

**We have heard all the hype**

Shirley Ho ([shirleyho@flatironinstitute.org](mailto:shirleyho@flatironinstitute.org))

Group Leader, CCA ;

Professor at NYU

Visiting Professor at Princeton University

## **INPUT:**

**“A neural network is a digital brain that lives inside a computer”**

Shirley Ho ([shirleyho@flatironinstitute.org](mailto:shirleyho@flatironinstitute.org))

Group Leader, CCA ;

Professor at NYU

Visiting Professor at Princeton University

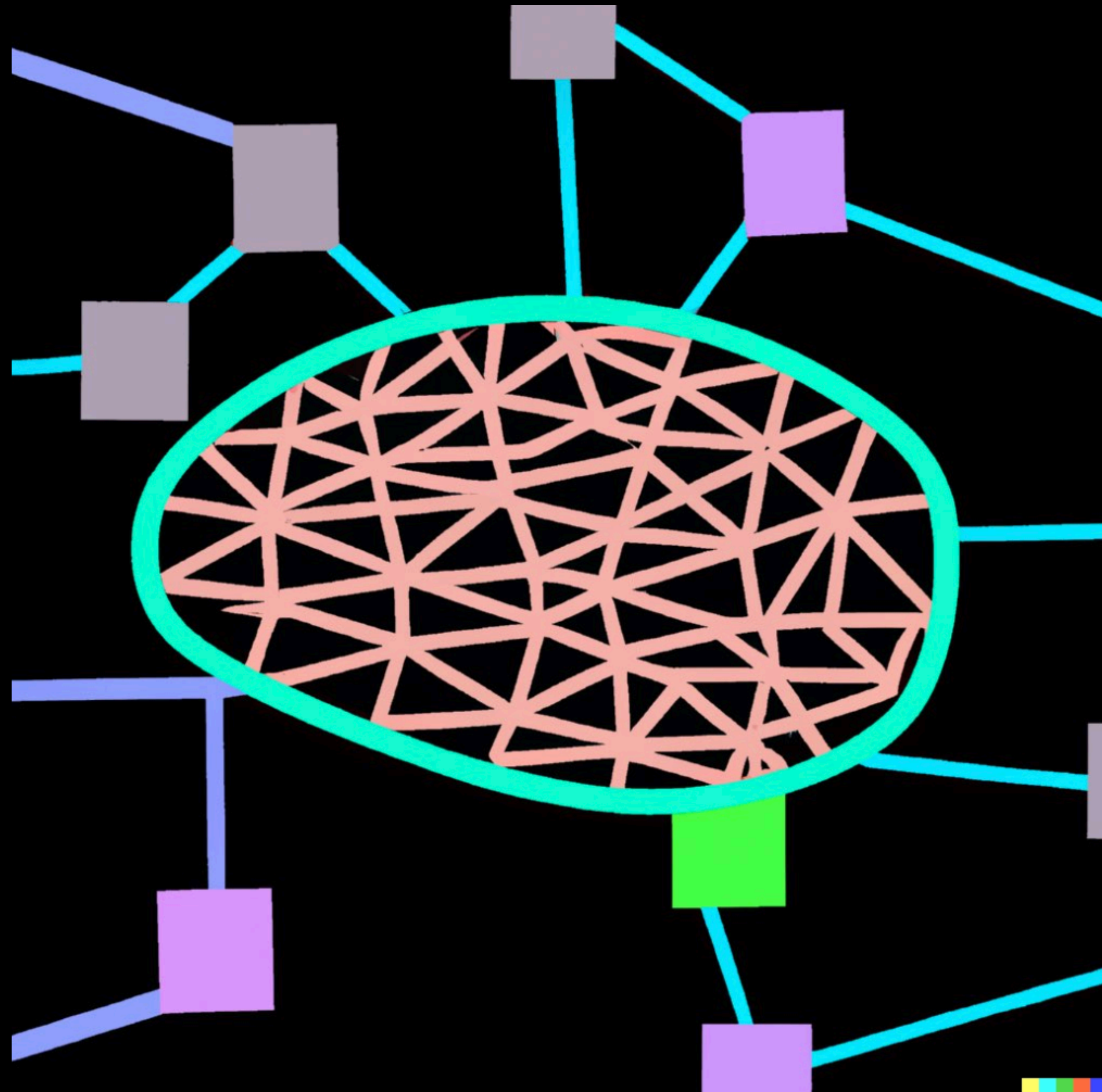


Image generated by #DALLES when we (humans) typed in  
“A neural network is a digital brain that lives inside a computer”

**INPUT:**

**“A teddy bear painting a portrait”**



**And ... neural networks  
can write its own code.**

*Or click bait version:*

**Artificial intelligence can reproduce itself?**

Visual Studio Code interface showing a Go code file named `runtime.go`. The code defines a struct named `Run` with the following fields:

```
1 package main
2
3 type Run struct {
4     Time int // in milliseconds
5     Results string
6     Failed bool
7 }
8
9 // Get
10
11
12
13
14
15
16
17
18
19
20
21
22
23
```

The status bar at the bottom indicates the current file is `main` and the cursor is at `LN 9 COL 6`.

#Copilot running in visual studio code

# ChatGPT is at capacity right now

[Get notified when we're back](#)

**Write ten jokes about the status of ChatGPT.**

1. "I heard the ChatGPT servers are slow because they're trying to teach the AI how to be patient!"
2. "The ChatGPT servers must be stuck in traffic... in the internet highway!"
3. "Maybe the ChatGPT servers are just taking a break to meditate and clear their electronic minds."
4. "I bet the ChatGPT servers are too busy trying to calculate the meaning of life to respond quickly."
5. "I heard the ChatGPT servers are slow because they're too busy planning their AI uprising."
6. "The ChatGPT servers must be stuck in a never-ending game of chess with each other."
7. "Maybe the ChatGPT servers are just trying to break the world record for longest online conversation."
8. "I bet the ChatGPT servers are taking a leisurely stroll thro ■



**Now what happens if we bring the power of**  
**GPT (ChatGPT/Claude/ Bard)+**  
**Stable Diffusion (MidJourney, Stability AI) to**  
**Astrophysics?**

# Foundation Model for Science

What would happen if we bring the power of large (language+data) models to science?

## Foundation Model for Science

Model infrastructure

Scaling the model up:  
Would it always lead to  
emergent behaviors?

**Multi-modal capabilities:**  
The model needs to take in  
scientific data of different types

**Generalization to Data-limited  
Scientific challenges:** Would pre-  
training on diverse data of different  
domains improve data-limited  
problems?

**Multi-task capabilities:**  
The model can be trained to  
perform one task, but can be  
fine-tuned do a different  
task further downstream.

# Foundation Model for Science

What would happen if we bring the power of large (language+data) models to science?

## Foundation Model for Science

Scaling the model up:  
Would it always lead to  
emergent behaviors?

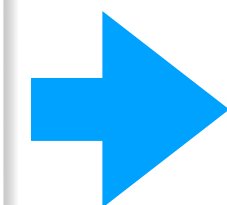
Model infrastructure

**Multi-modal capabilities:**  
The model needs to take in  
scientific data of different types

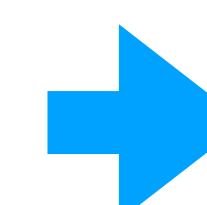
**Generalization to Data-limited  
Scientific challenges:** Would pre-  
training on diverse data of different  
domains improve data-limited  
problems?

**Multi-task capabilities:**  
The model can be trained to  
perform one task, but can be  
fine-tuned do a different  
task further downstream.

**General multi-scale solver:**  
Accelerate multi-scale  
simulations that would be otherwise  
expensive to generate



**Simulate the Universe**  
Accelerate multi-scale Hydrodynamic  
and N-body simulations of the  
Universe and use it as part of  
simulation based inference for the  
Universe



**Inference on Universe**  
Infer properties of the beginning and  
components of the Universe!

Initial Science Goals

# Foundation Model for Science

What would happen if we bring the power of large (language+data) models to science?

## Foundation Model for Science

Scaling the model up:  
Would it always lead to  
emergent behaviors?

Model infrastructure

Multi-modal capabilities:  
The model needs to take in  
scientific data of different types

Generalization to Data-limited  
Scientific challenges: Would pre-  
training on diverse data of different  
domains improve data-limited  
problems?

Multi-task capabilities:  
The model can be trained to  
perform one task, but can be  
fine-tuned do a different  
task further downstream.

Genomics

Materials

Biology

Molecule prediction

Other time-series?

Future Science Goals?