

DOE National Science Bowl  
April, 2024

FERMILAB-SLIDES-24-0074-CSAID

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

Aleksandra Ćiprijanović  
(she/her/hers)

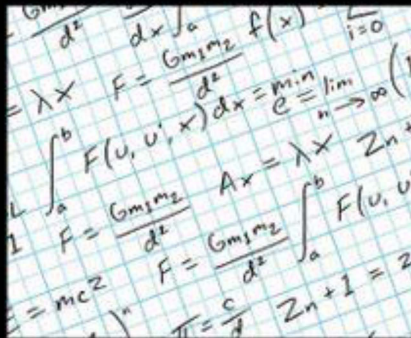
Wilson Fellow Associate Scientist  
Fermilab  
aleksand@fnal.gov

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 convertIsamI_db.getFromDB(parameter, database, view, *args)  
  
def main():  
    #print samI_db.count('main', 'blueSample')  
    #samI_db.createDB(DBfile)  
    data = np.genfromtxt(DBfile, delimiter=',', comments='#', dtype = [('objID', 'S24'), ('ra', float), ('dec',  
    ('petroMag_g', float), ('petroMag_g', float), ('petroMag_r', float), ('petroMag_i', float), ('petroMag_z', float),  
    ('petroMagErr_g', float), ('petroMagErr_g', float), ('petroMagErr_r', float), ('petroMagErr_z', float), ('exti  
    ('extinction_r', float), ('extinction_i', float), ('extinction_z', float), ('isaa_r', float), ('isaa_z', float),  
    ('lnExp_r', float), ('expMag_r', float), ('lnDev_r', float), ('devMag_r', float)], names=True) #, missing_va  
    plotData = [data[:,1:2], data[:,1:10]]  
    print plotData  
    plotScatter(plotData, ('red', 'black'), 'r_vs_z', 'Apparent magnitude vs. redshift', 'z', 'r')  
    exit()
```

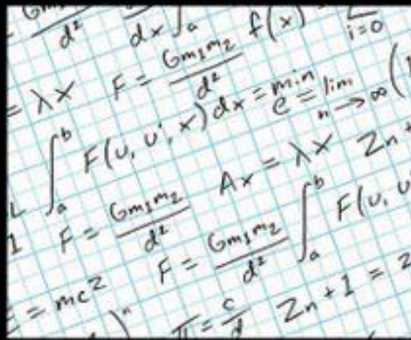
what I really do



# ASTROPHYSICIST



## what people think I do



## what friends think I do



### what Mom thinks I do

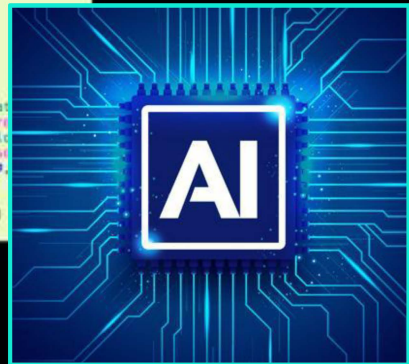


## what I think I do

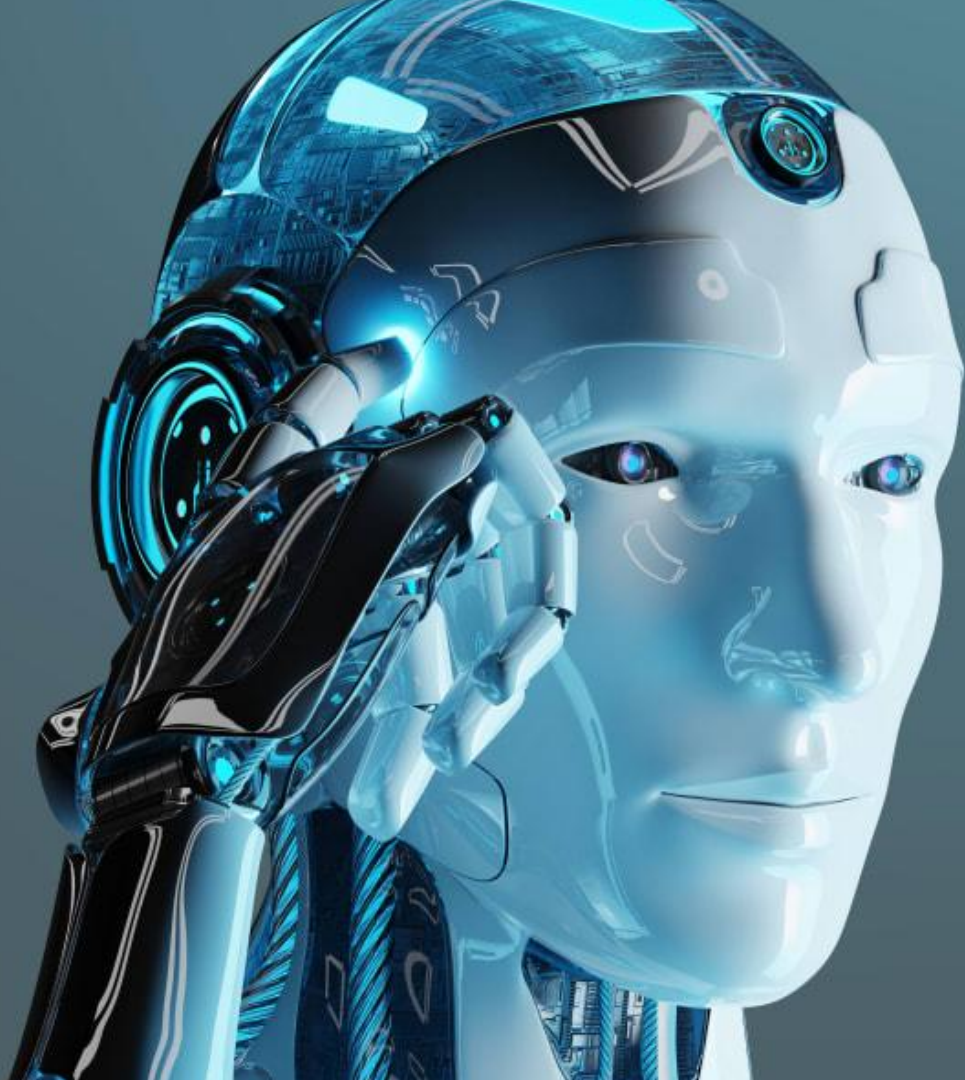
```
def getParameter(parameter, database, view, *args):
    return convert(isami_db.getFromDB(parameter, database, view, *args))

def main():
    #print isami_db.count('main', 'blueSample')
    #isami_db.createDB(DBfile)
    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 g', float), ('petroMagErr r', float), ('petroMagErr i', float), ('petroMagErr z', float), ('extinction r', float), ('extinction i', float), ('extinction z', float), ('isaa r', float), ('isaa i', float), ('isaa z', float), ('expRad r', float), ('inDev r', float), ('devRad r', float)], names=True)
    plotData = data[:,1:2], data[:,10:'petroMag r']
    print plotData
    plotScatter(plotData, ('red', 'black'), 'r vs z', 'Apparent magnitude vs. redshift', 'z', 'm r')
    exit()
```

## what i really do



# What is Artificial Intelligence (AI)?







# What is Artificial Intelligence (AI)?



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:

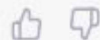




Can you help me prepare a lecture about AI in Astrophysics?



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?**

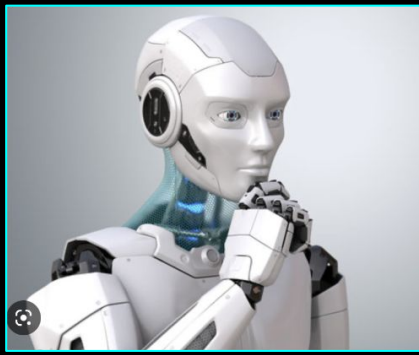
**Case Studies**

**Challenges and Opportunities**

**Conclusion**

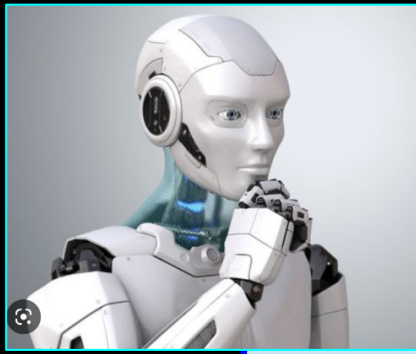


So, if not



what is AI?

So, if not



what is AI?

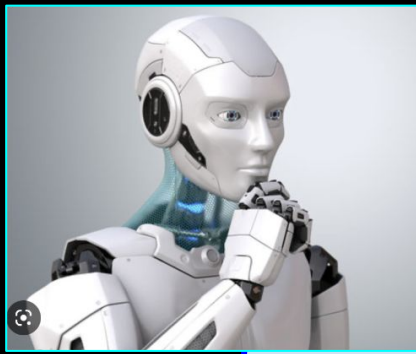
## Artificial Intelligence

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?

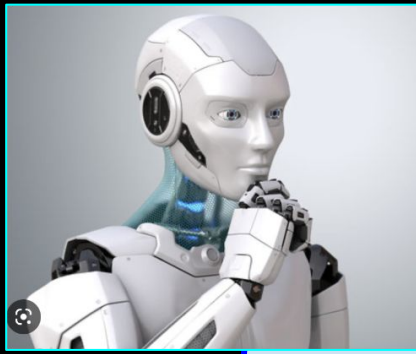
## Artificial Intelligence

### 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

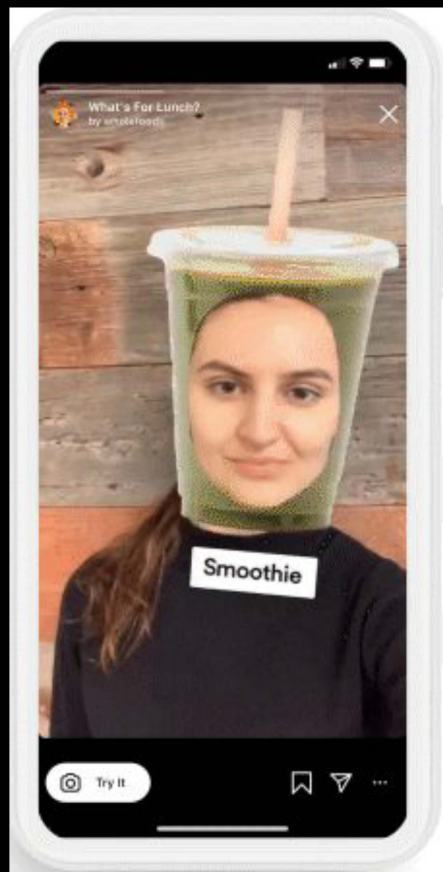


Machine Learning

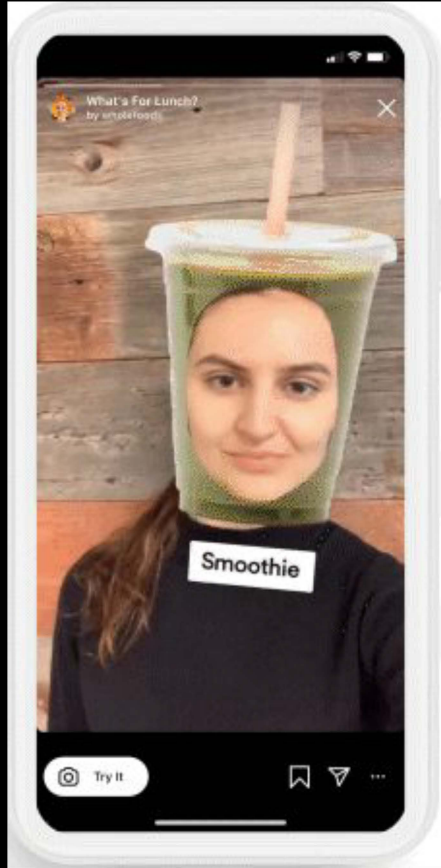
Deep Learning

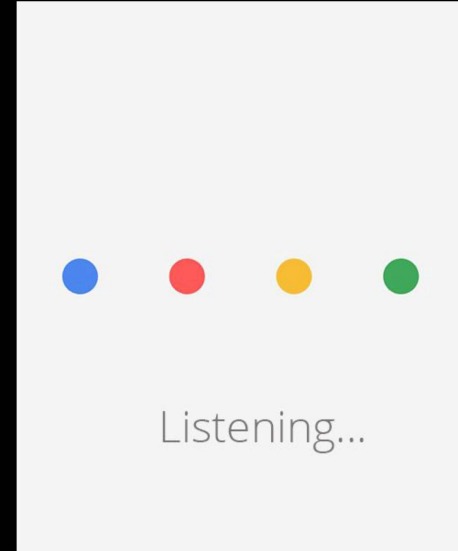
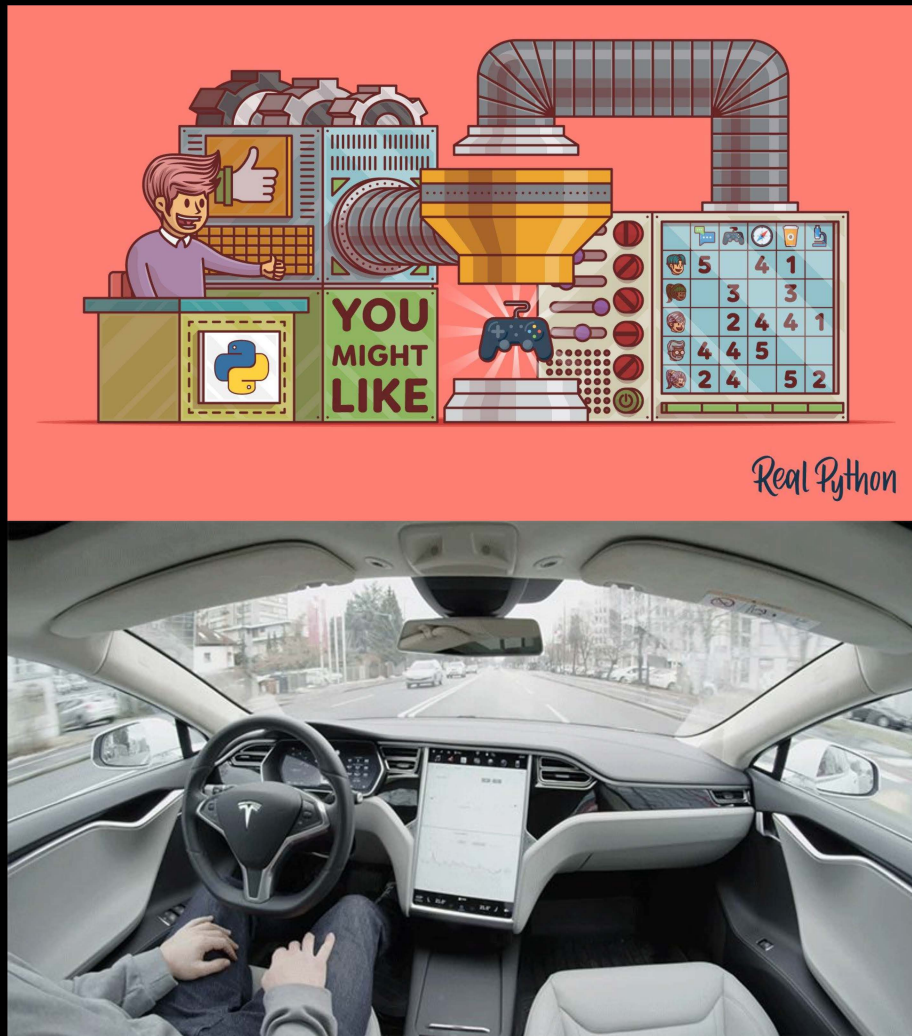
Learn **complex tasks using deep neural networks** - image and video analysis, speech recognition, language translation, autonomous vehicles, robotics...









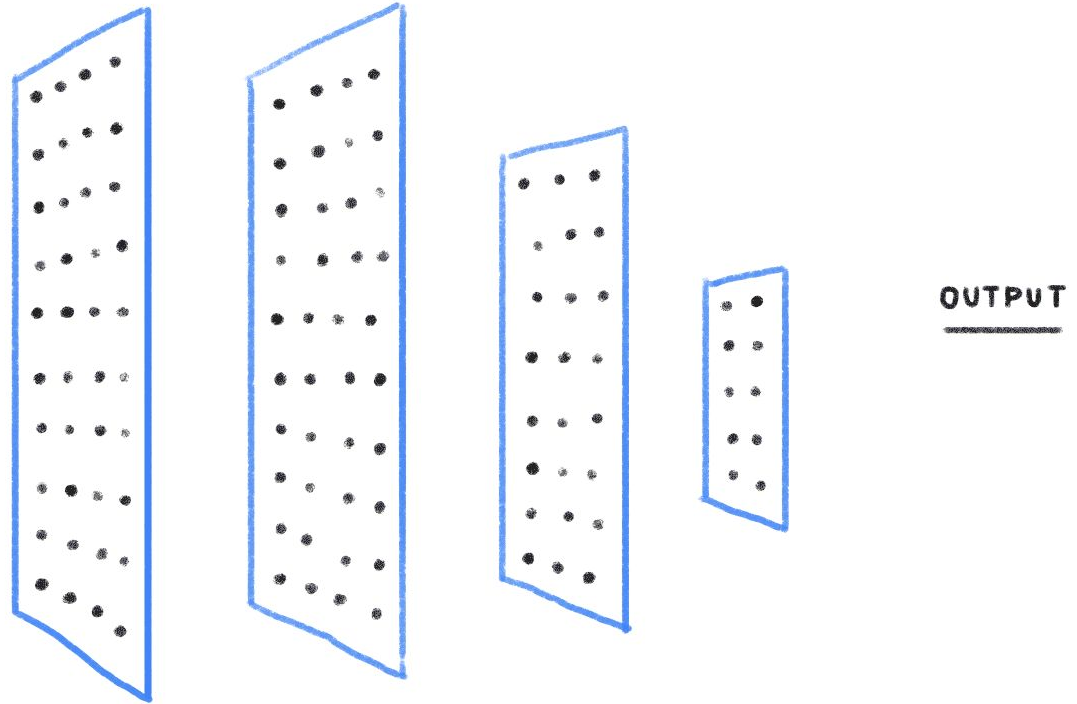
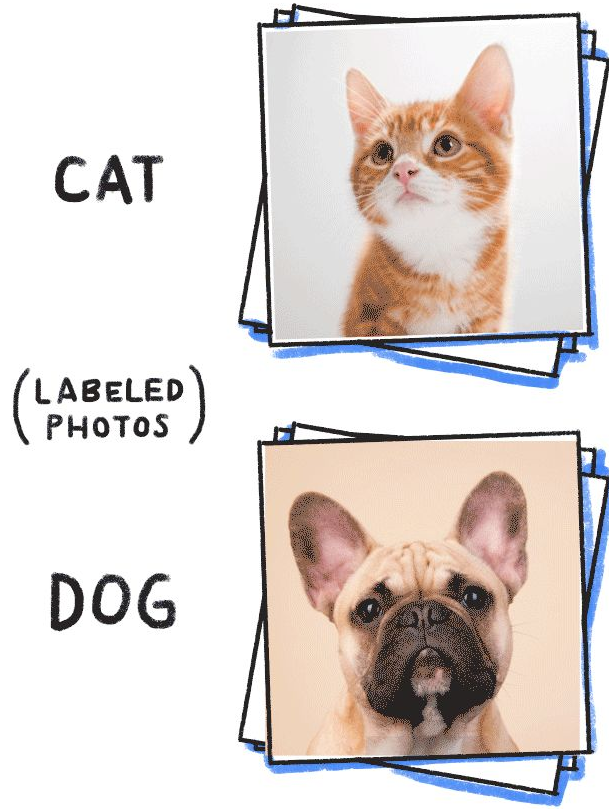




AI methods are inspired by neurons in our brains.



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





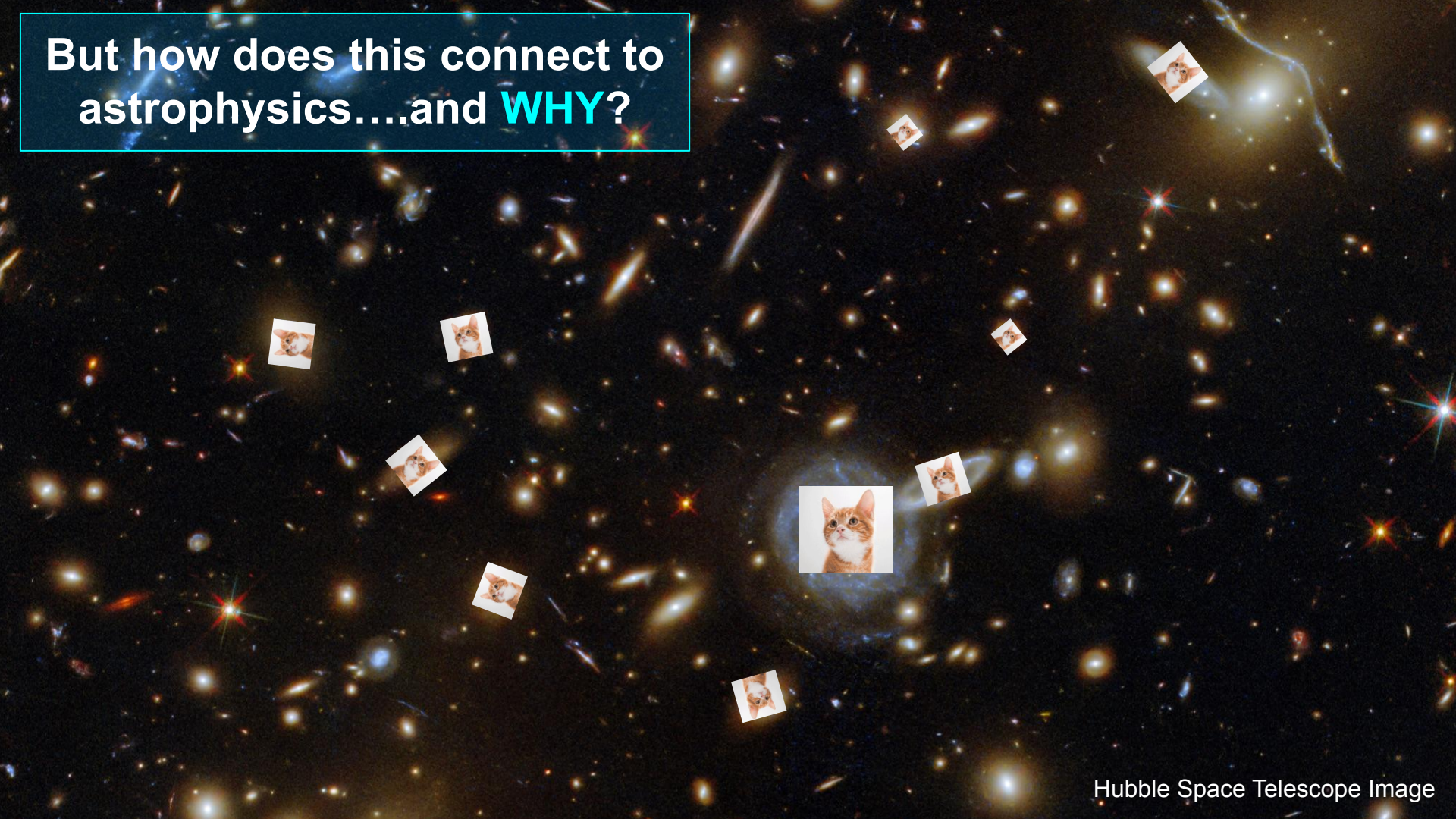
But how does this connect to  
astrophysics....and **WHY?**



Hubble Space Telescope Image



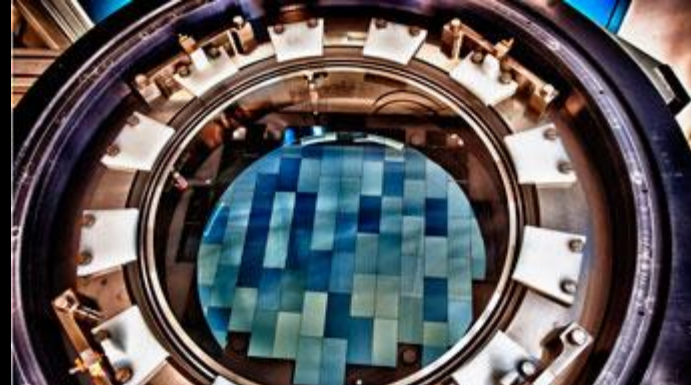
But how does this connect to  
astrophysics....and **WHY?**







iPhone camera



Our Camera!!!



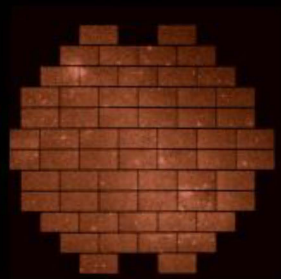
iPhone camera

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

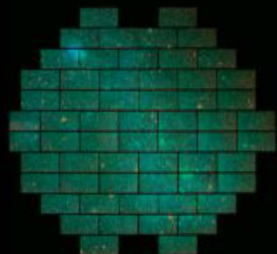
■ iPhone  
Camera



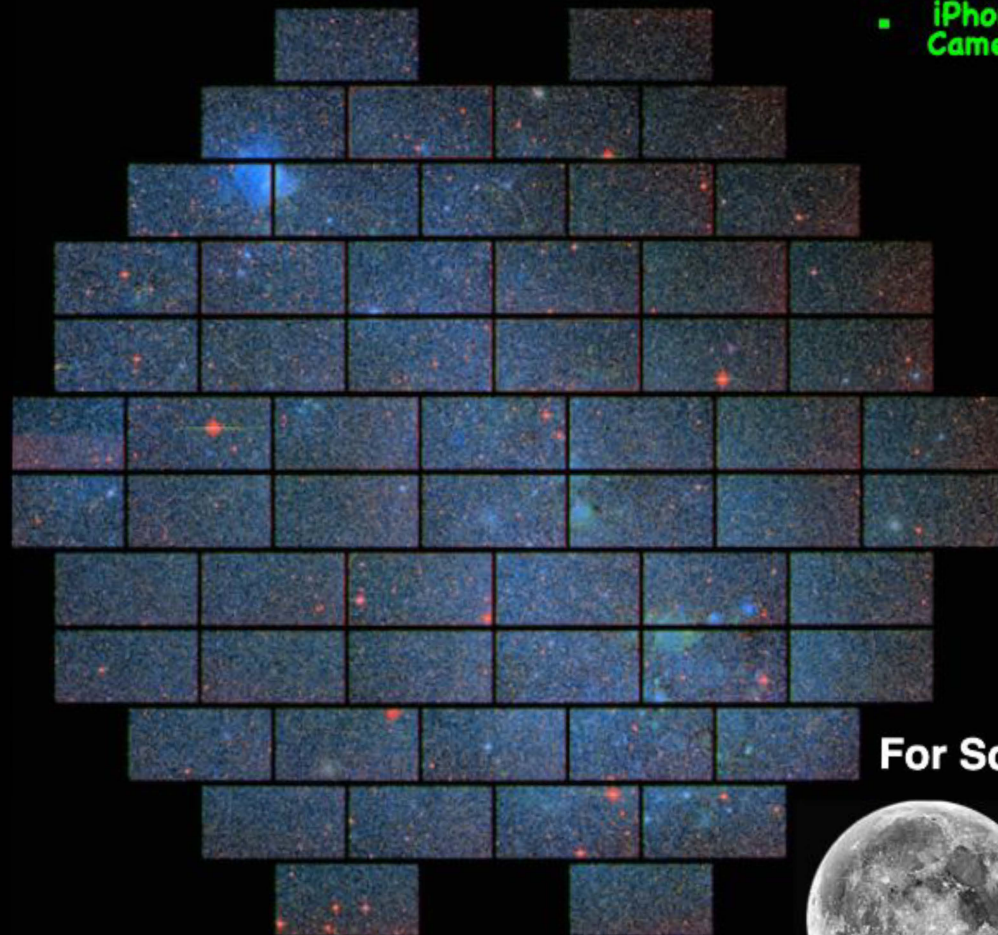
■ iPhone  
Camera



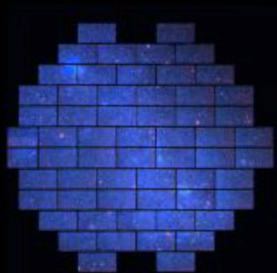
+



=



+

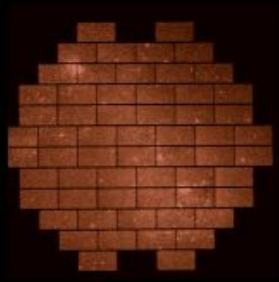


+ 2 more filters

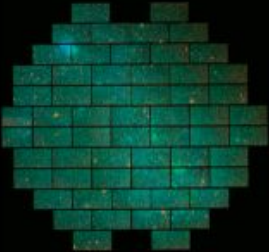
For Scale



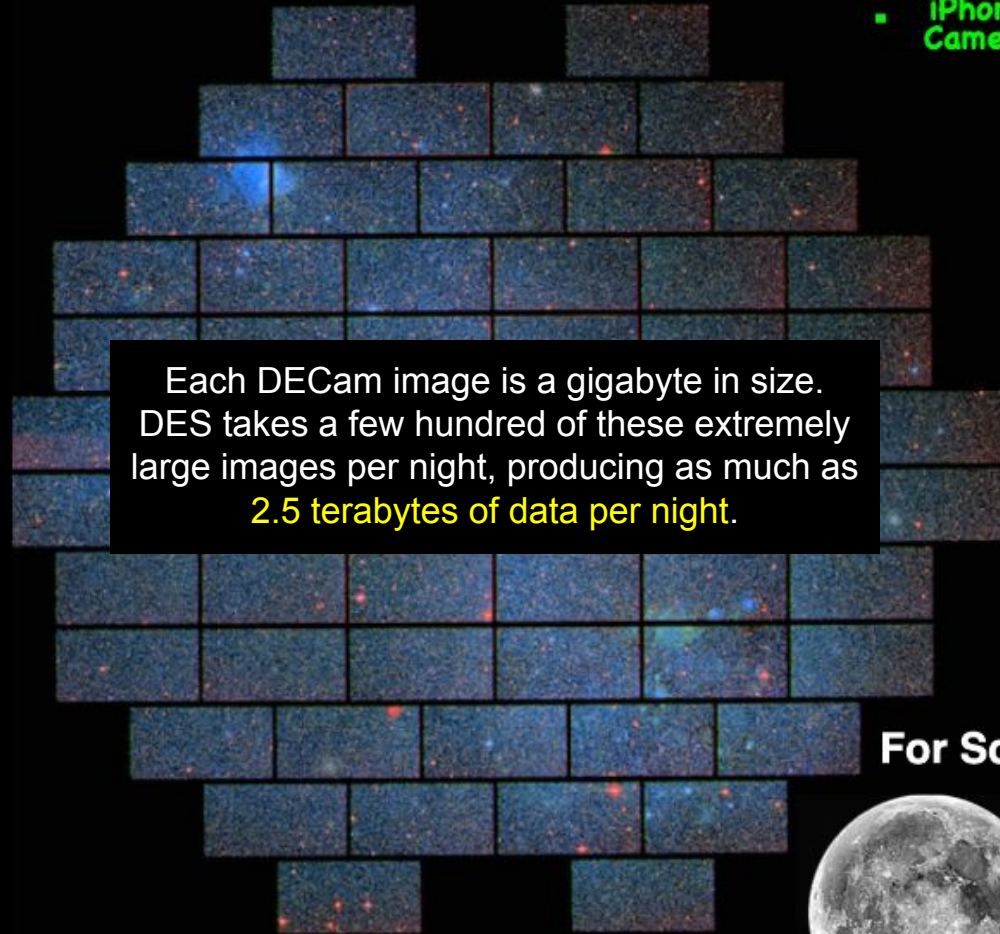




+

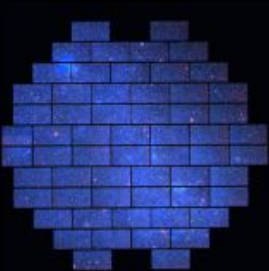


=



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.**

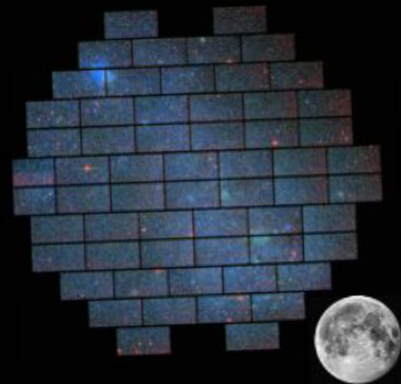
+



**+ 2 more filters**

**For Scale**



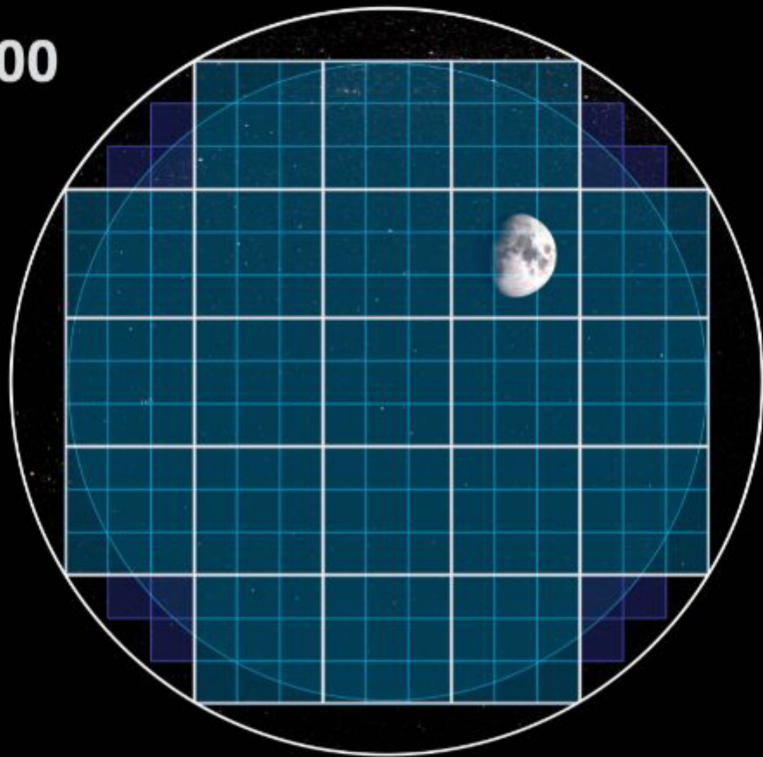


**DECam**  
**570,000,000**  
**pixels**

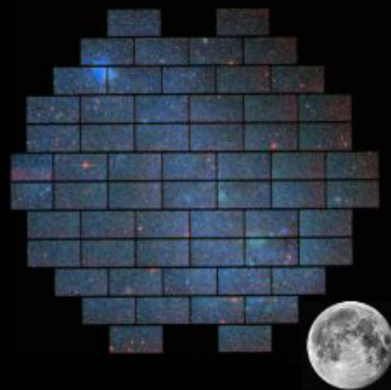


**HDTV**  
**2,000,000**  
**pixels**

LSSTCam  
3,200,000,000  
pixels



—  
HDTV  
2,000,000  
pixels



DECam  
570,000,000  
pixels

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

Enough 4k TVs to cover half a basketball court

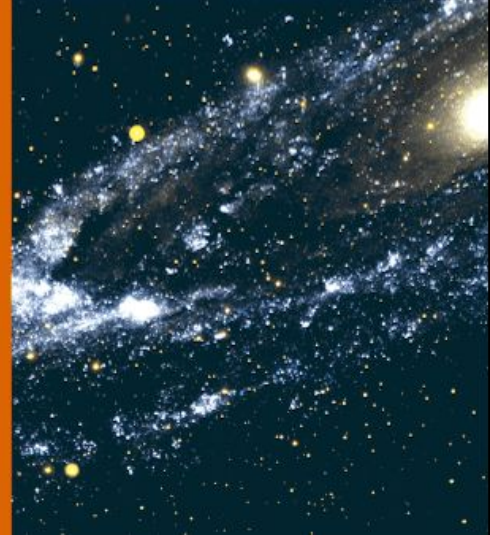




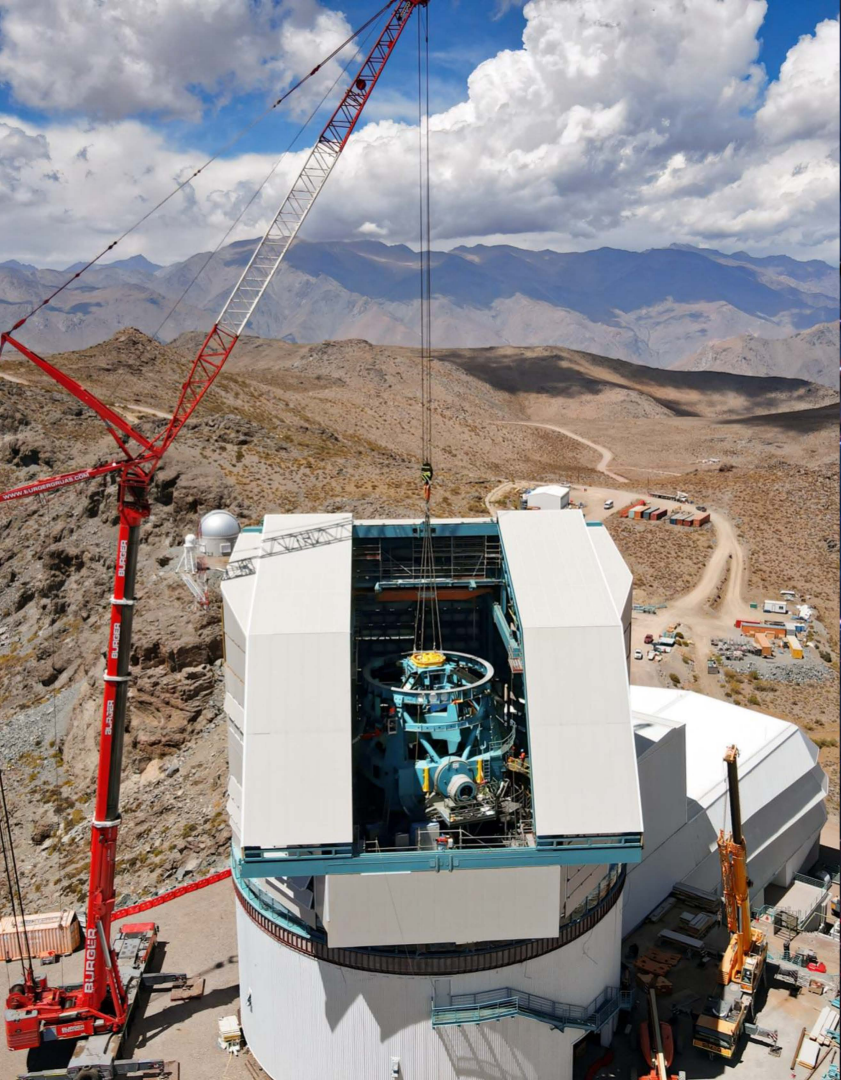
El Peñón peak of Cerro Pachón in northern Chile, alongside the existing Gemini South and Southern Astrophysical Research Telescope.



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







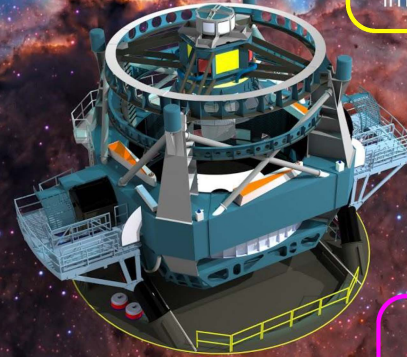
# LSST Project in Numbers

8,4 meters  
Primary mirror diameter

3 200 Megapixels  
Resolution of the Telescope Camera

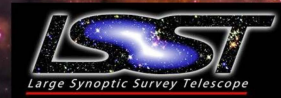
3 Nights  
Time needed for an all-sky imaging

1.23 F/D  
Telescope aperture



15 seconds  
Exposure time needed to capture an image

800 times  
Number of times a same object will be captured



37 Billion  
Number of celestial objects detected after 10 years

15 TB  
Amount of data collected every night

Credit image: Rolf Wahl Olsen  
[www.rolfolsenastronomyphotography.com](http://www.rolfolsenastronomyphotography.com)



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









## How do we...

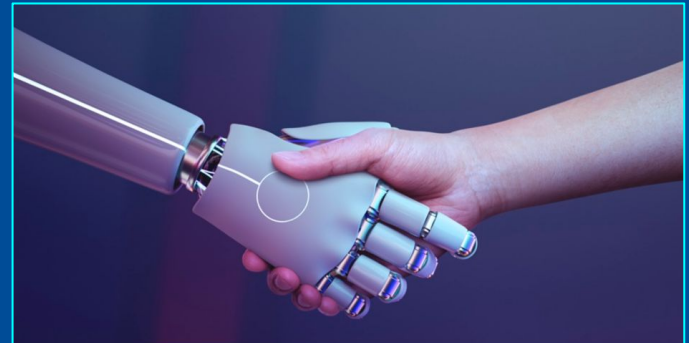
- ... process the data?
- ... search through the data?
- ... learn from the data?
- ... find unusual objects?
- ... send alerts quickly?
- ... combine/compare datasets?
- ...
- ...
- ....





## 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?

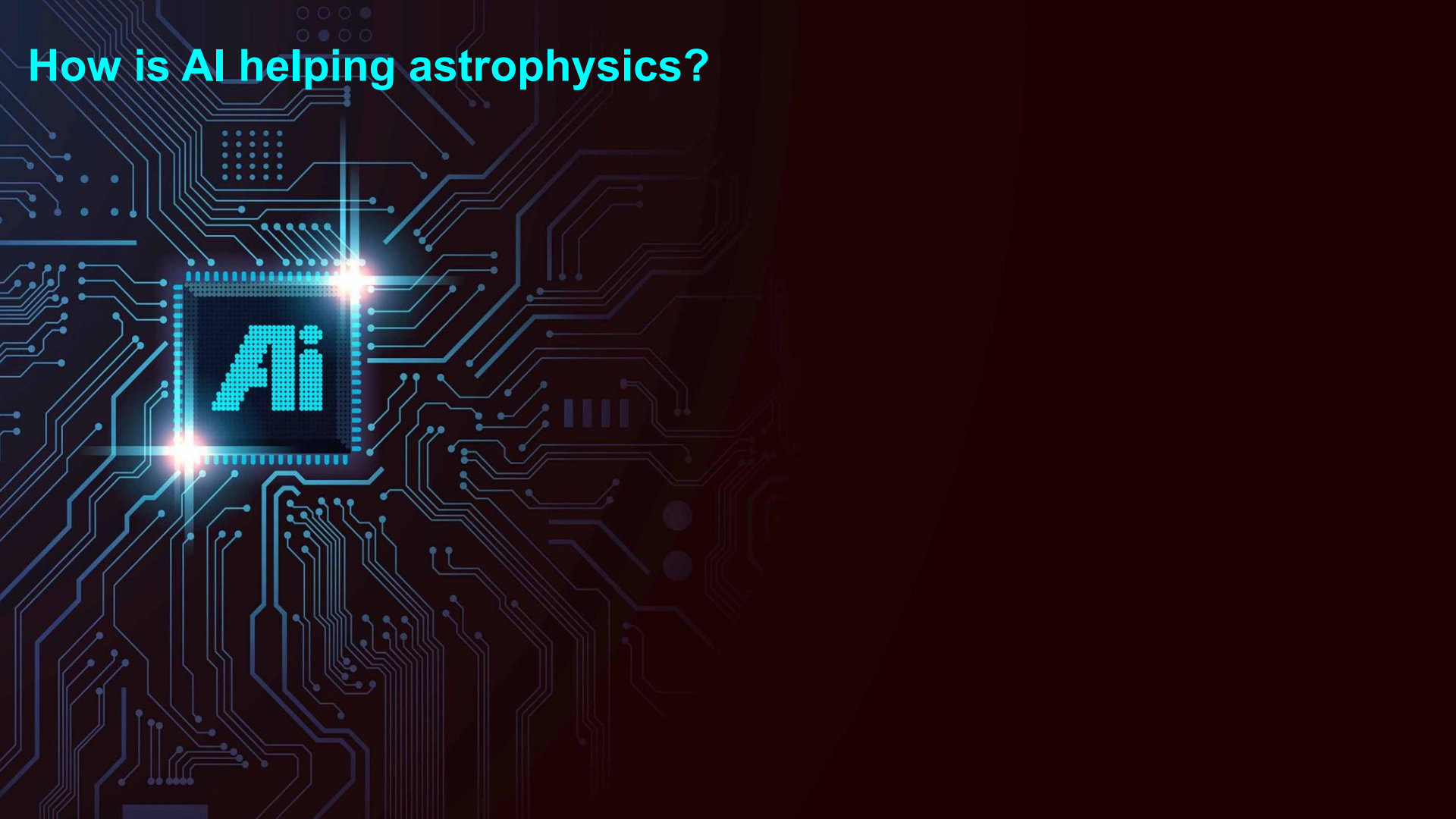
Case Studies

Challenges and Opportunities

Conclusion



# How is AI helping astrophysics?



# How is AI helping astrophysics?

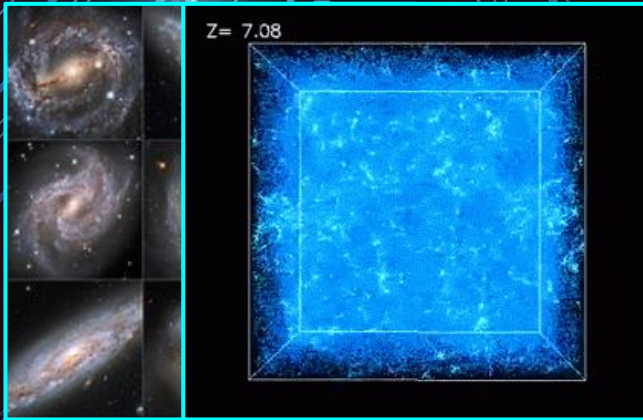
Image/data processing and analysis



# 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**





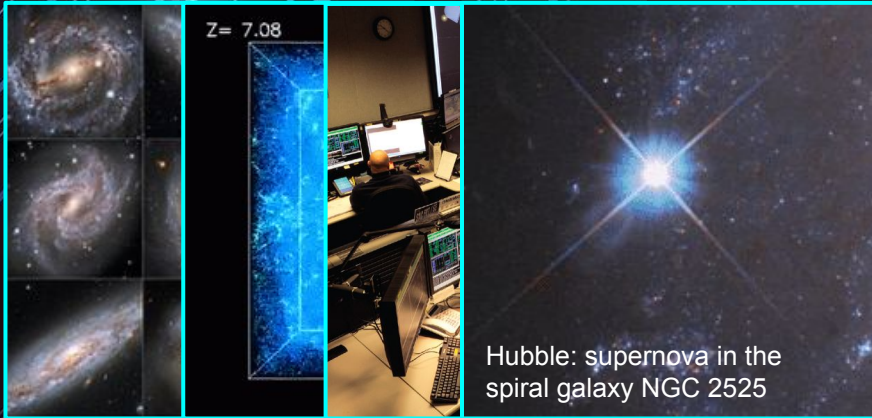
# How is AI helping astrophysics?

**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

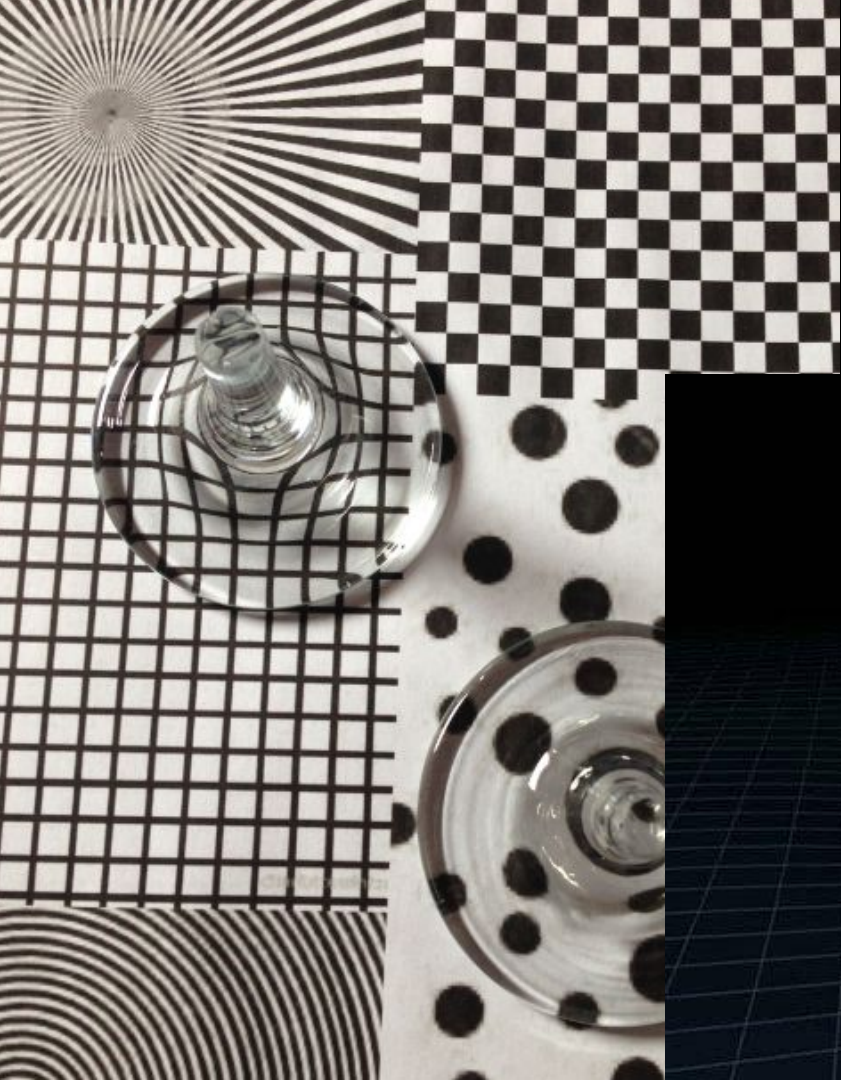




**Galaxy images are  
distorted!**

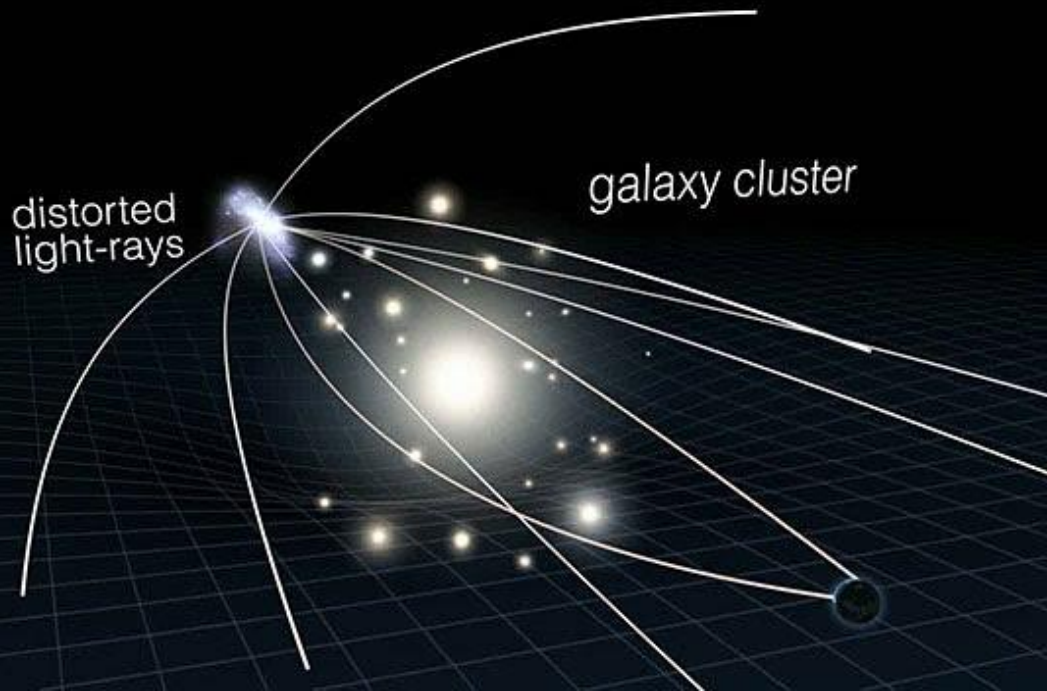




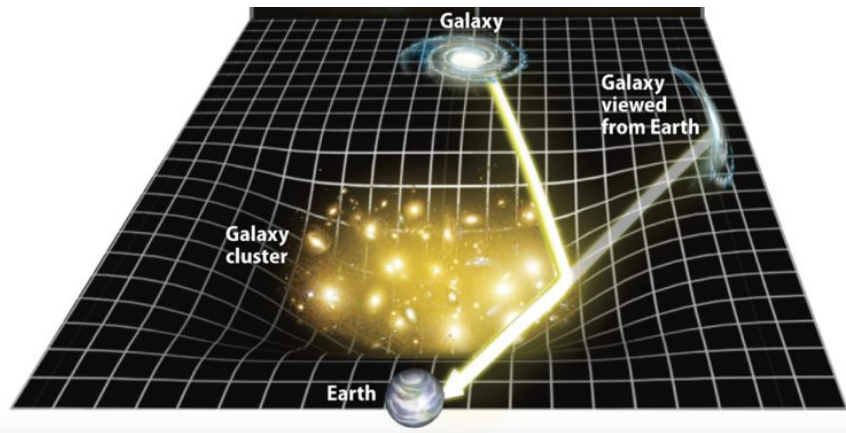


## Gravitational Lensing

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



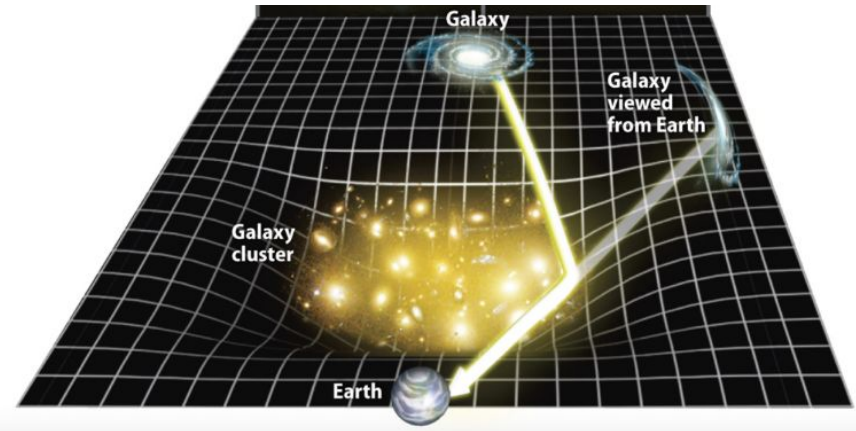
Compared to regular galaxies, lensed galaxies are quite rare!





Compared to regular galaxies, lensed galaxies are quite rare!

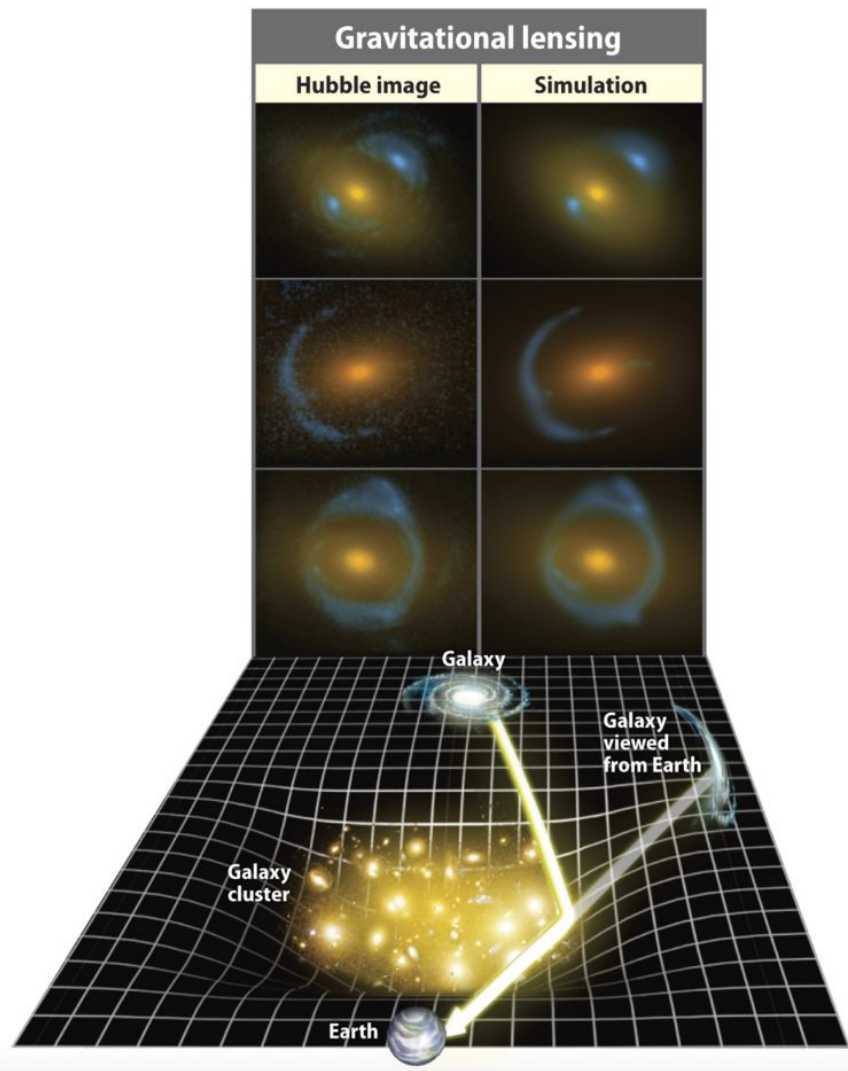
We can learn a lot about the evolution of our Universe and it's content!



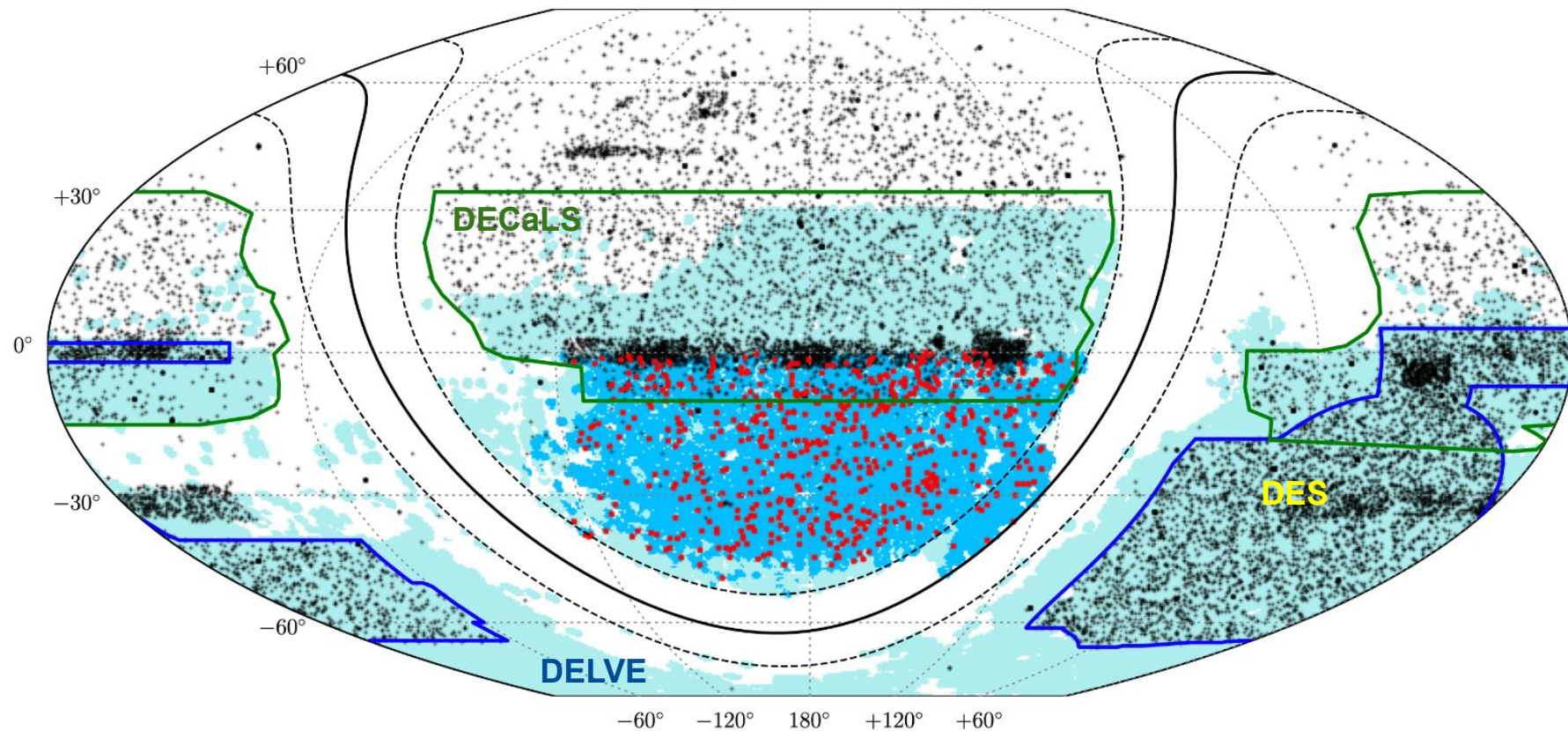
Compared to regular galaxies, lensed galaxies are quite rare!

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)

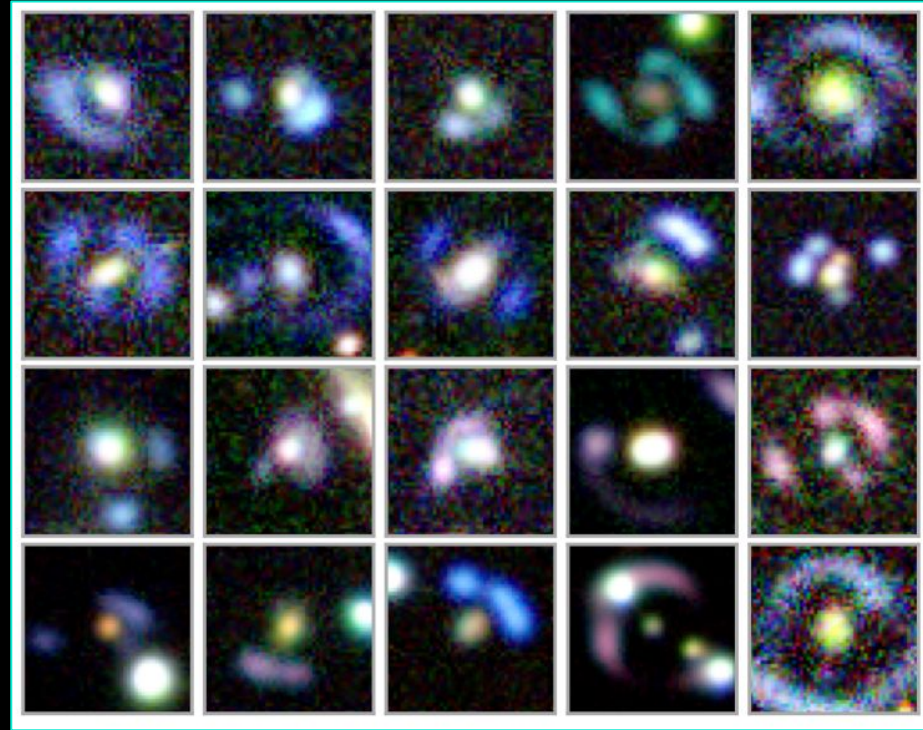




Non-lens

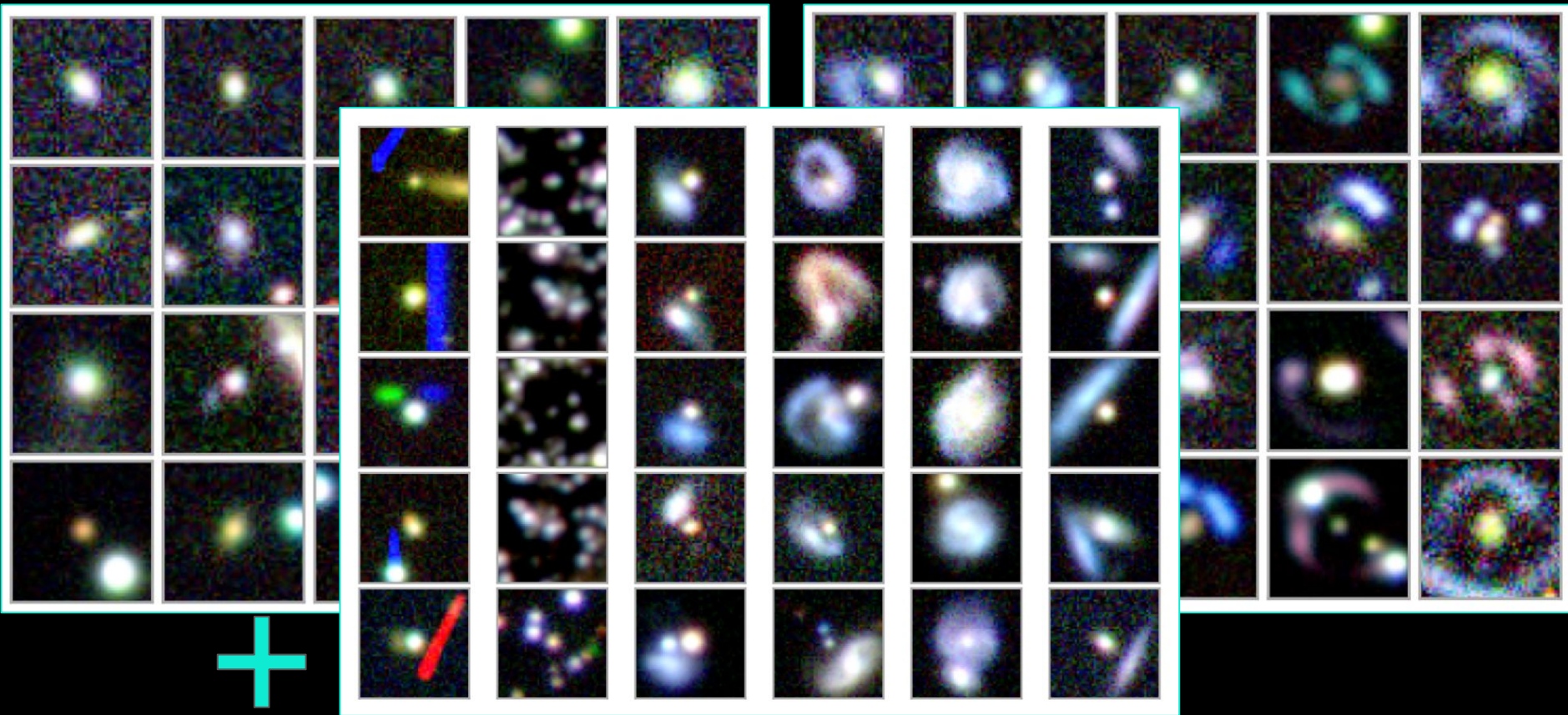


Lens



Non-lens

Lens





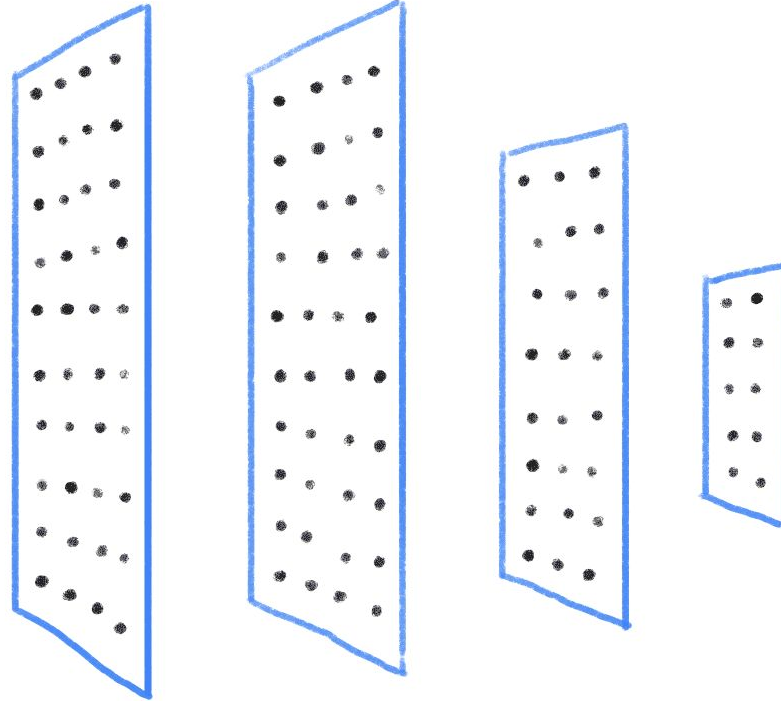
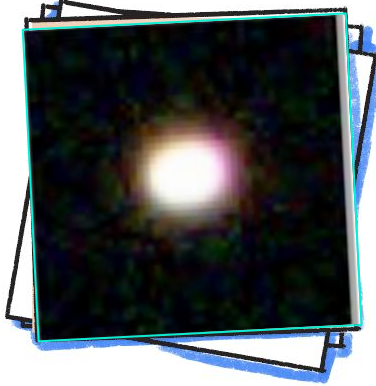
Model is trained with examples of lensed and non-lensed galaxies to learn to distinguish between them.

LENS



(LABLED)  
PHOTOS

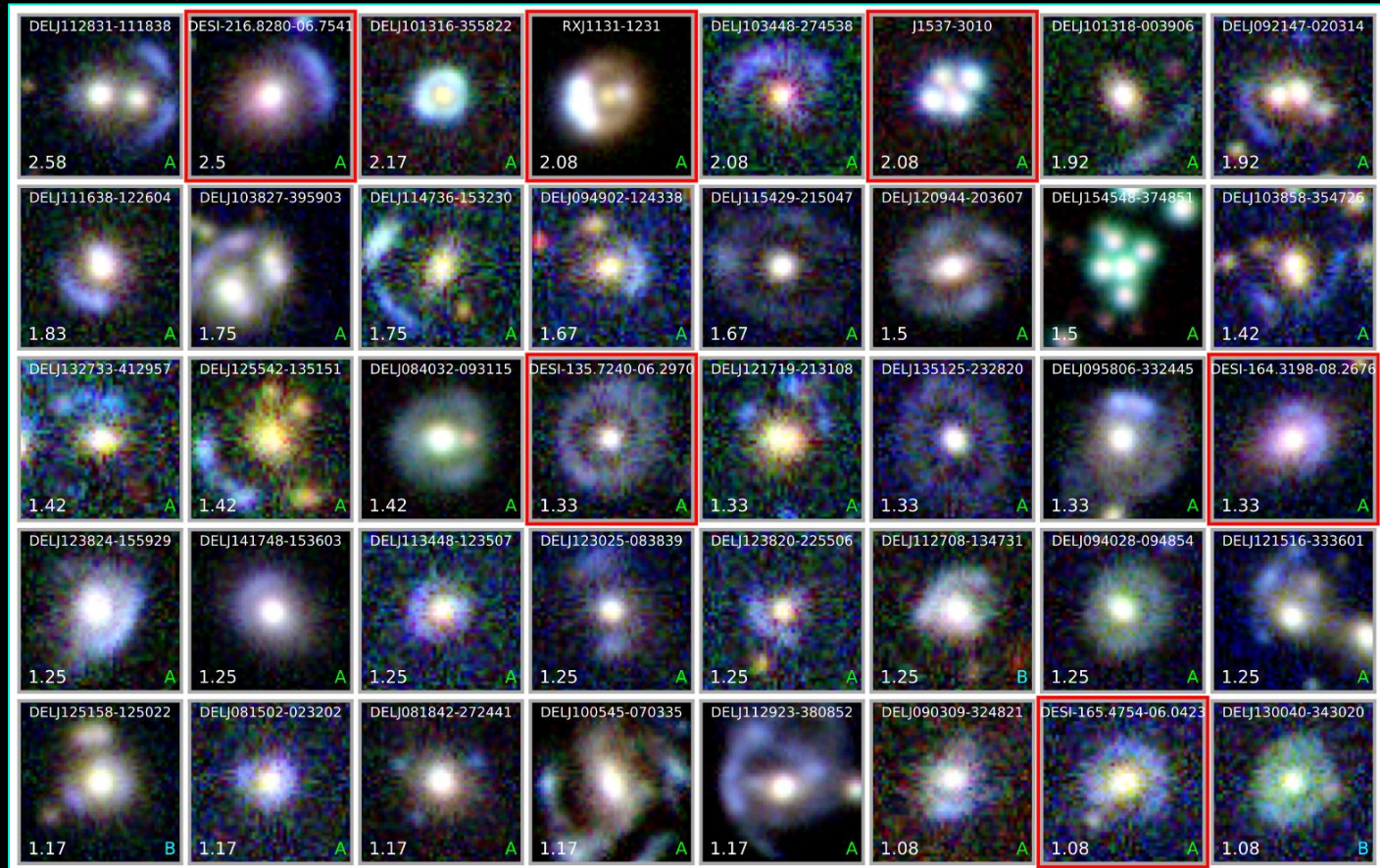
NON-LENS



OUTPUT

LENS





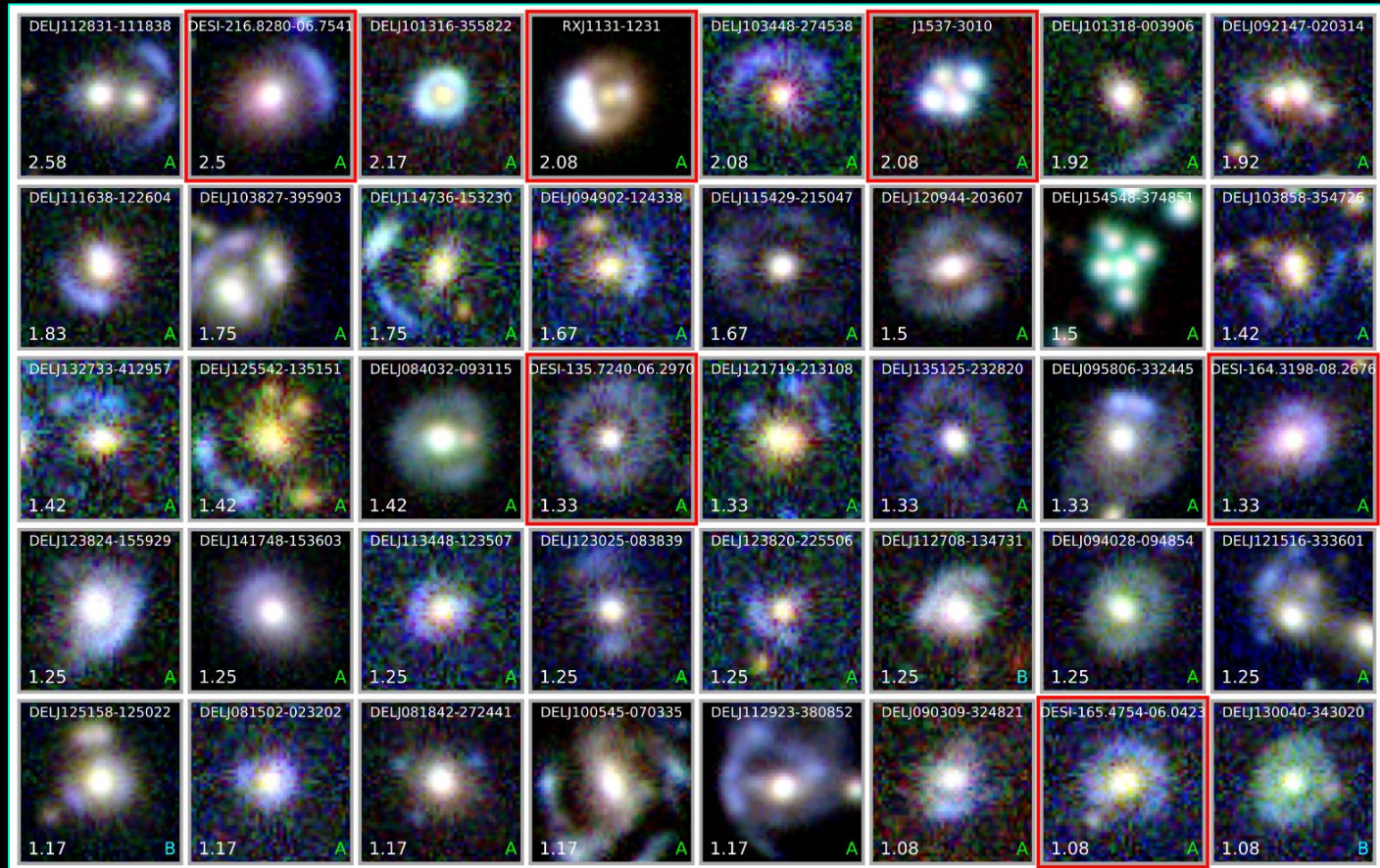
CNN finds 50,000  
candidates



Two rounds of  
visual inspection



617 final  
candidates



CNN finds 50,000  
candidates



Two rounds of  
visual inspection

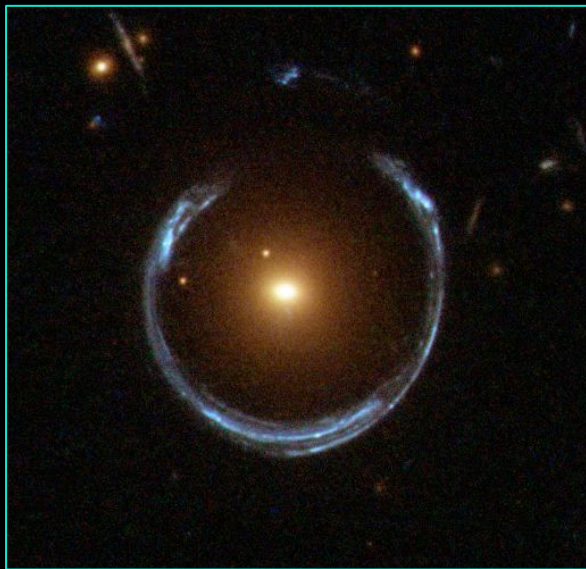


617 final  
candidates

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



# Why is this important?



**Probe galaxy structure**

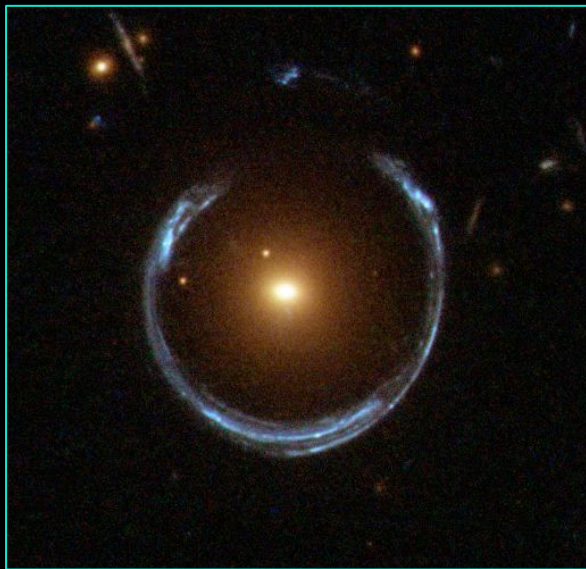
**Probe distribution of matter**

**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**

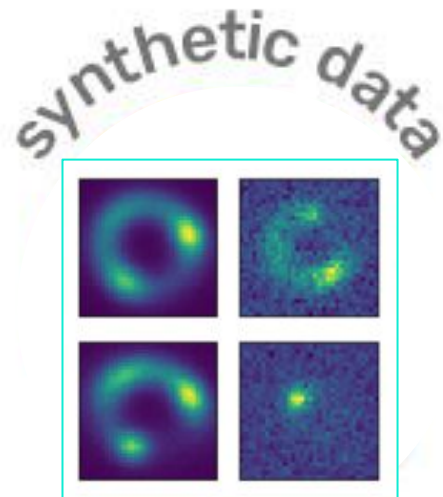
**Cosmology and how the universe evolves**

## How?

By inferring properties of the lens system.

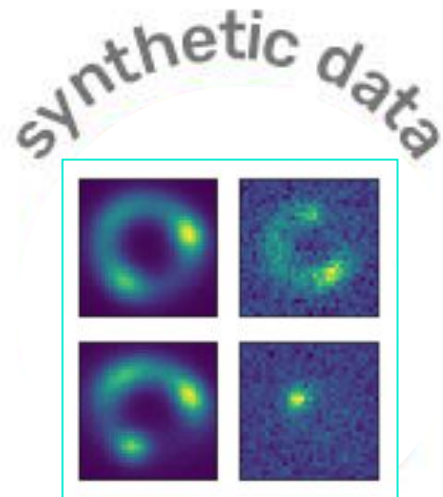
# Simulation is easy....inference is hard!

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



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

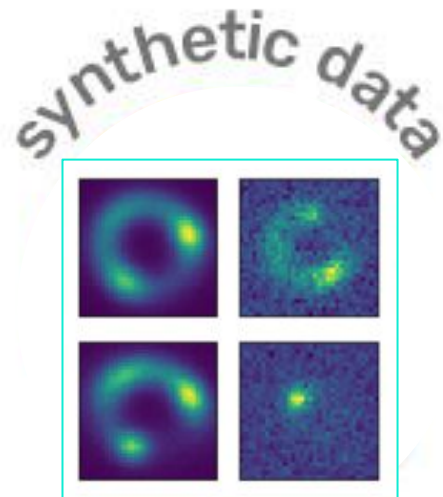
Galaxy source and lens masses, shapes, positions, gravity, observational effects...





# Simulation 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...

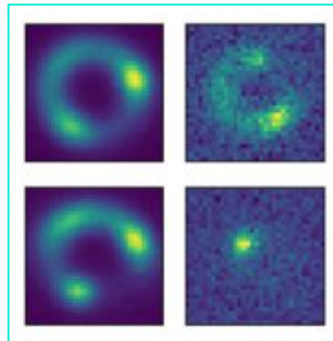
Detector properties, light propagation, galaxy properties, cosmology....

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

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....

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





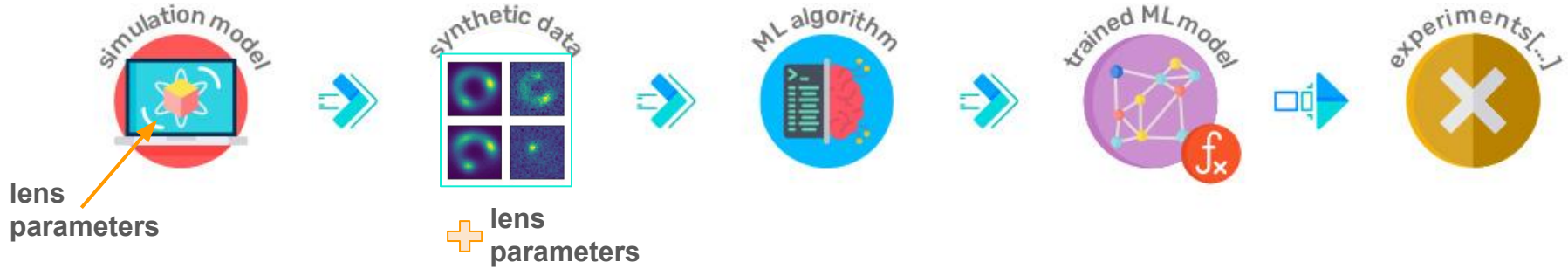
But with AI...

It's not necessary to explicitly calculate the integral!



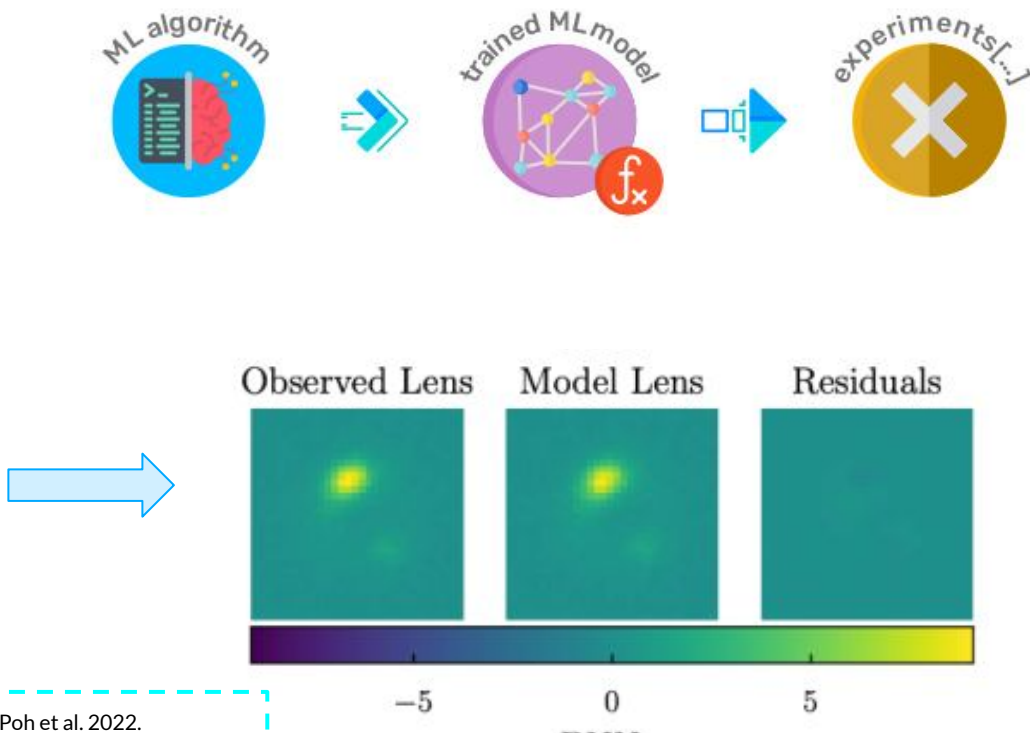
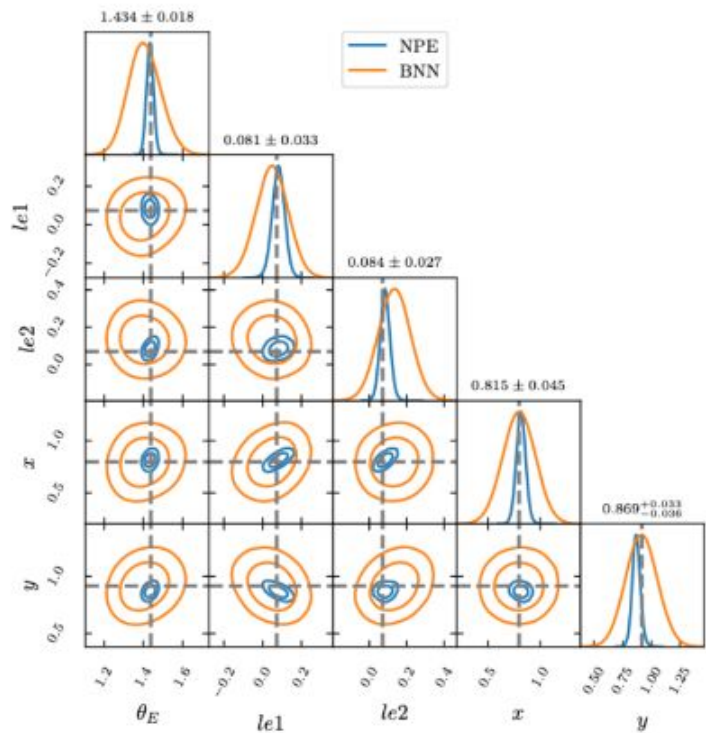
## But with AI...

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.

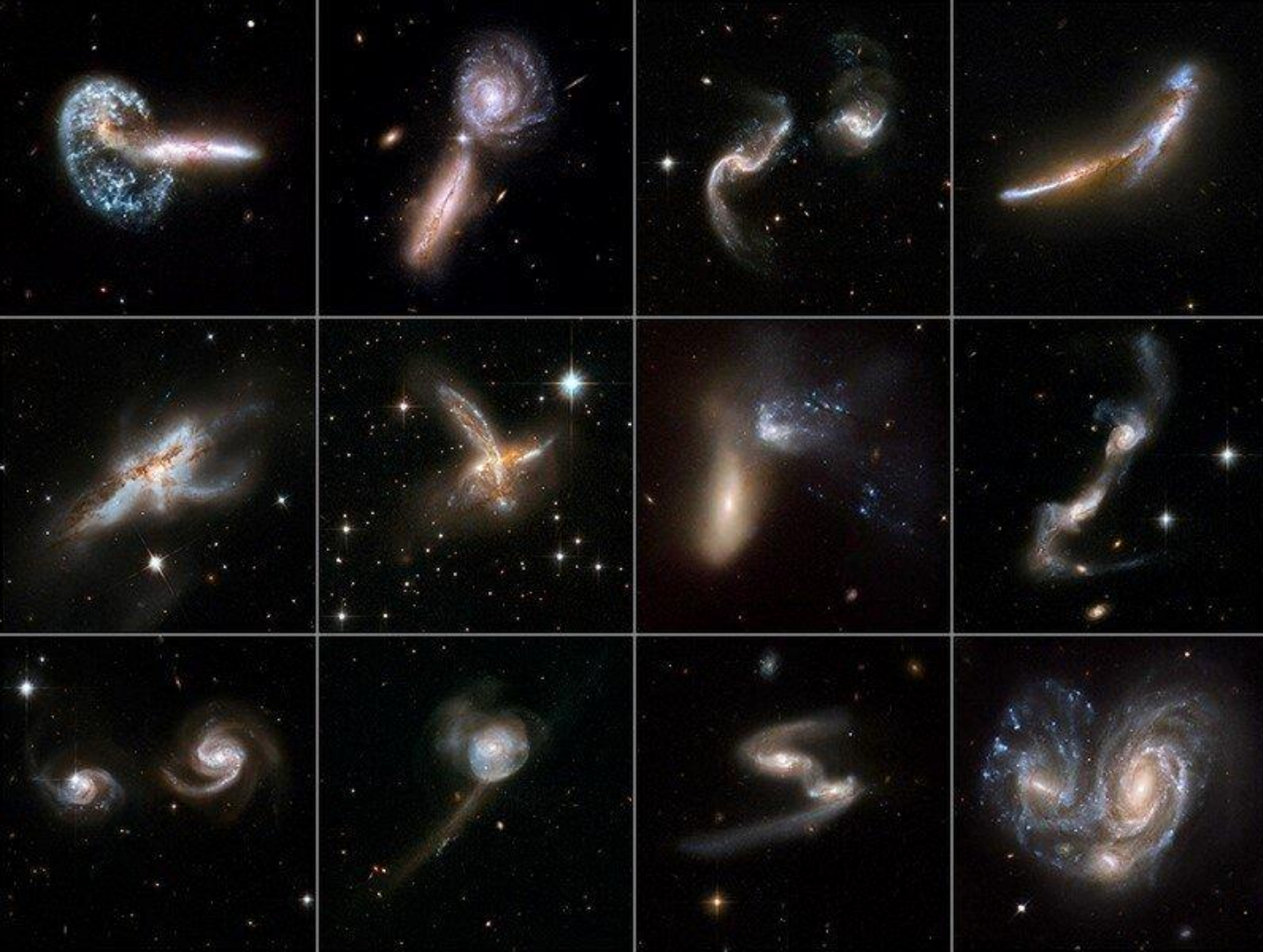




## Example 2

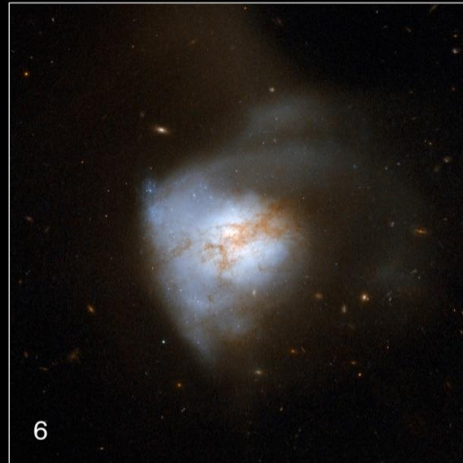
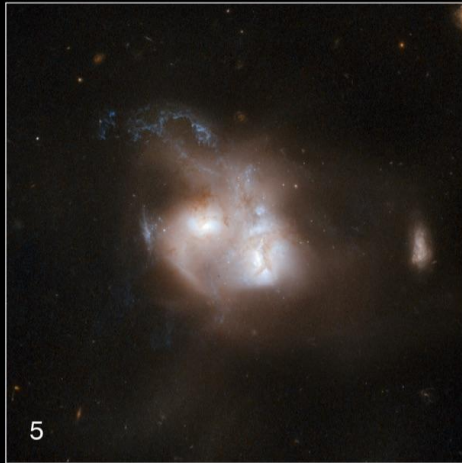
How about  
these  
weird-looking  
galaxies?

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



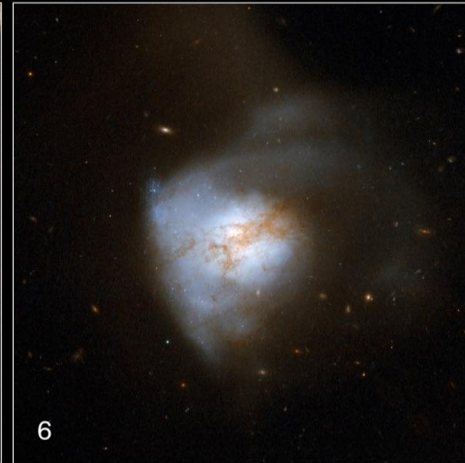
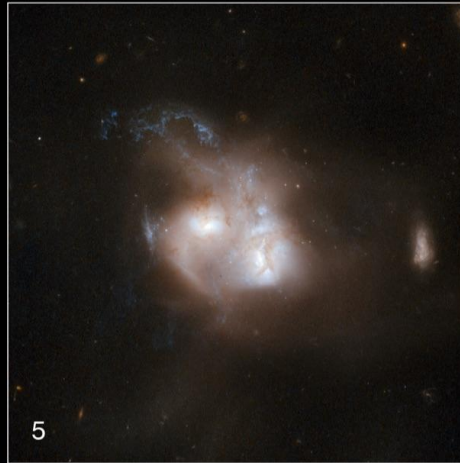
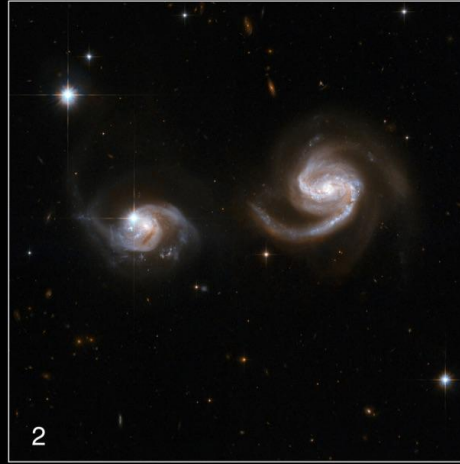
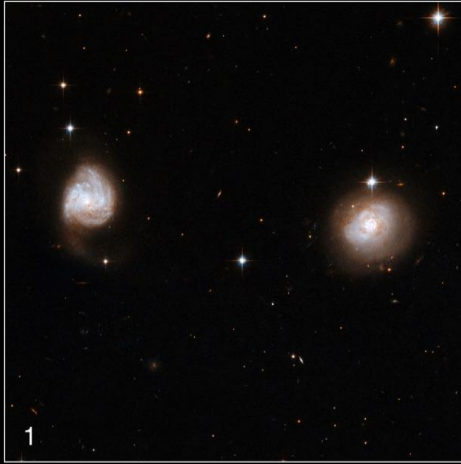
**How about  
these  
weird-looking  
galaxies?**

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!

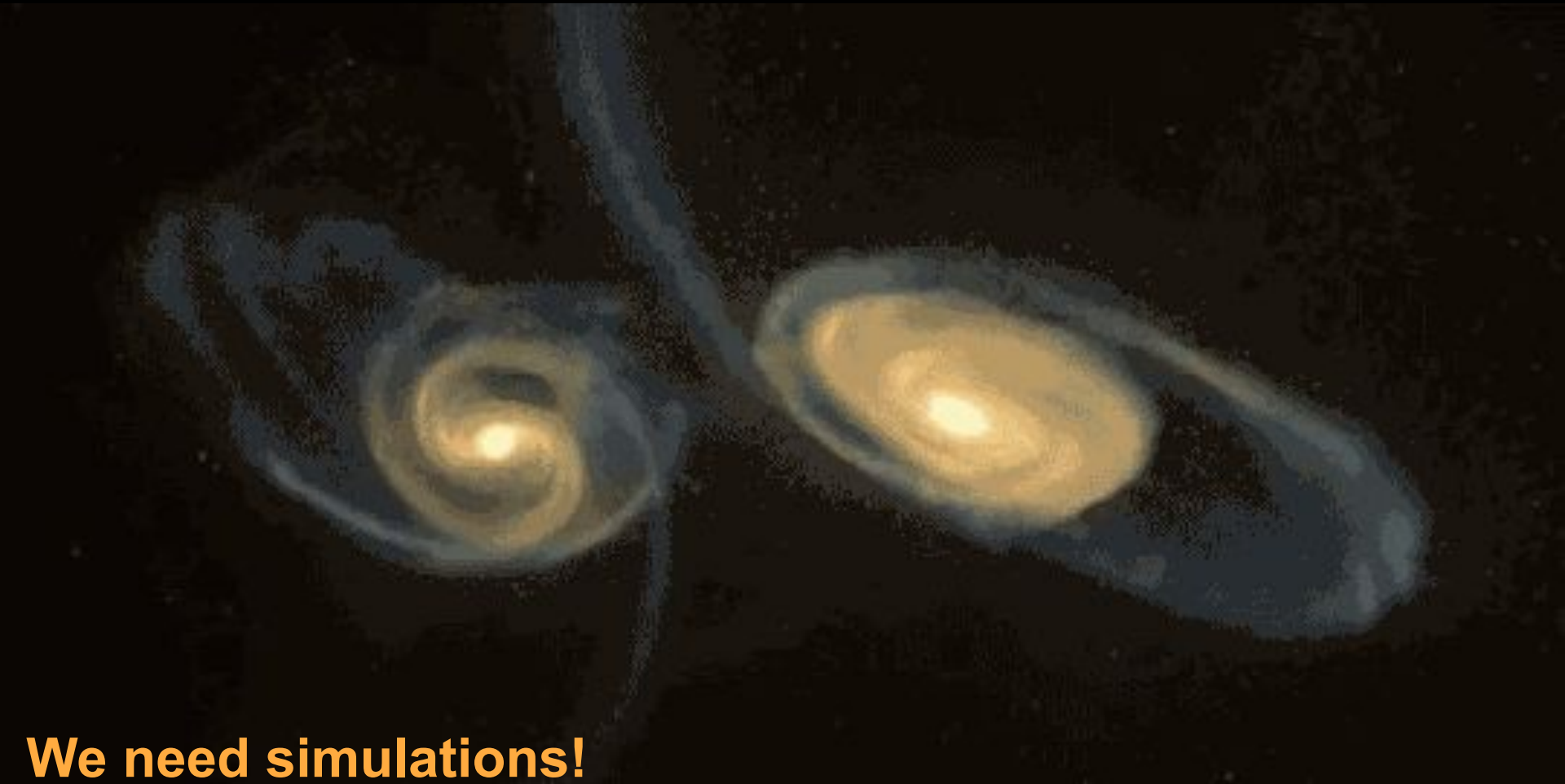




If we want to study the process we need a lot of galaxy pairs in different merger stages!

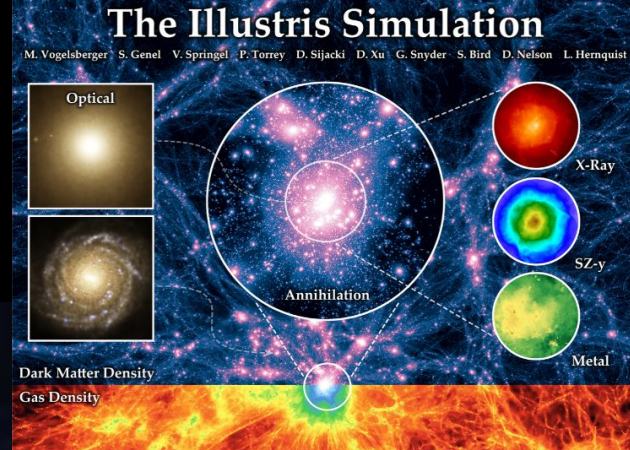
Unfortunately we cannot easily use observations for that...

Are galaxies really merging or just visually overlapping?



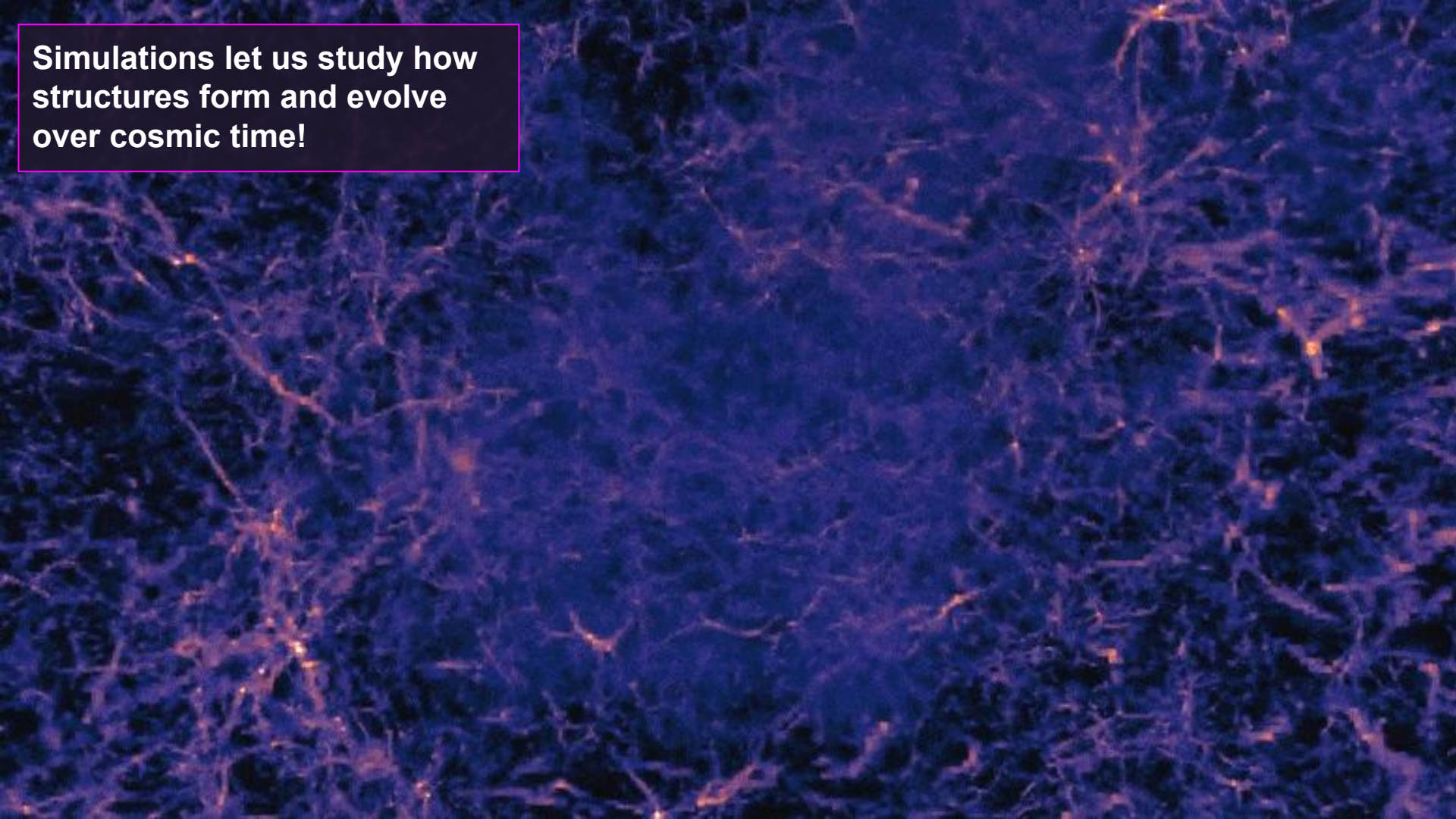
**We need simulations!**

**We need even bigger simulations  
with many galaxies!**

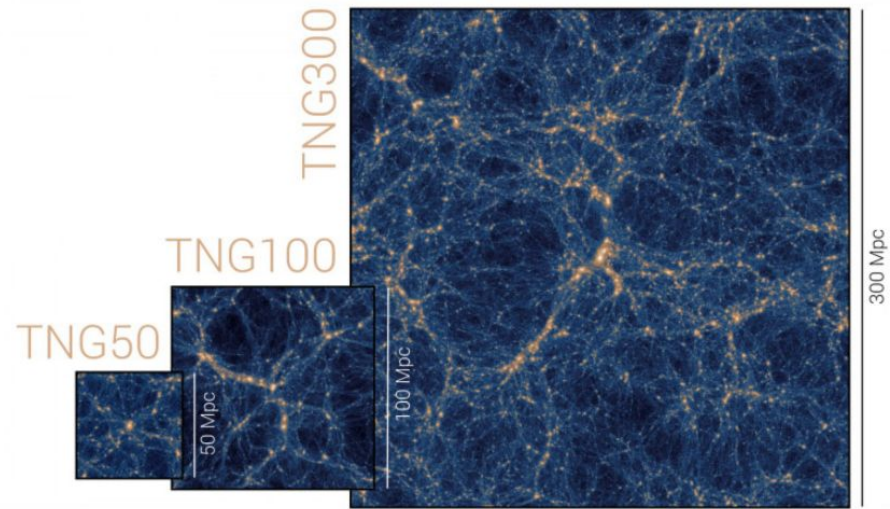


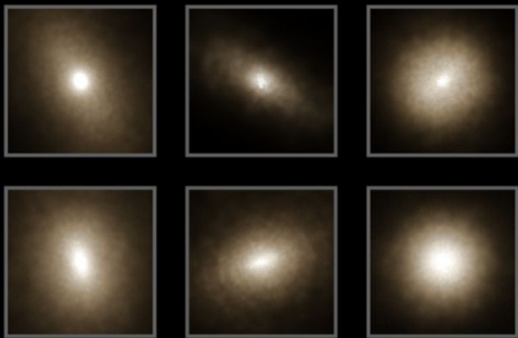


**Simulations let us study how  
structures form and evolve  
over cosmic time!**

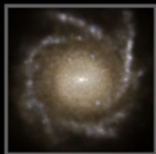
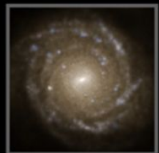


1pc = 20 trillion miles

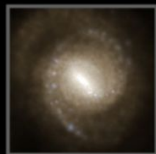
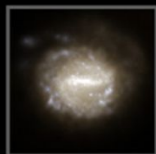




ellipticals



disk galaxies



irregular

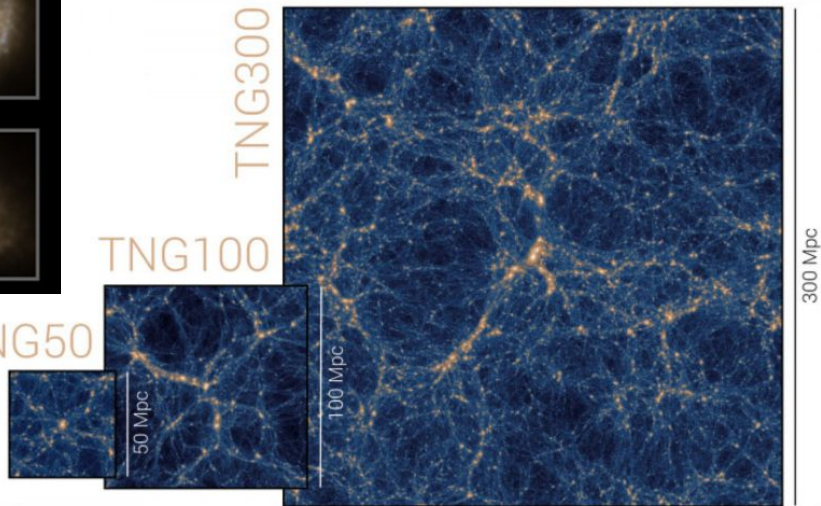


1pc = 20 trillion miles

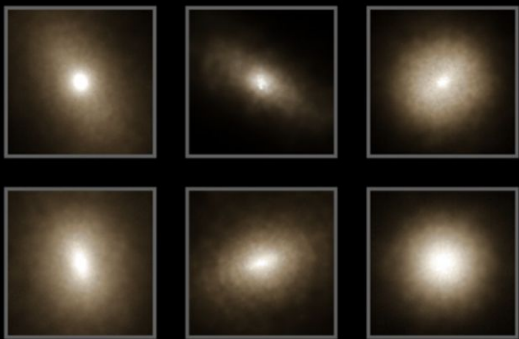
TNG50

TNG100

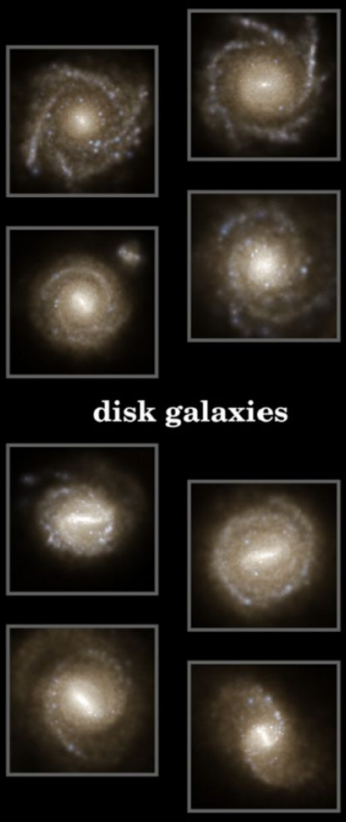
TNG300





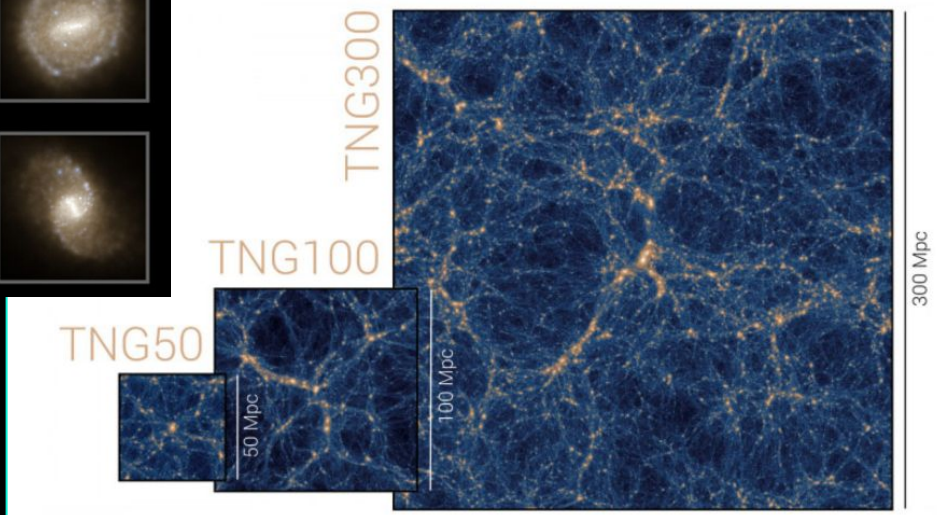
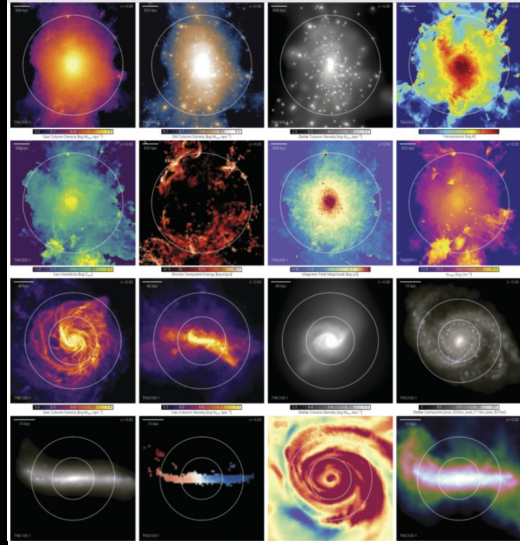
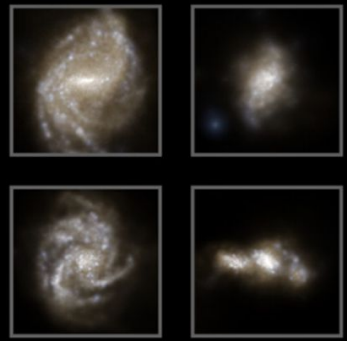


ellipticals



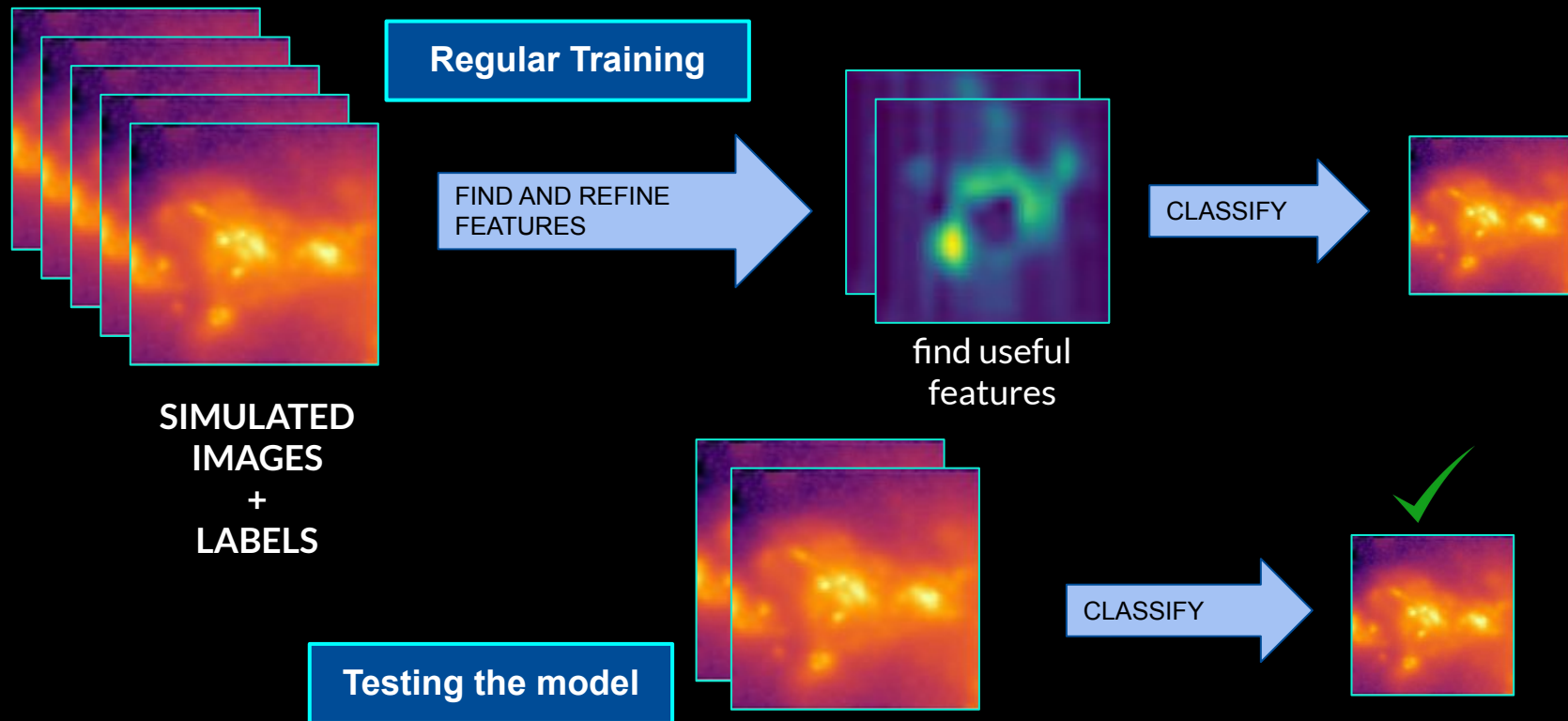
disk galaxies

irregular

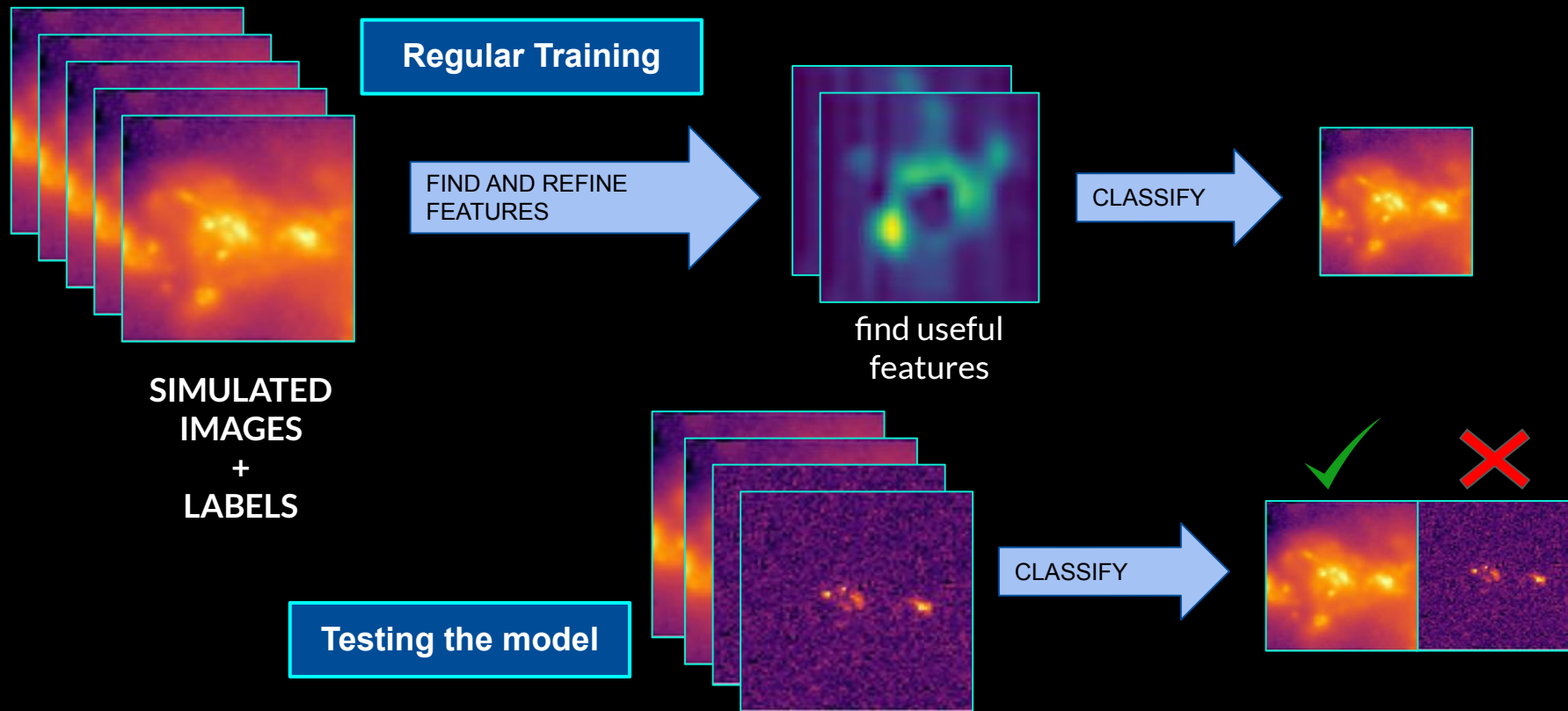


1pc = 20 trillion miles

# Let's train the model:



# Let's train the model:





# Combining Datasets

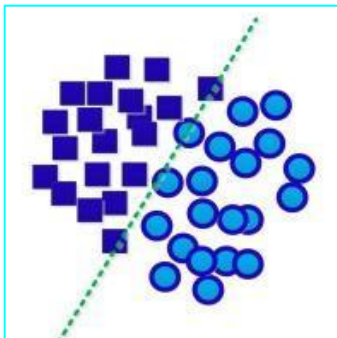
Why does this happen?

# Combining Datasets

Why does this happen?

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

Source Domain

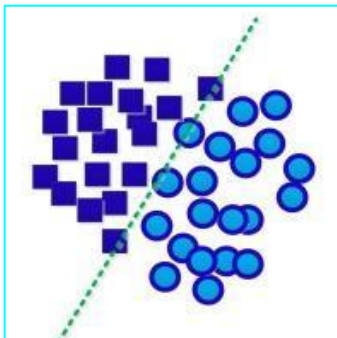


# Combining Datasets

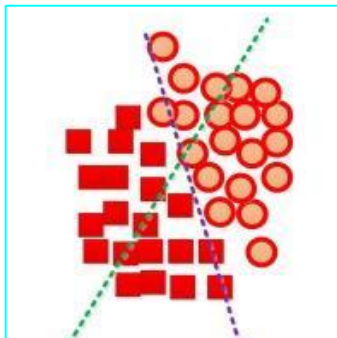
## Why does this happen?

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

Source Domain



Target Domain



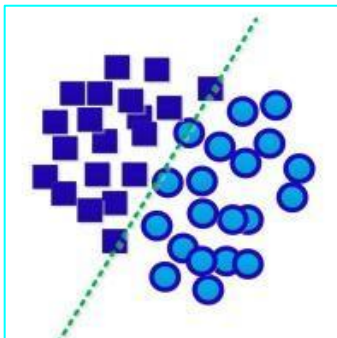


# Combining Datasets

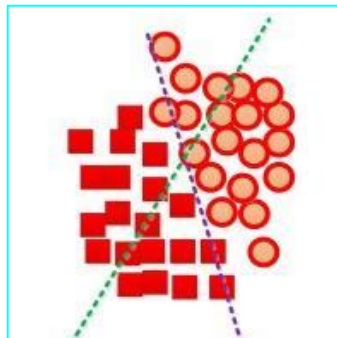
Why does this happen?

We need to align  
the data during  
training!

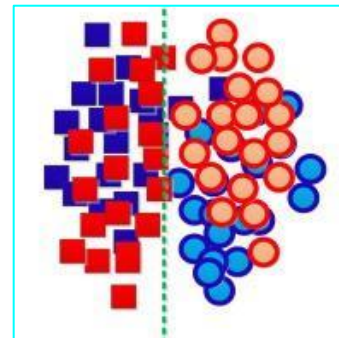
Source Domain



Target Domain



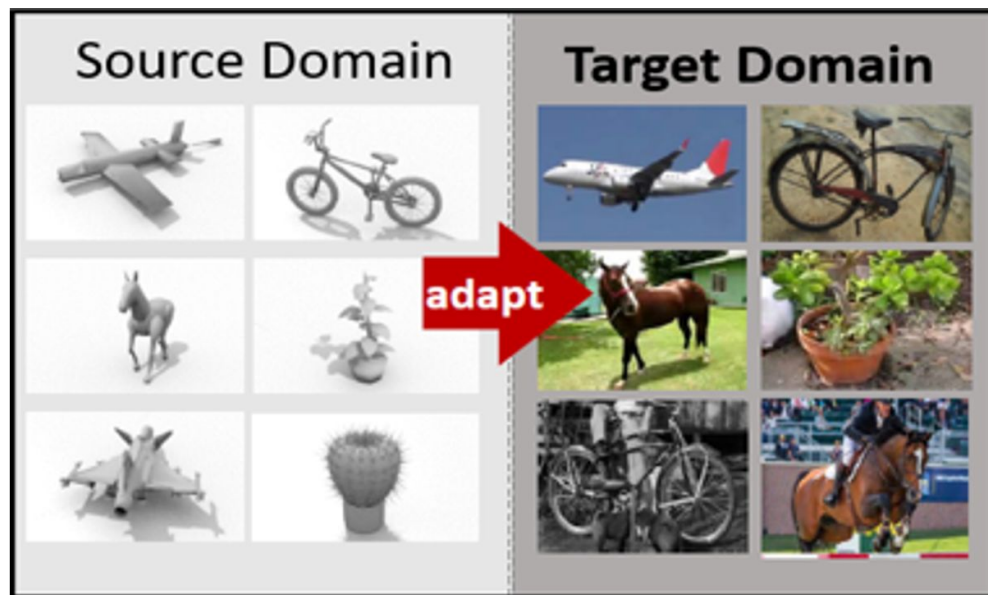
Domain Alignment



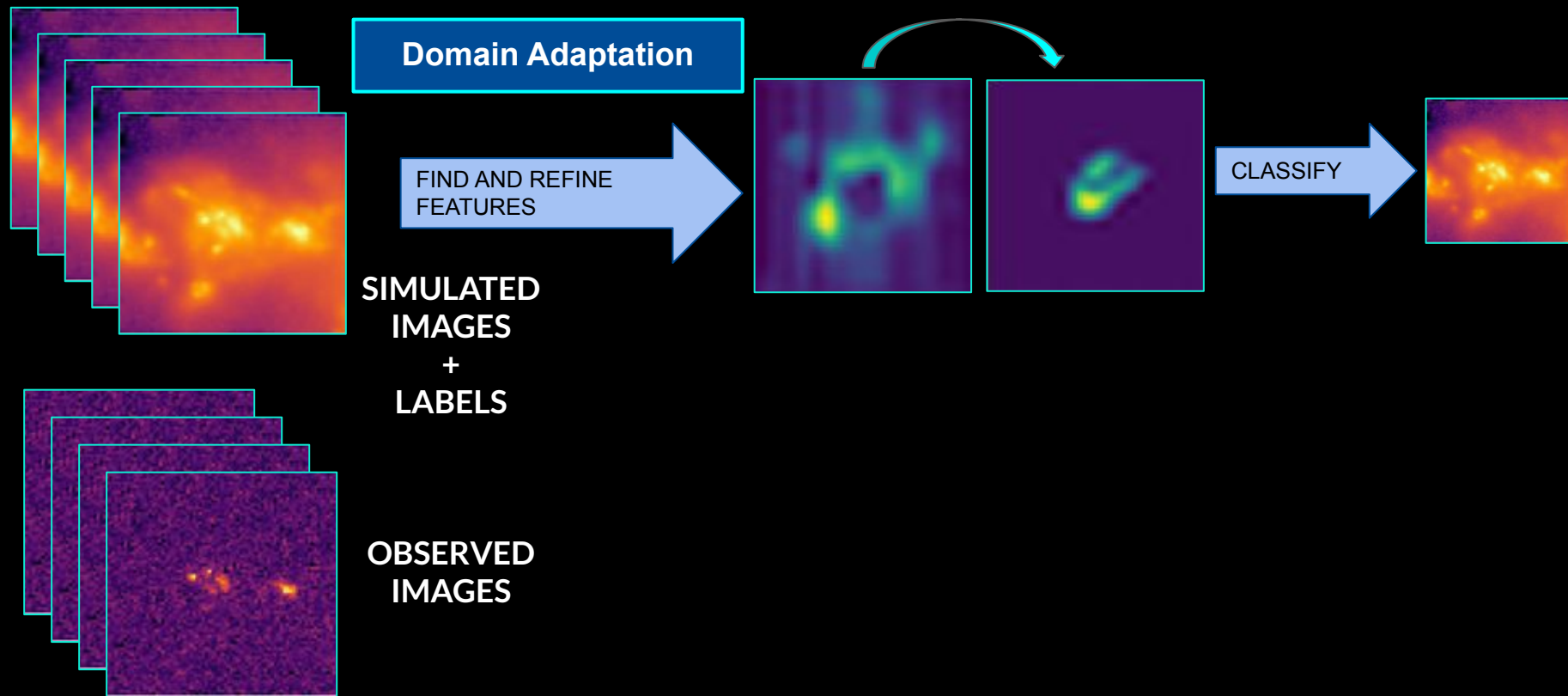
# Solving the dataset shift problem!

Train together and learn to use domain-invariant features.

Source - day      Target - night, rain

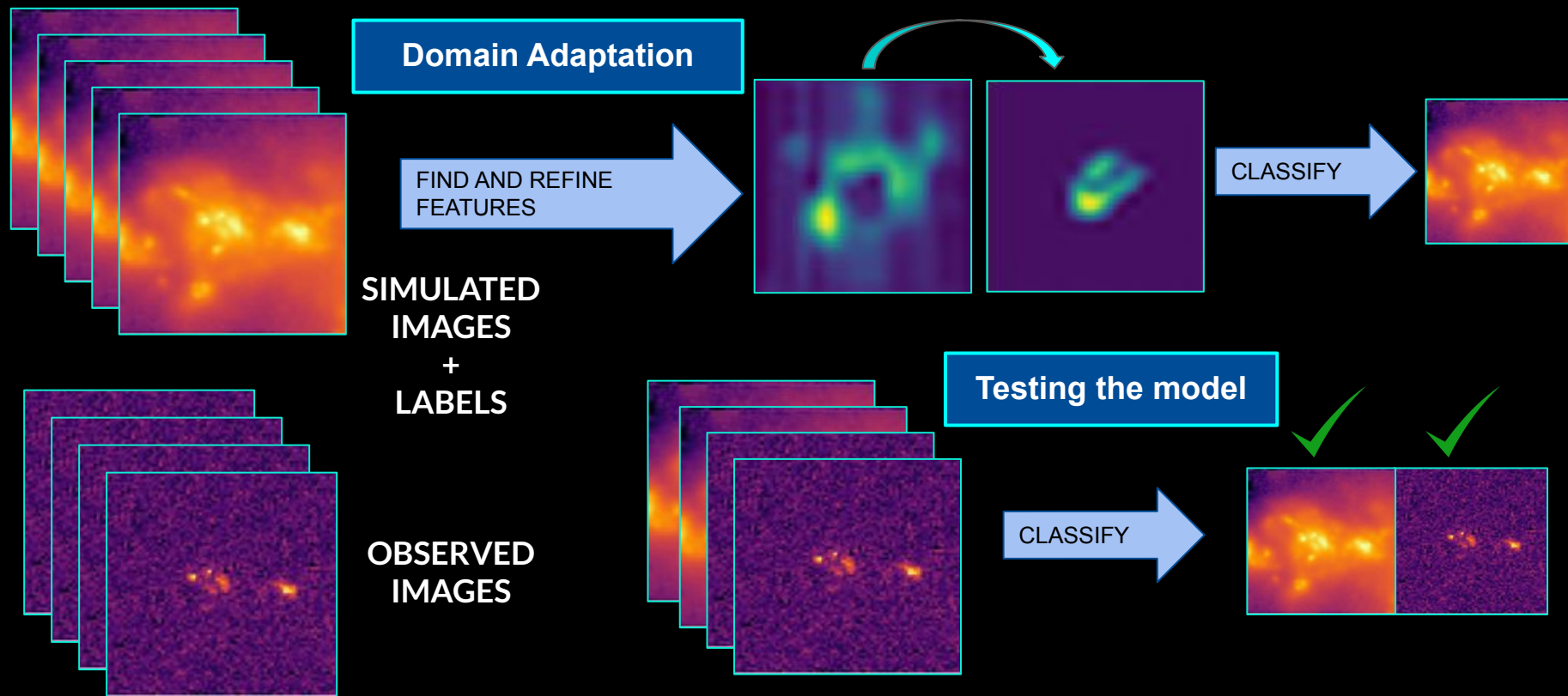


# Combining Datasets

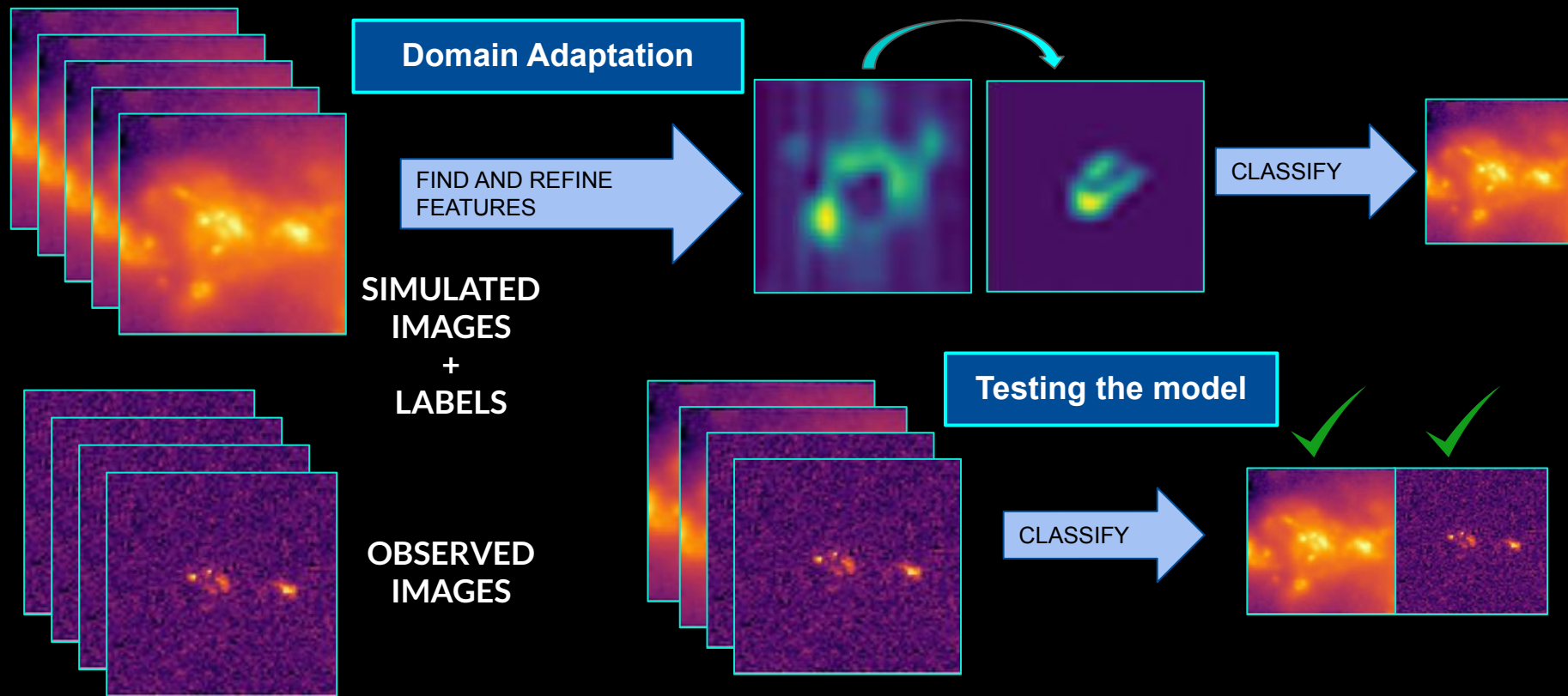




# Combining Datasets



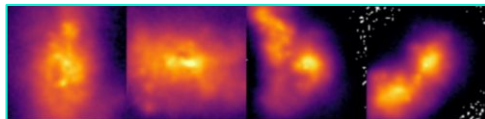
# Combining Datasets



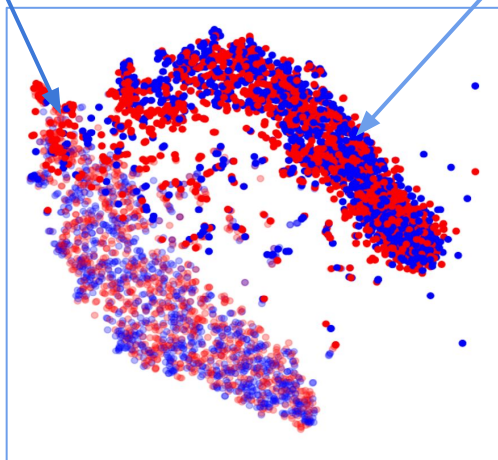
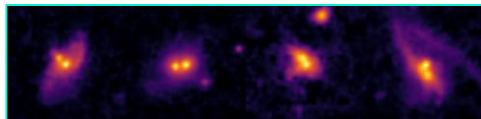
**Enable merger searches in both clean and real noisy images!**

# Combining Datasets

Source - Illustris



Target - SDSS observations



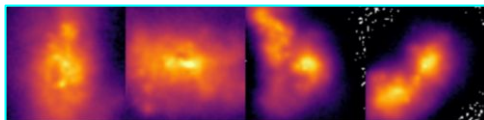
← This is how the network sees the data.  
2D representation of network's latent space.

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

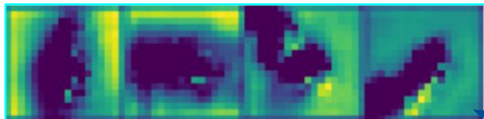


# Combining Datasets

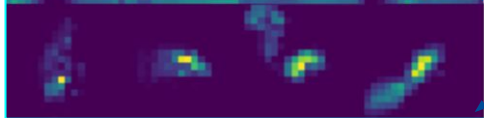
Source - Illustris



M



NM



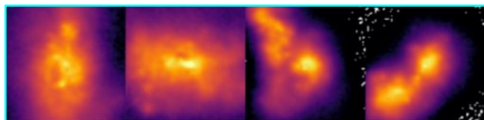
Important regions are highlighted!

Regular Training

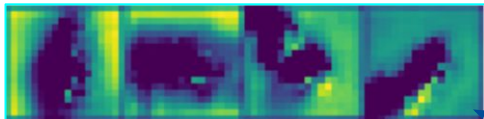
Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

# Combining Datasets

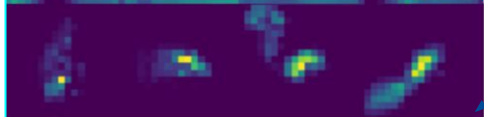
Source - Illustris



M



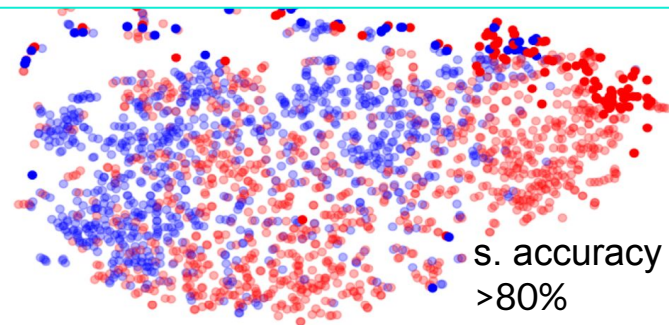
NM



Important regions are highlighted!

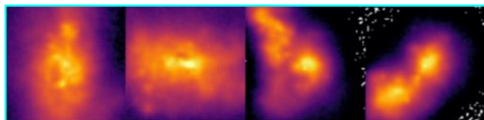
Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

Regular Training

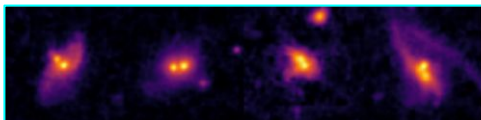


# Combining Datasets

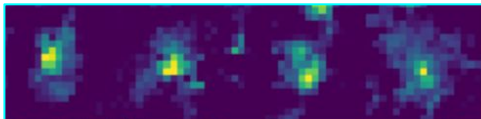
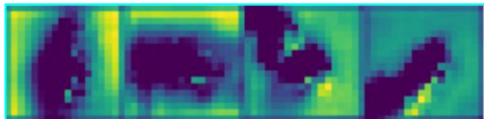
Source - Illustris



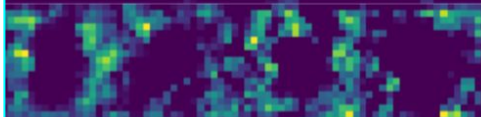
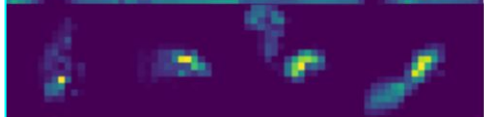
Target - SDSS observations



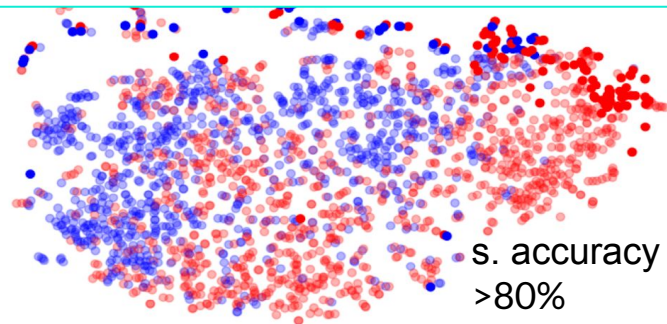
M



NM



Regular Training

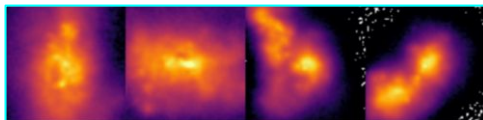


Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

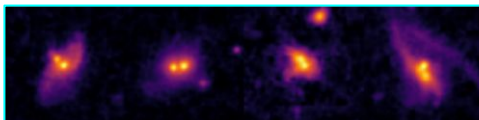


# Combining Datasets

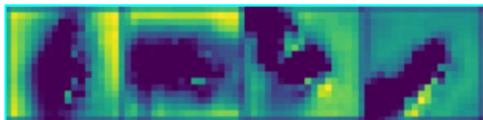
Source - Illustris



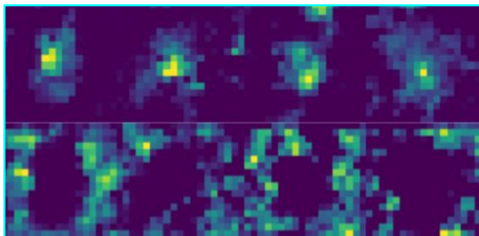
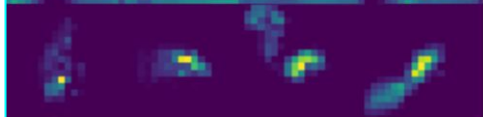
Target - SDSS observations



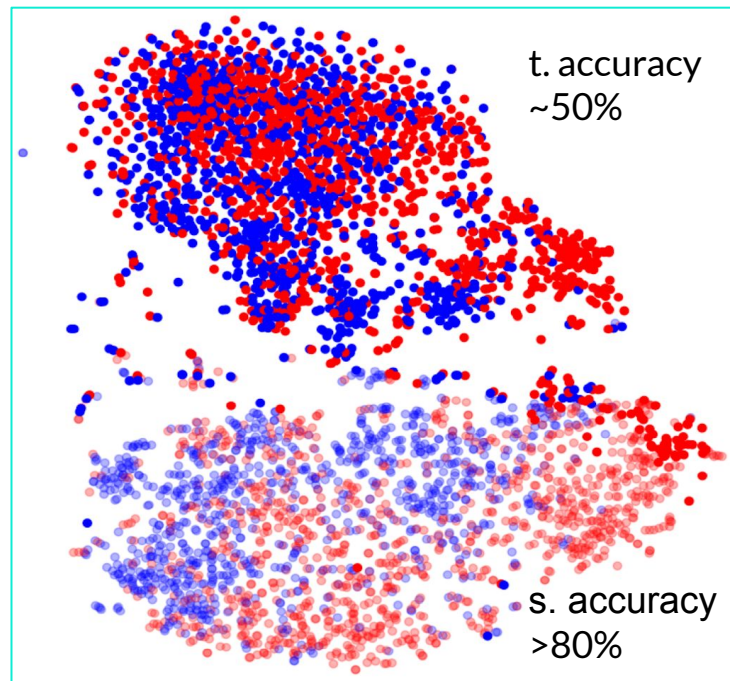
M



NM

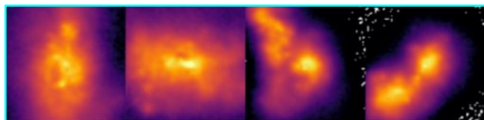


Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

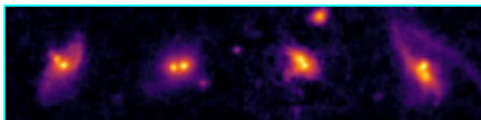


# Combining Datasets

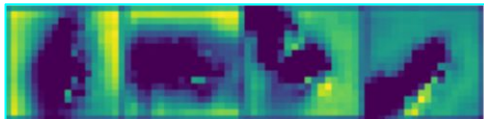
Source - Illustris



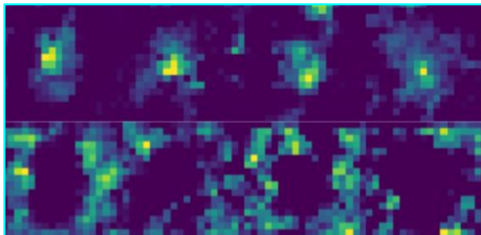
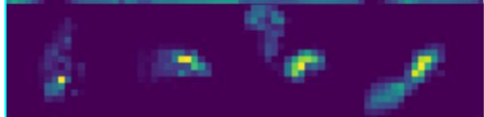
Target - SDSS observations



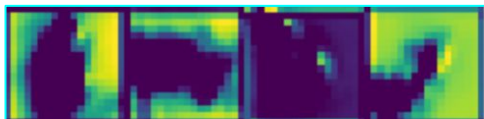
M



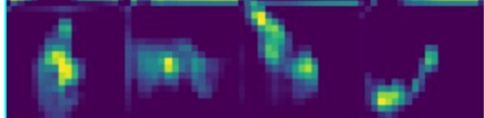
NM



M



NM

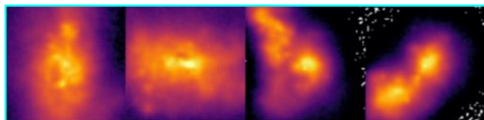


Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

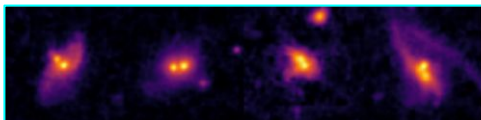
Domain Adaptation

# Combining Datasets

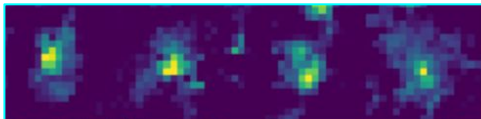
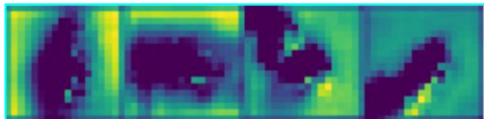
Source - Illustris



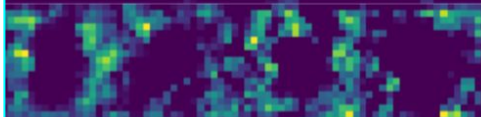
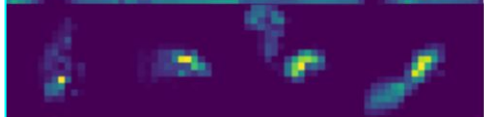
Target - SDSS observations



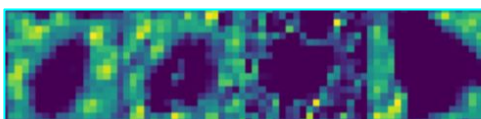
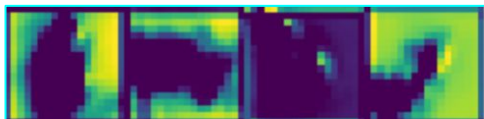
M



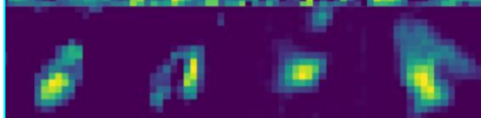
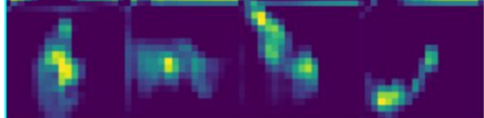
NM



M



NM



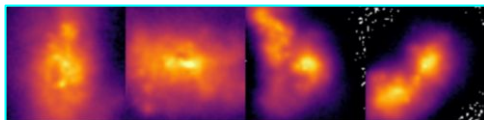
Ćiprijanović et al. 2020.  
Ćiprijanović et al. 2021.

Domain Adaptation

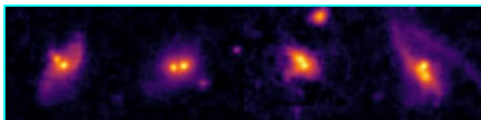


# Combining Datasets

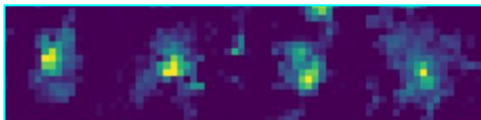
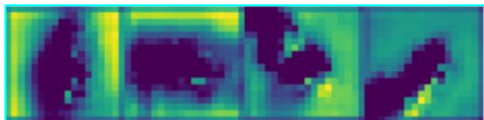
Source - Illustris



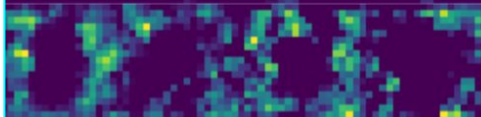
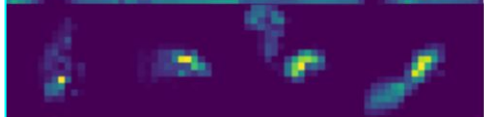
Target - SDSS observations



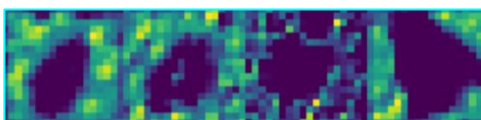
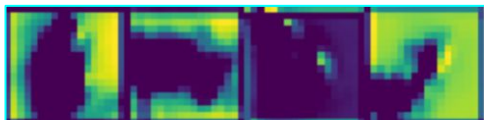
M



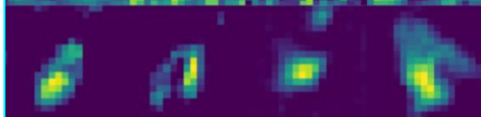
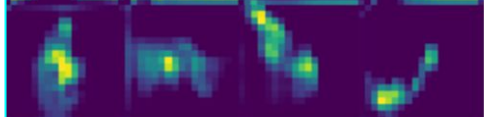
NM



M

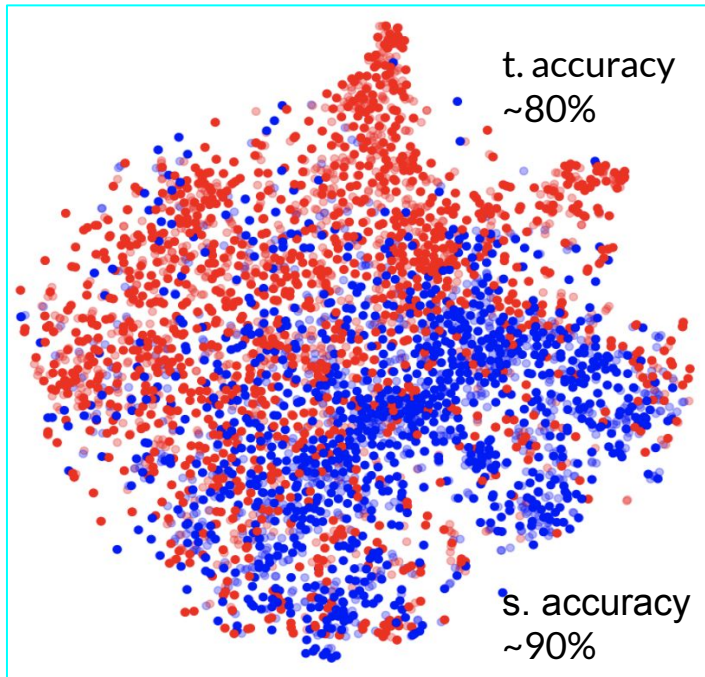


NM



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

Up to 30% increase!



# Talk Outline



**Introduction**

**How is AI used in Astrophysics?**

**Case Studies**

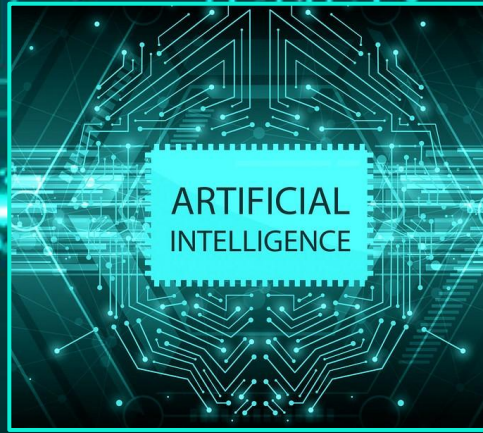
**Challenges and Opportunities**

**Conclusion**





+

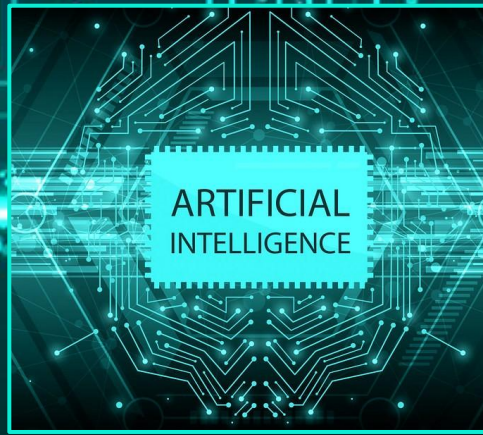


**Complex models  
based on data**





+



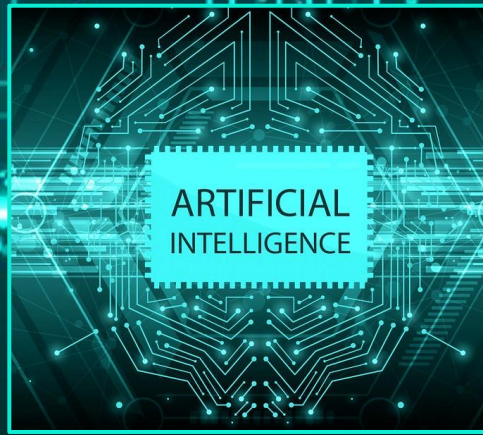
**Complex models  
based on data**

### 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

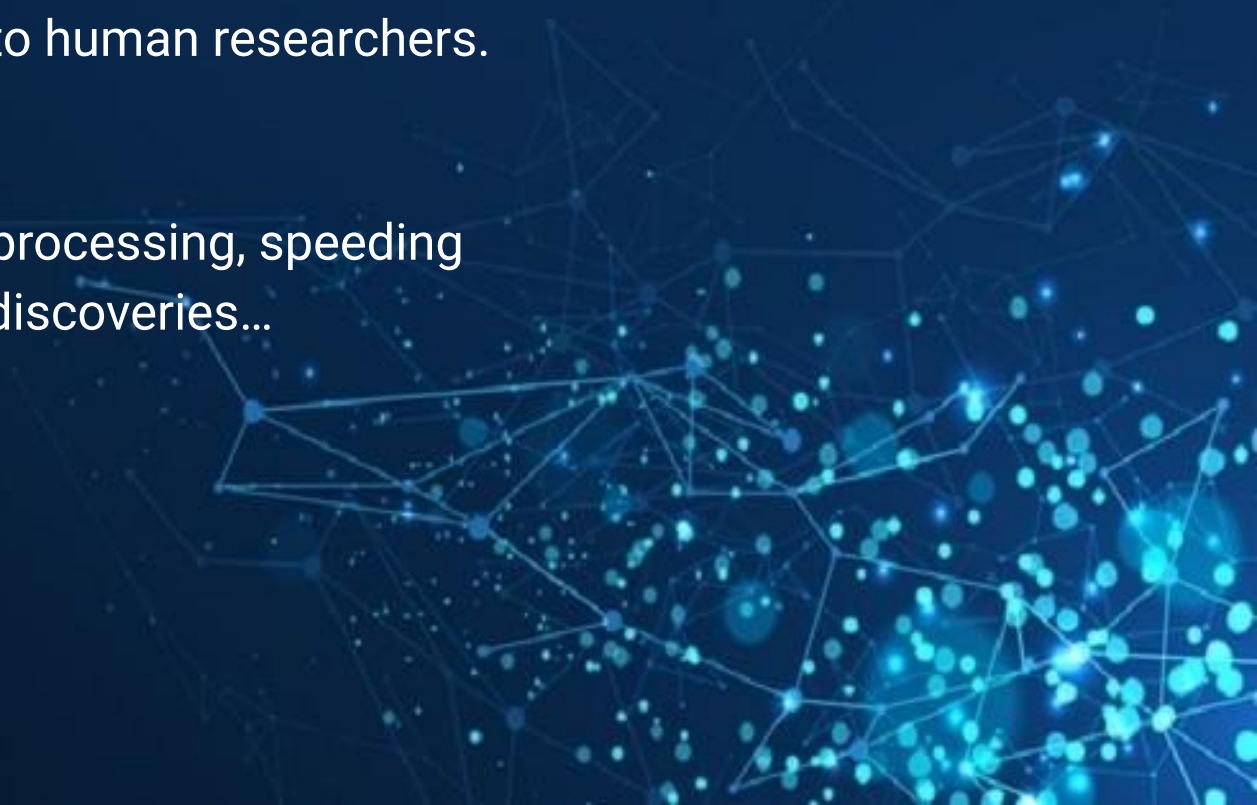
Conclusion



- 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