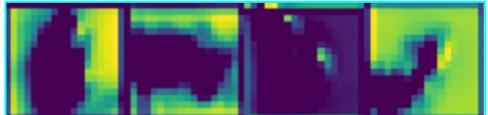
M



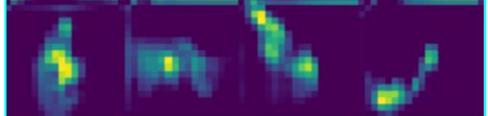
NM



M



NM

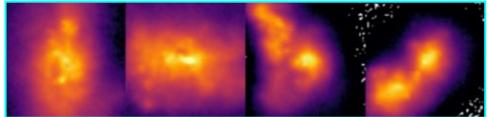


Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

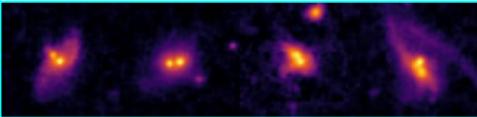
Domain Adaptation

# Combining Datasets

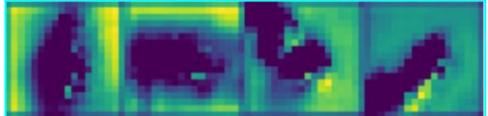
Source - Illustris



Target - SDSS observations



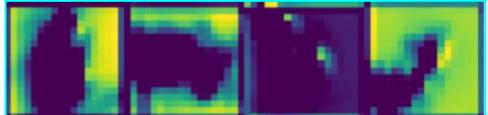
M



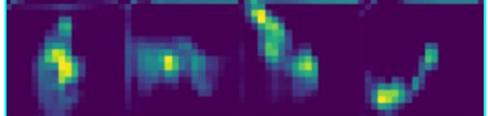
NM



M



NM

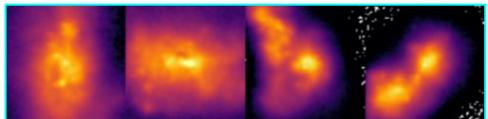


Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

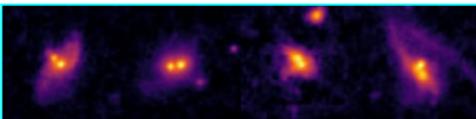
Domain Adaptation

# Combining Datasets

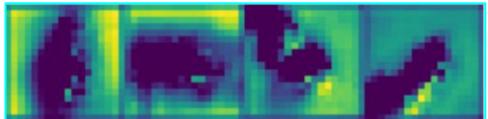
Source - Illustris



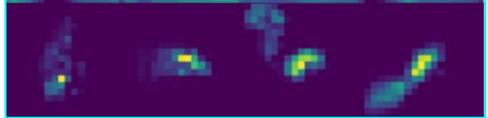
Target - SDSS observations



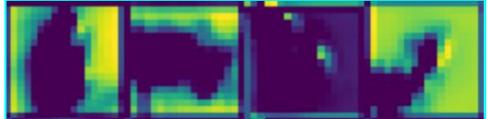
M



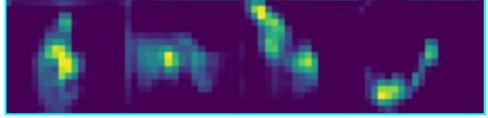
NM



M



NM



Čiprijanović et al. 2020.  
Čiprijanović et al. 2021.

Up to 30% increase!

t. accuracy  
~80%

s. accuracy  
~90%

# Talk Outline

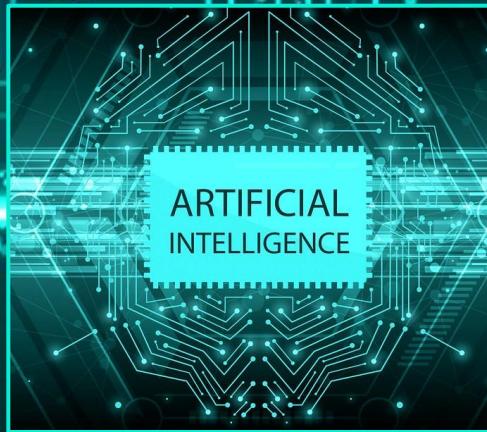
Introduction

How is AI used in Astrophysics?

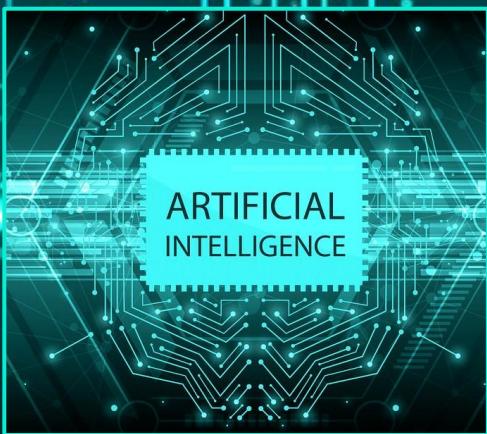
Case Studies

Challenges and Opportunities

Conclusion



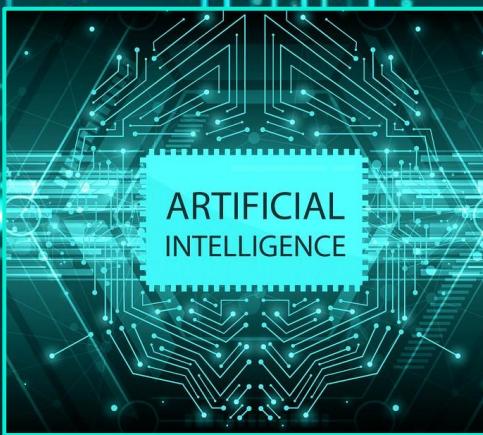
**Complex models  
based on data**



**Complex models  
based on data**

## OPPORTUNITIES

- Enabling work with huge datasets.
- Speed of analysis like never before.
- Avoid compound biases in analysis.
- Potential for new discoveries.
- Models include details, no need for approximations.



**Complex models  
based on data**

## OPPORTUNITIES

- Enabling work with huge datasets.
- Speed of analysis like never before.
- Avoid compound biases in analysis.
- Potential for new discoveries.
- Models include details, no need for approximations.

## CHALLENGES

- Model is as good as the data.
- Watch out for biased data!
- Often do not work for out-of-distribution data.
- We have to carefully think about the data and how to apply AI methods.
- It will learn even the biases we are not aware of.

# Talk Outline

Introduction

How is AI used in Astrophysics?

Case Studies

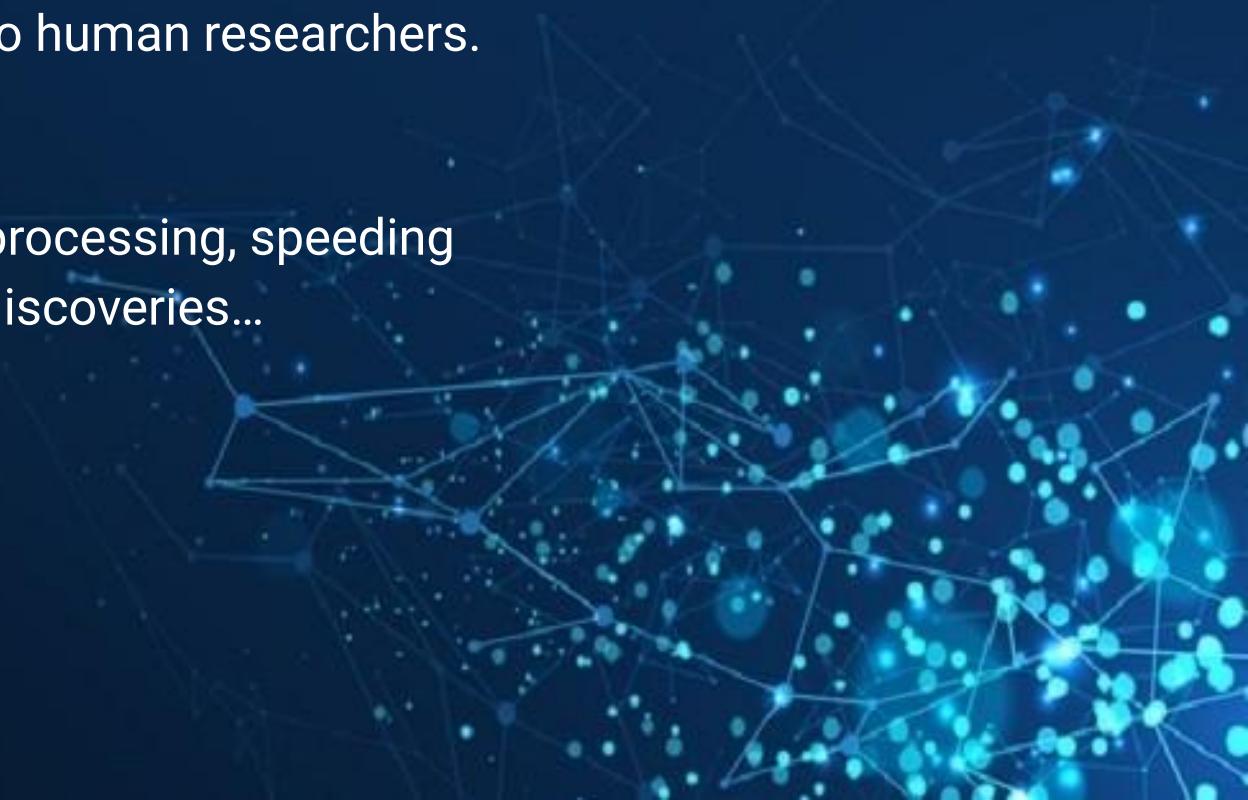
Challenges and Opportunities

Conclusion

- AI has great potential for astrophysics research, as it can be used to analyze large amounts of data and identify patterns that may not be apparent to human researchers.



- AI has great potential for astrophysics research, as it can be used to analyze large amounts of data and identify patterns that may not be apparent to human researchers.
- Data analysis, image processing, speeding up simulations, new discoveries...



- AI has great potential for astrophysics research, as it can be used to analyze **large amounts of data** and **identify patterns** that may not be apparent to human researchers.

- Data analysis, image processing, speeding up simulations, new discoveries...

- **We still have long way to go**  
**...but stay tuned!**



# THANK YOU!

Aleksandra Ćiprijanović  
(she/her/hers)

Wilson Fellow Associate Scientist  
Fermilab  
[aleksand@fnal.gov](mailto:aleksand@fnal.gov)