Deep Learning for Science
Opportunities & Challenges

Maxwell Cai, Ph.D. (SURFsara)

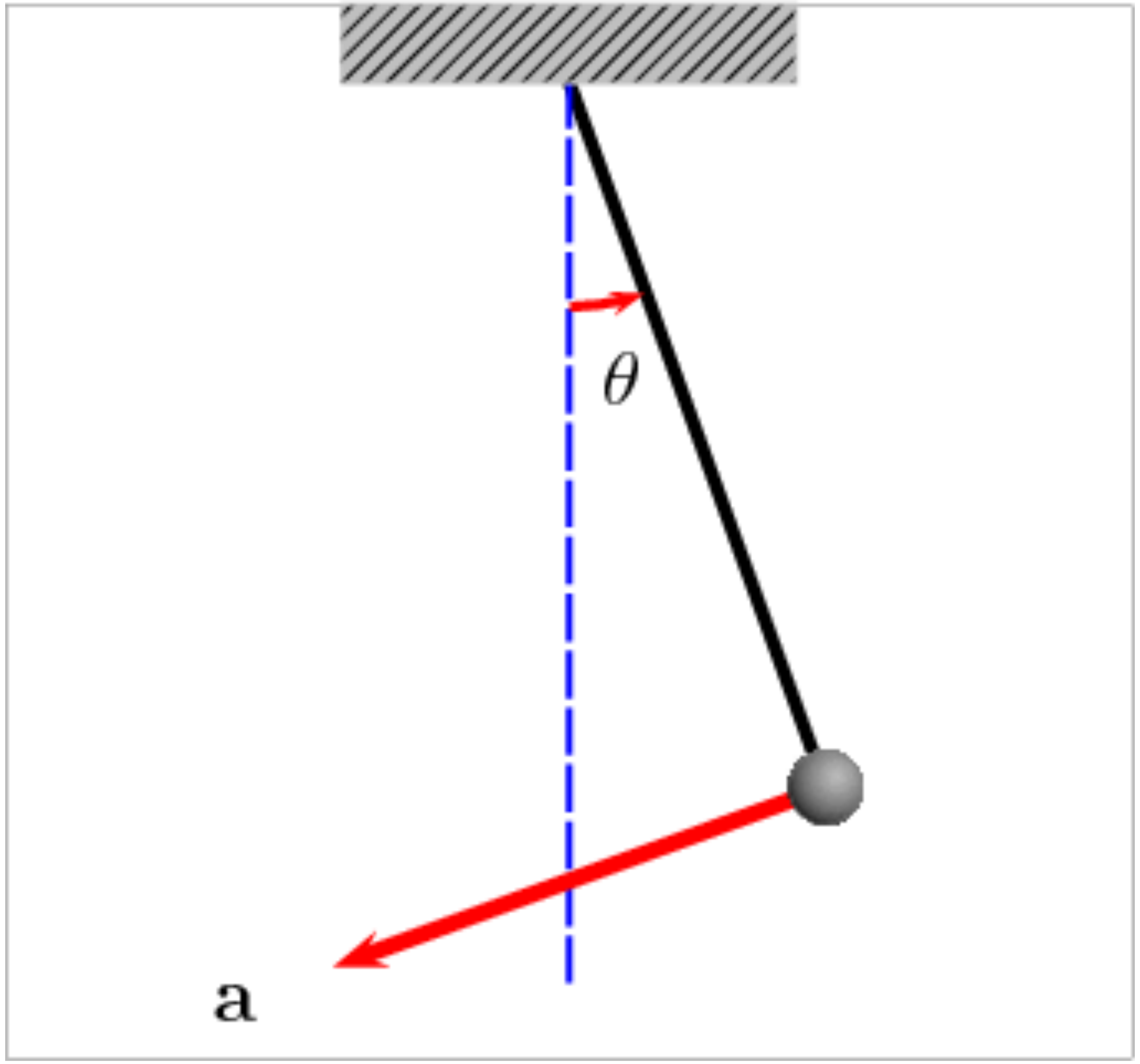
Deep Learning for Science
Opportunities & Challenges

Maxwell Cai, Ph.D. (SURFsara)

Analytical models

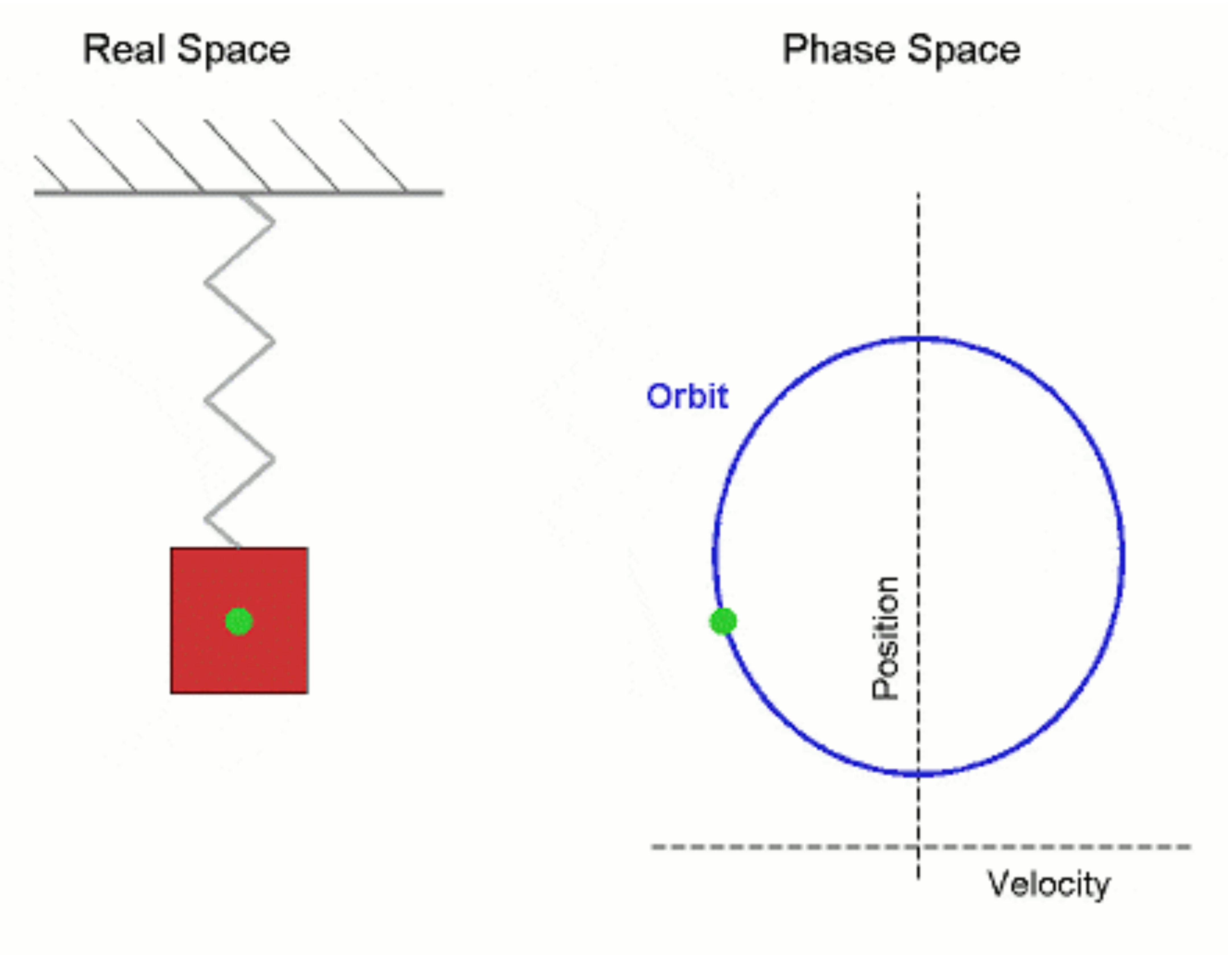
Elegant & Abstract

Analytical model: pendulum & harmonic oscillator

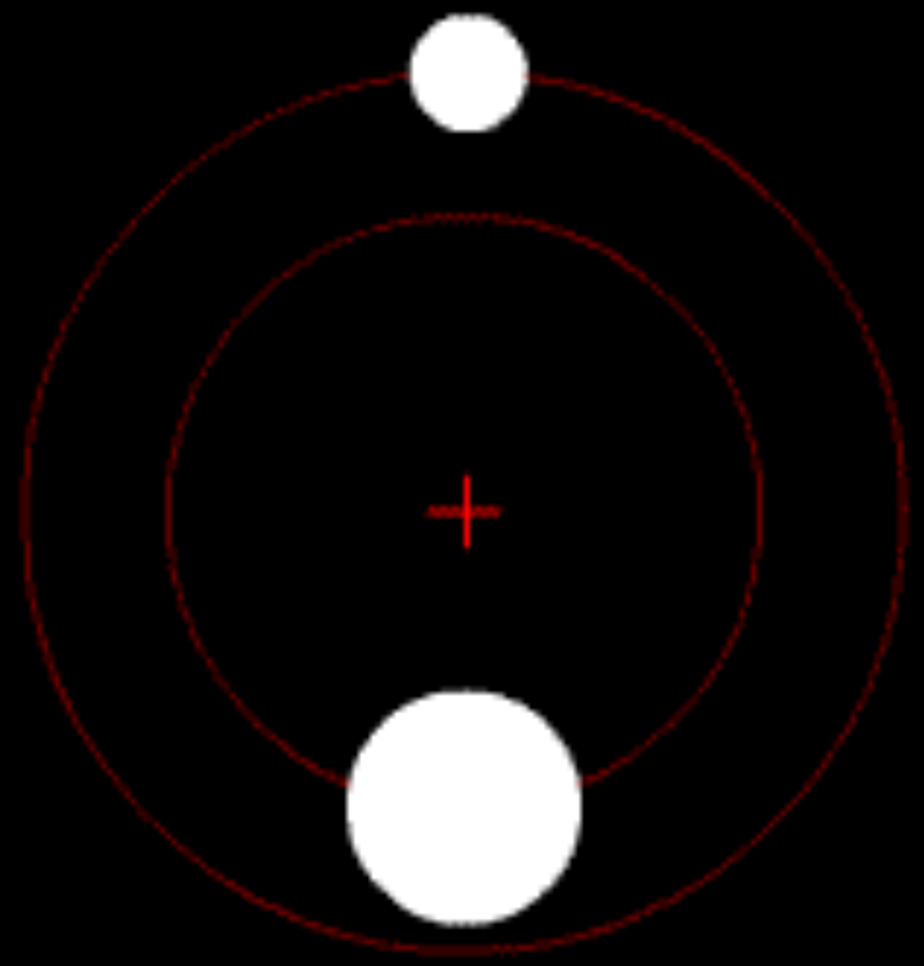


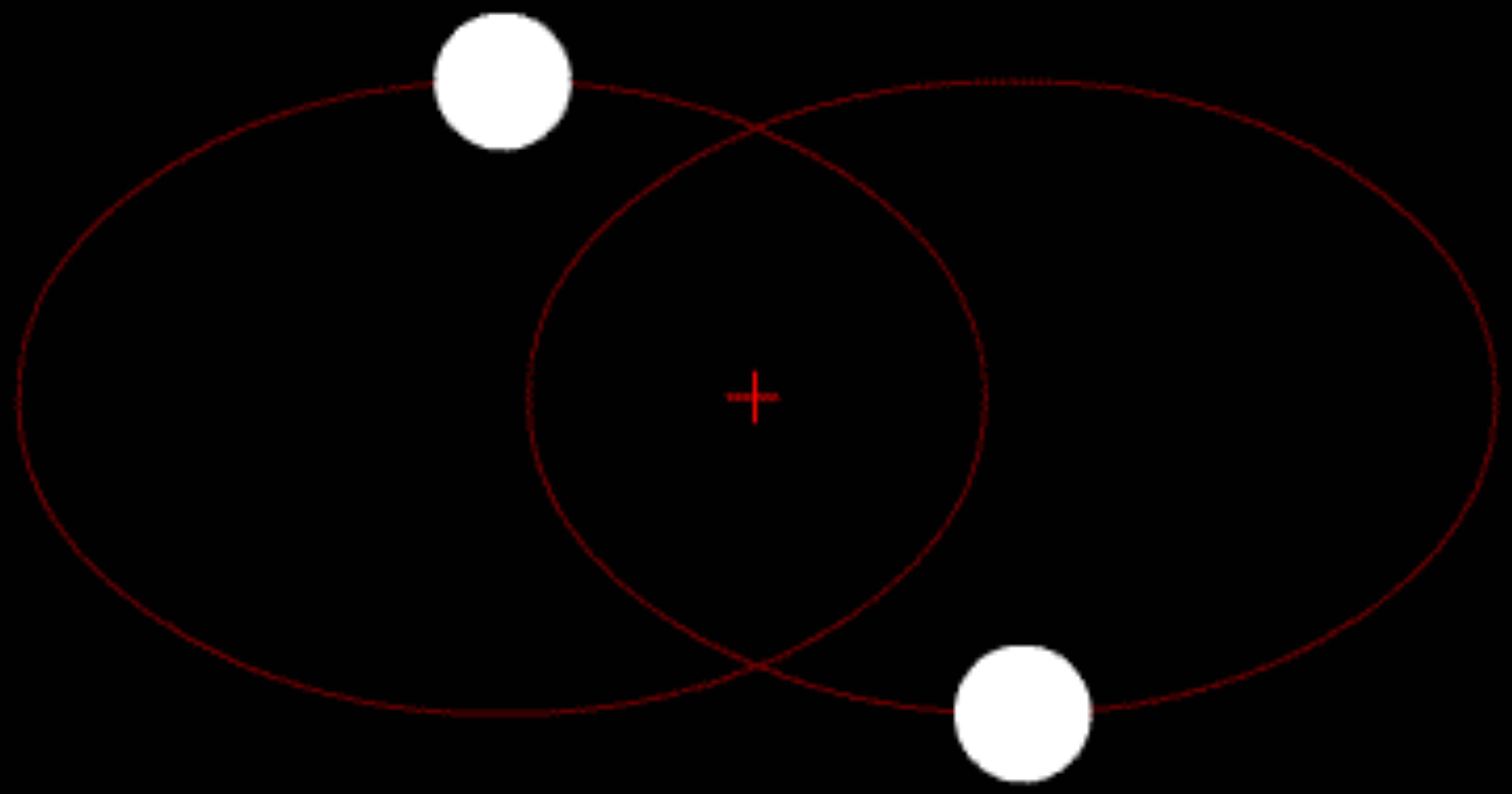
$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\sin\theta = 0$$

$$T_0 \approx 2\pi\sqrt{\frac{l}{g}}$$



Analytical model: two-body problem





$$\mathbf{F}_{12}(\mathbf{x}_1, \mathbf{x}_2) = m_1 \ddot{\mathbf{x}}_1 \qquad \mathbf{F}_{21}(\mathbf{x}_1, \mathbf{x}_2) = m_2 \ddot{\mathbf{x}}_2$$

$$\ddot{\mathbf{R}} \equiv \frac{m_1 \ddot{\mathbf{x}}_1 + m_2 \ddot{\mathbf{x}}_2}{m_1 + m_2} = 0 \qquad \mu = \frac{1}{\frac{1}{m_1} + \frac{1}{m_2}} = \frac{m_1 m_2}{m_1 + m_2}$$

$$\ddot{\mathbf{r}} = \ddot{\mathbf{x}}_1 - \ddot{\mathbf{x}}_2 = \left(\frac{\mathbf{F}_{12}}{m_1} - \frac{\mathbf{F}_{21}}{m_2} \right) = \left(\frac{1}{m_1} + \frac{1}{m_2} \right) \mathbf{F}_{12}$$

$$\mathbf{x}_1(t) = \mathbf{R}(t) + \frac{m_2}{m_1 + m_2} \mathbf{r}(t) \qquad \mathbf{L} = \mathbf{r} \times \mathbf{p} = \mathbf{r} \times \mu \frac{d\mathbf{r}}{dt}$$

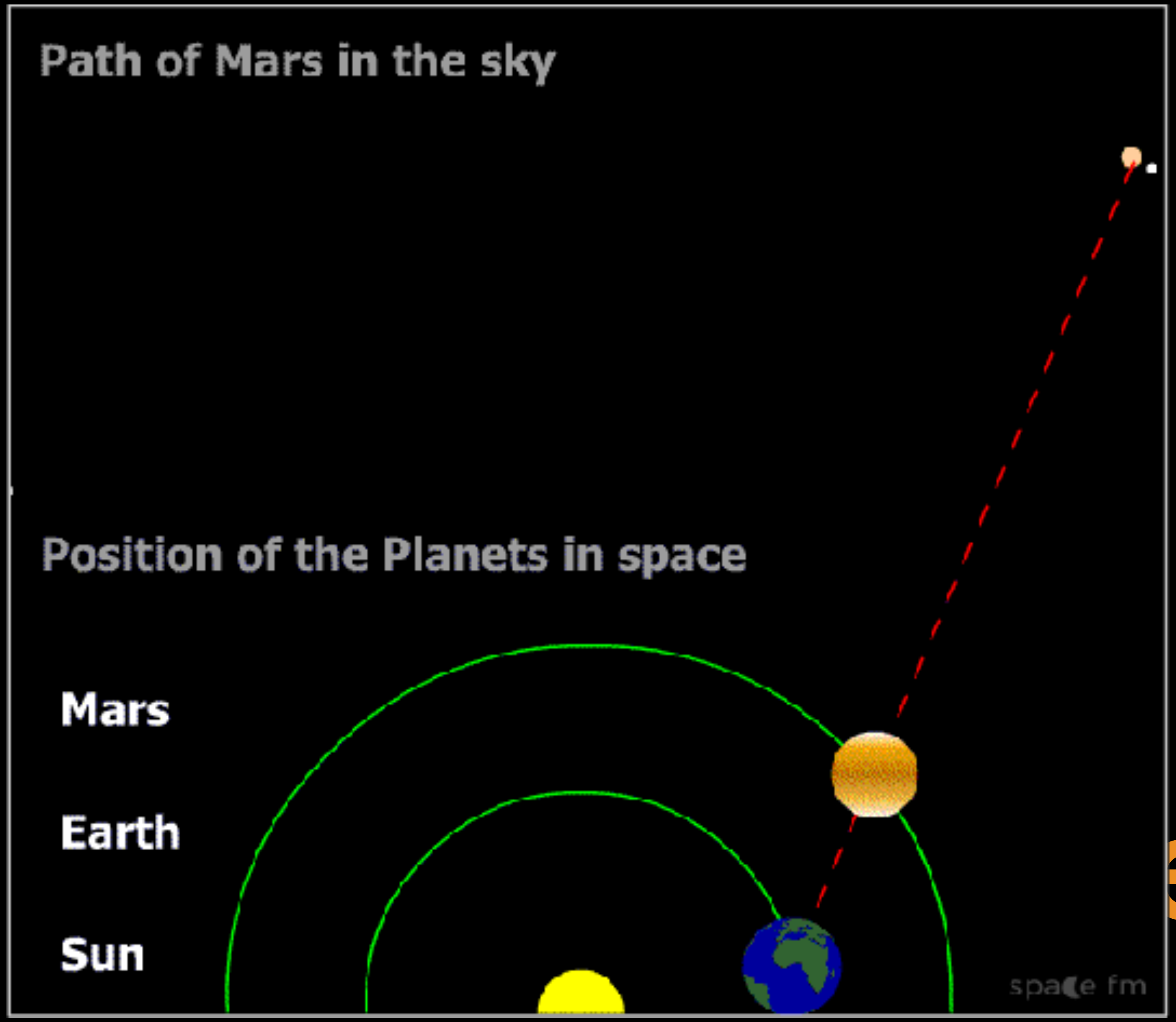
$$\mathbf{x}_2(t) = \mathbf{R}(t) - \frac{m_1}{m_1 + m_2} \mathbf{r}(t) \qquad \mathbf{N} = \frac{d\mathbf{L}}{dt} = \dot{\mathbf{r}} \times \mu \dot{\mathbf{r}} + \mathbf{r} \times \mu \ddot{\mathbf{r}}$$

Analytical models → Numerical models

Computers are very good at this!

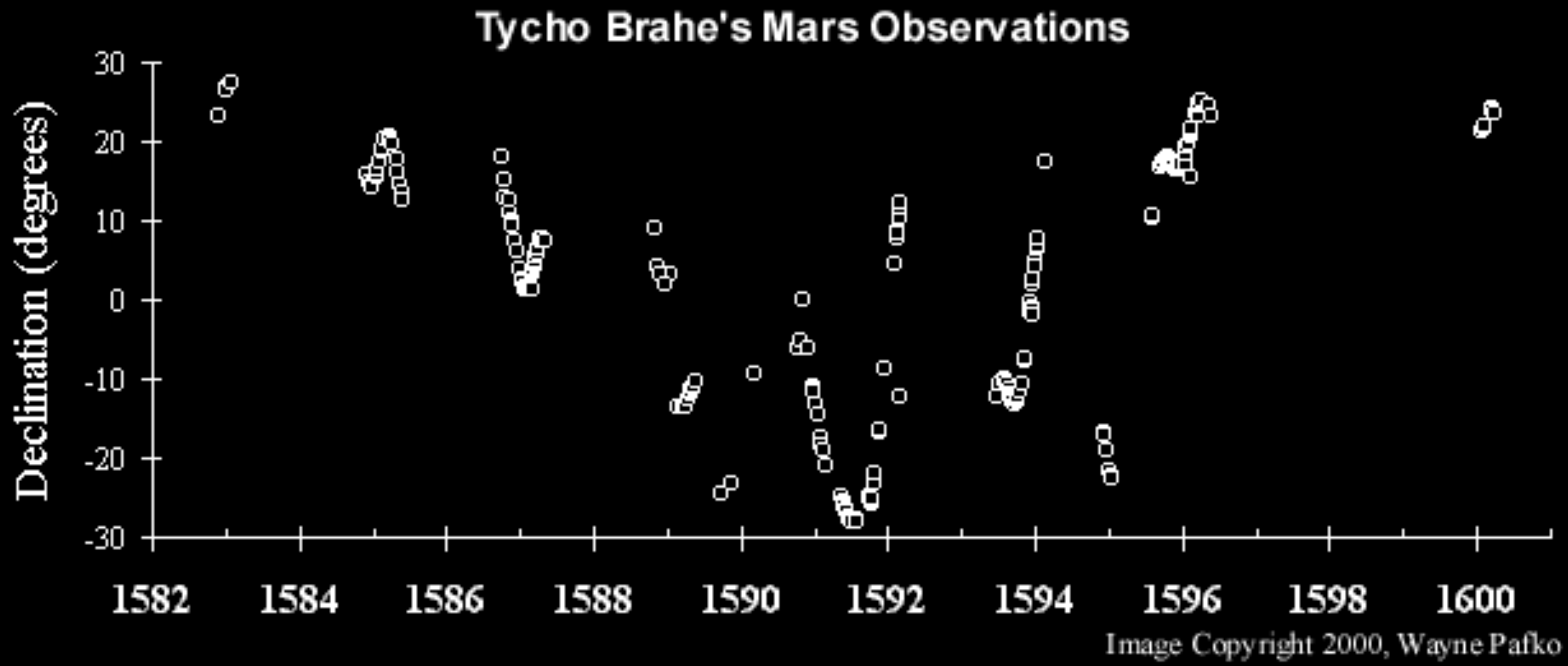
But sometimes we don’t know how to model…

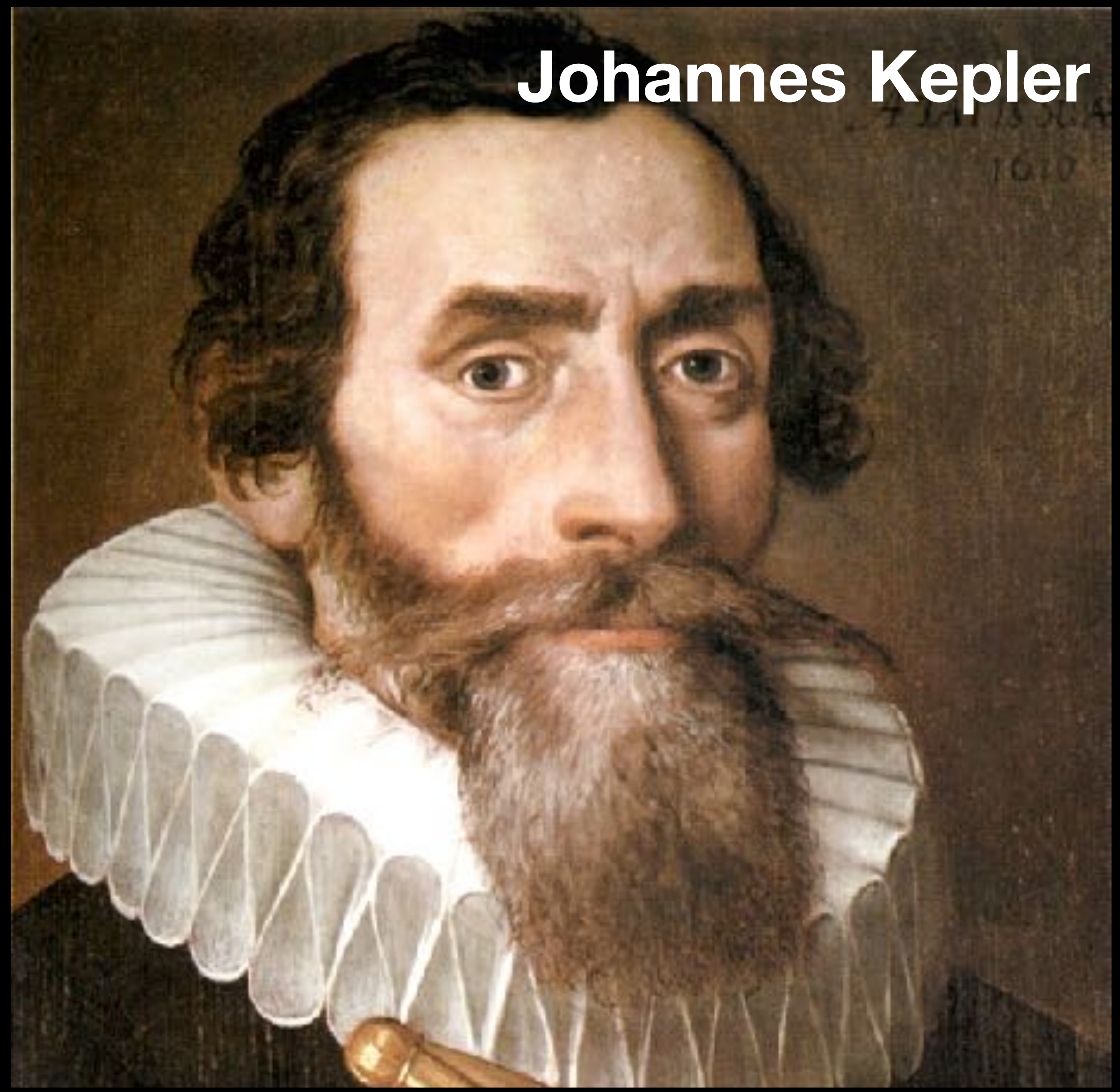
The world is complicated



$$\frac{a^3}{T^2} = \text{const} = \frac{G(M+m)}{4\pi^2}$$

e rise of empirical moc





The rise of empirical models

Analytical/numerical models	Empirical models
Driven by **direct knowledge**	Driven by **data**
Formulation required (expensive human efforts)	**Observation** required (can be automated)
Based on certain assumptions	No/little assumptions
Subject to the quality of the **assumptions**	Subject to the quality of the **data**
Seek to **emulate** the real-world system	Seek to **find patterns** in the data
Making predictions can be **expensive**	Making predictions is **straightforward**

Nature, I want to **understand** how you think and do, fully...

Analytical models

You wish!

Nature

Nature, I want to **see** how you think and do, more and more...

Empirical models

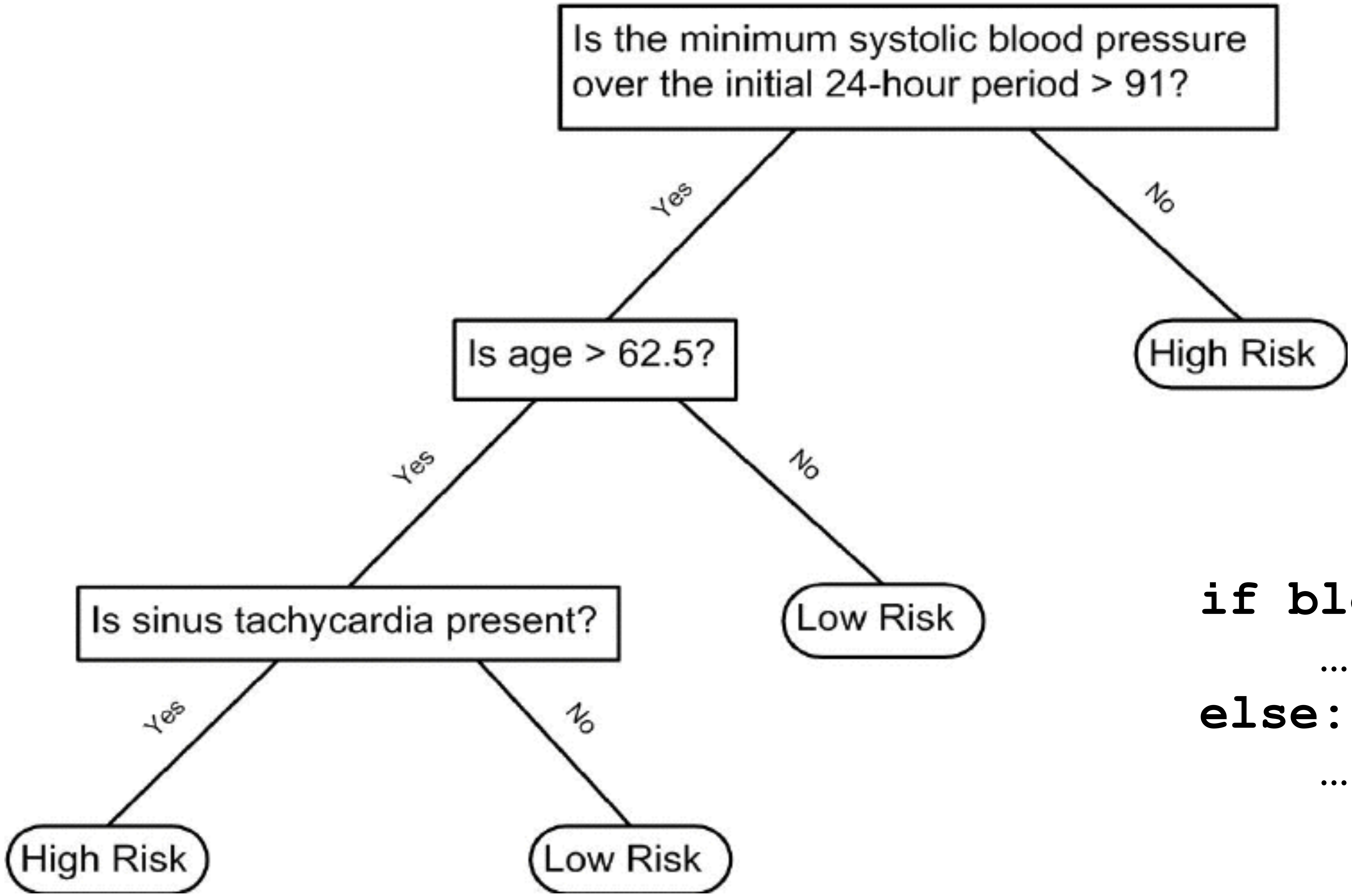
You may.

Nature

Complex problem, (huge amount of) complex data

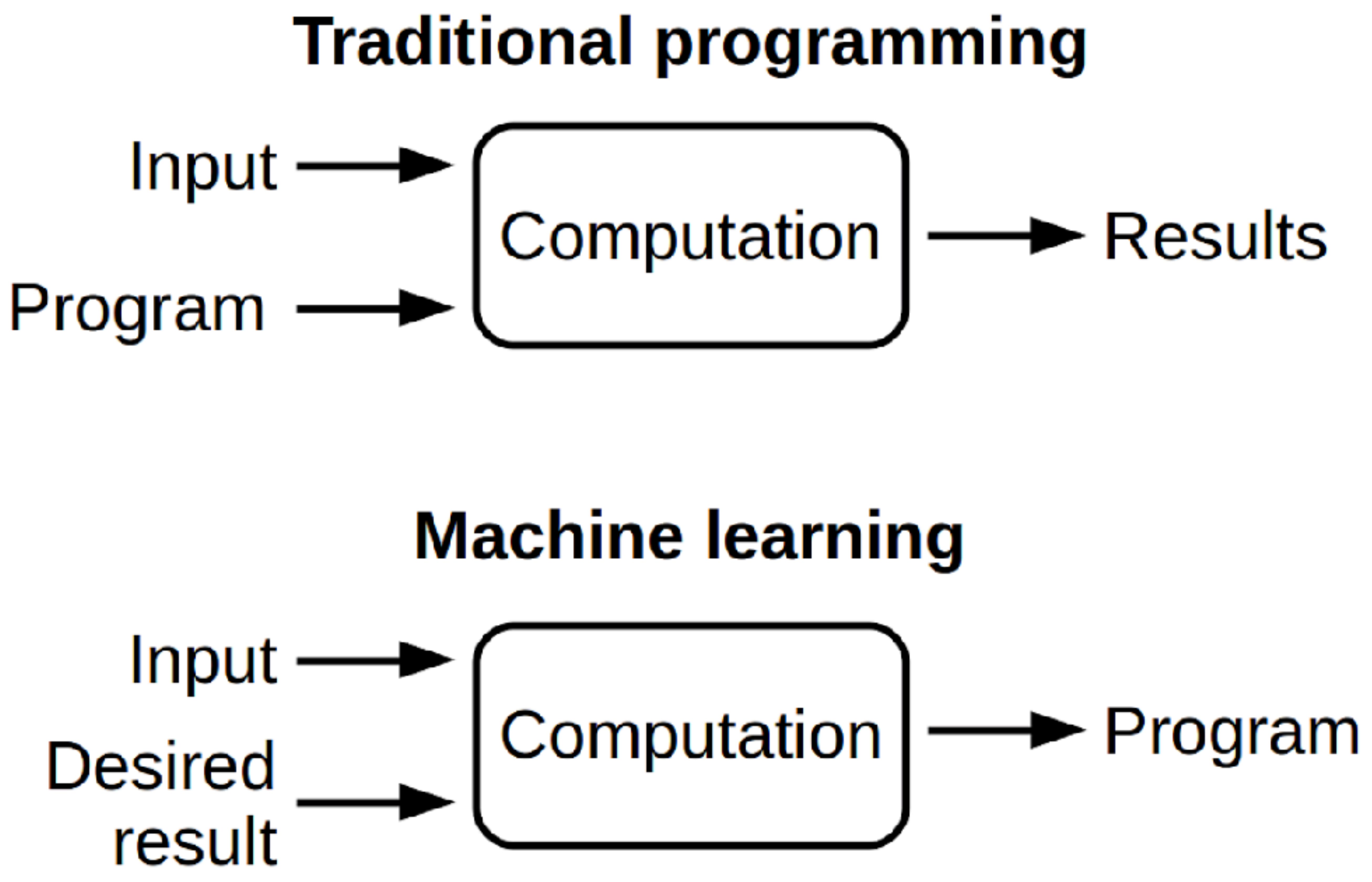
New programming paradigm is needed!

Decision Tree(s)




```
if blood_pressure > 91:
    …
else:
    …
```

Machine learning: a new programming paradigm



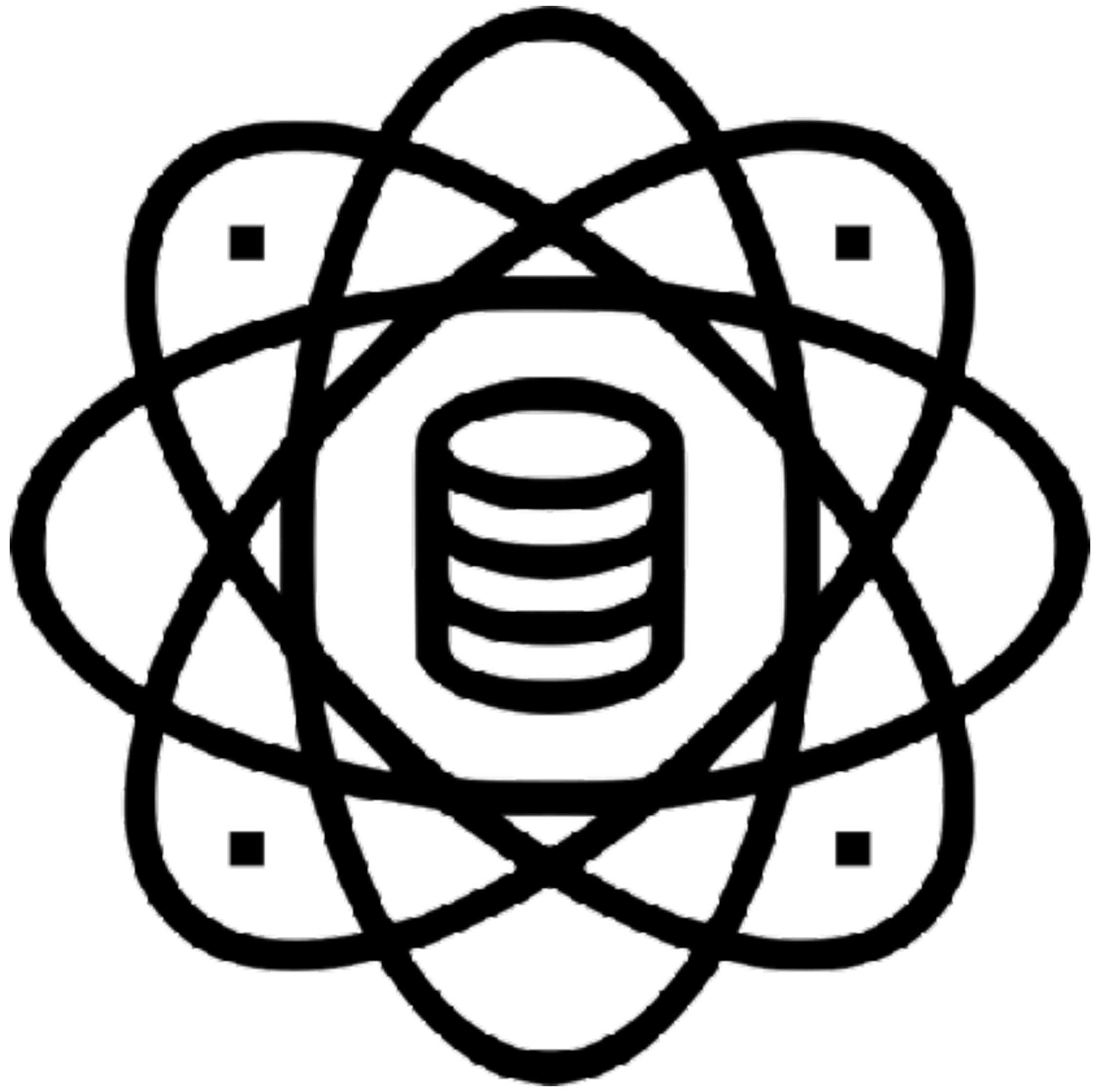
Knowledge base & rules
Expert systems
Human input

Features and result
'Decision' system
Limited human input

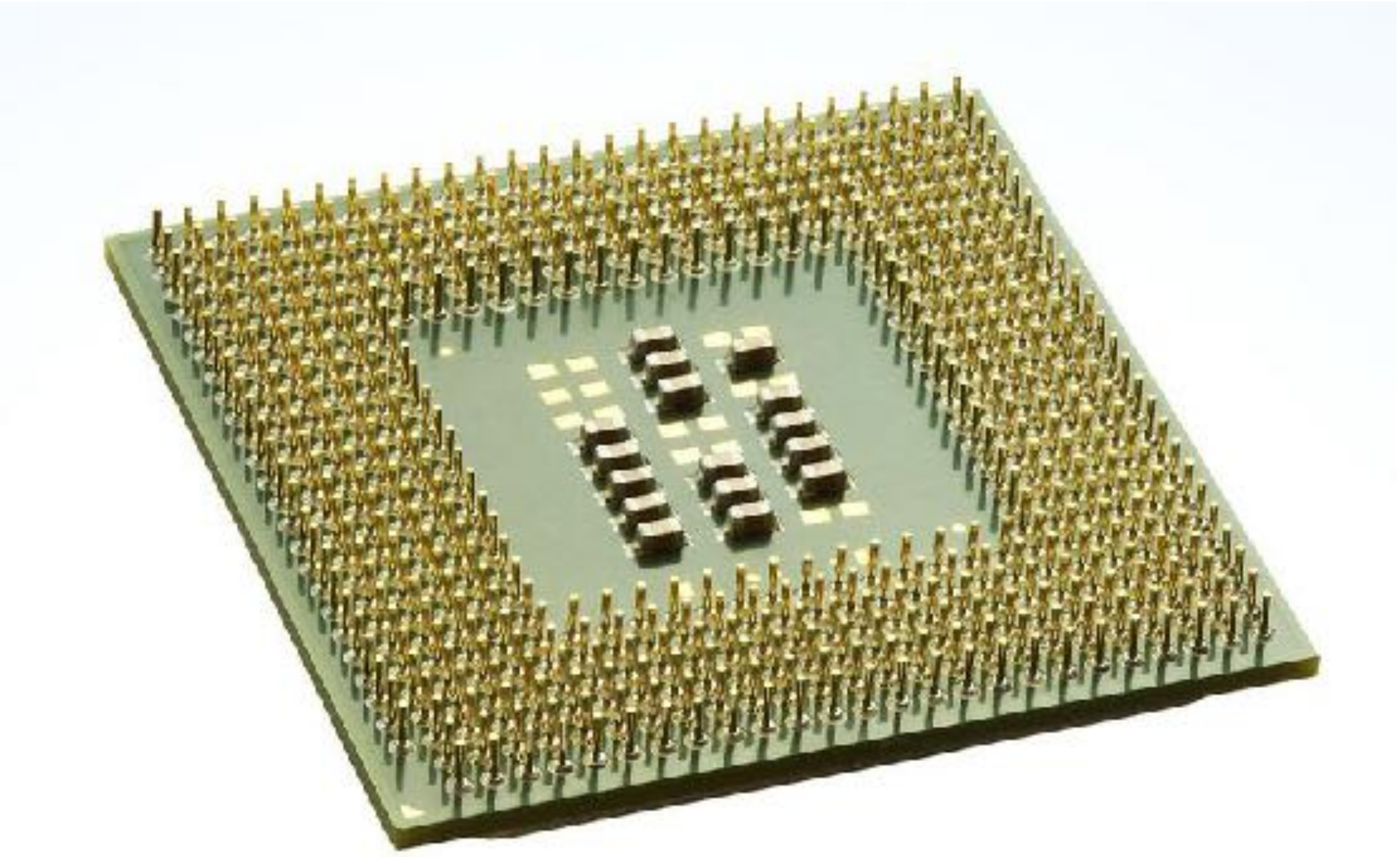
Source: Machine Learning for an Expert System to Predict Preterm Birth Risk , Woolery et al. (1994)

Three ingredients of AI

(Scientific) data



Computational Facilities



Algorithms

TensorFlow

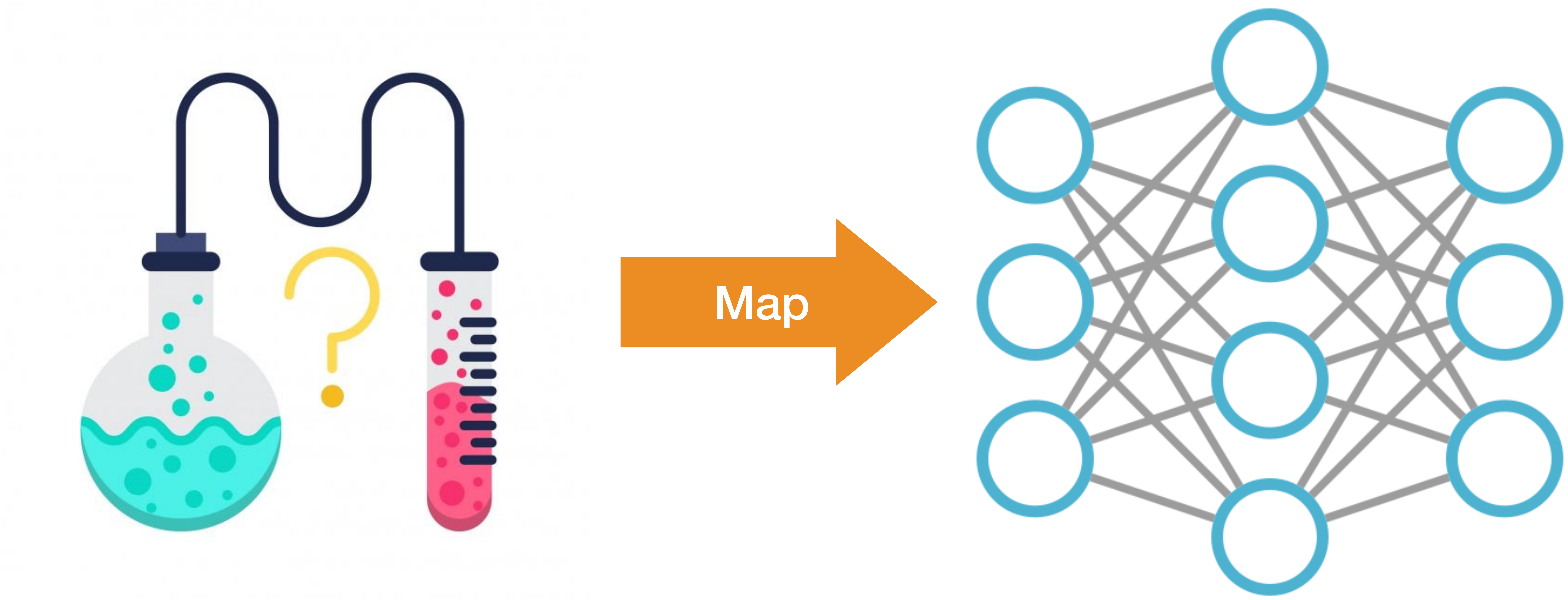
mxnet

PyTorch

Microsoft CNTK

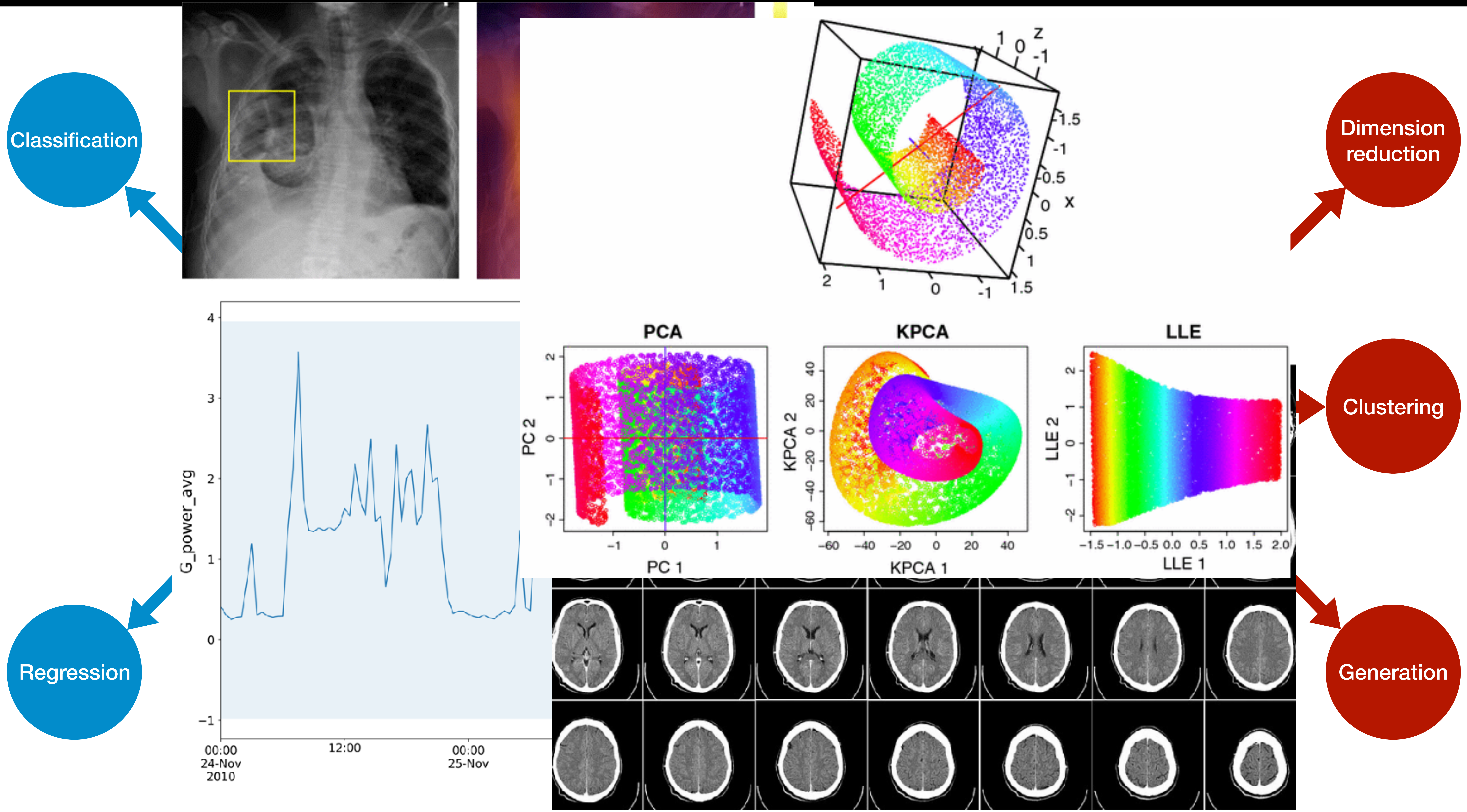
The AI opportunity

A new tool for scientific research

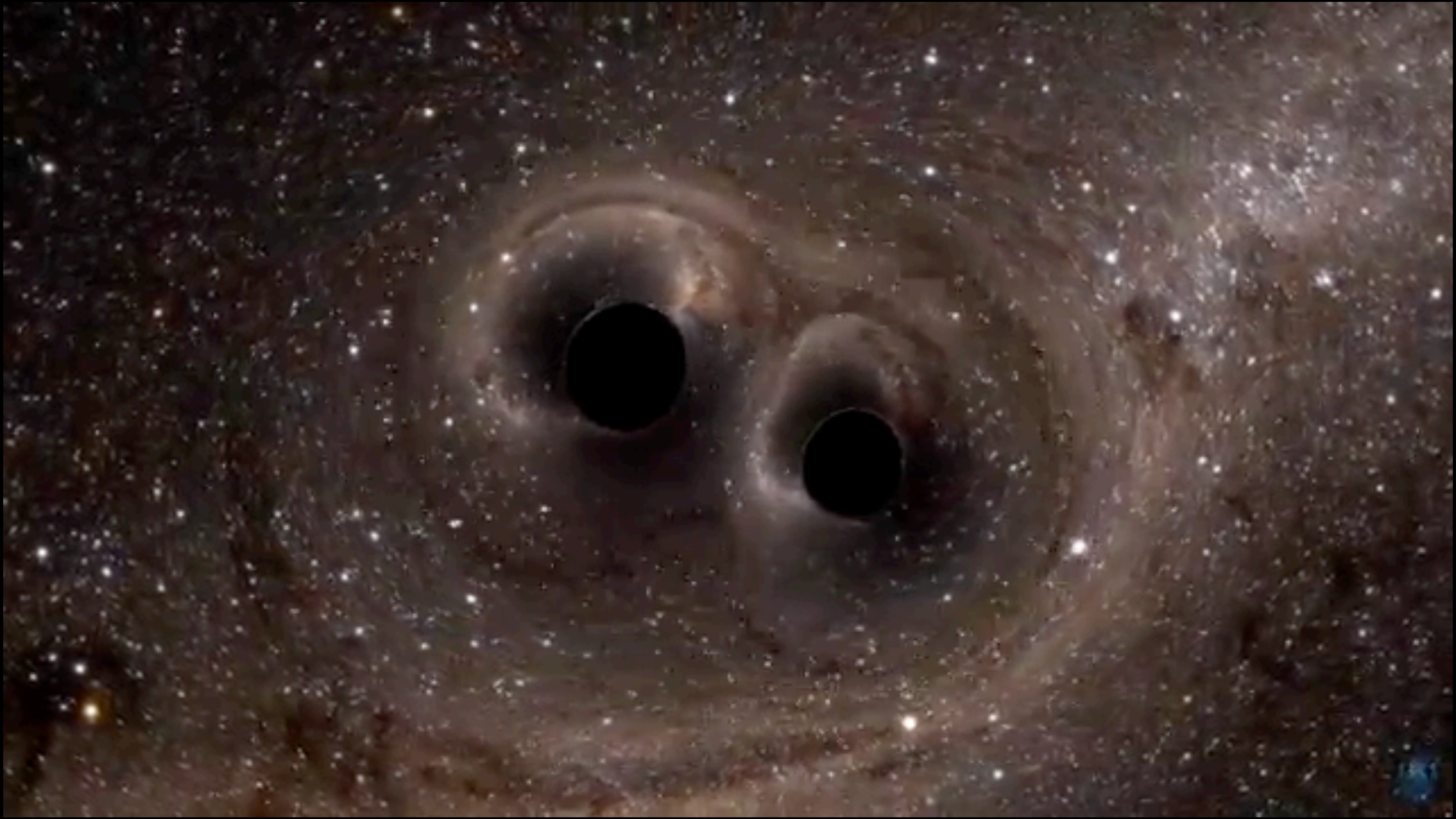


Scientific Problem **AI Problem**

Scientific Problem → AI Problem



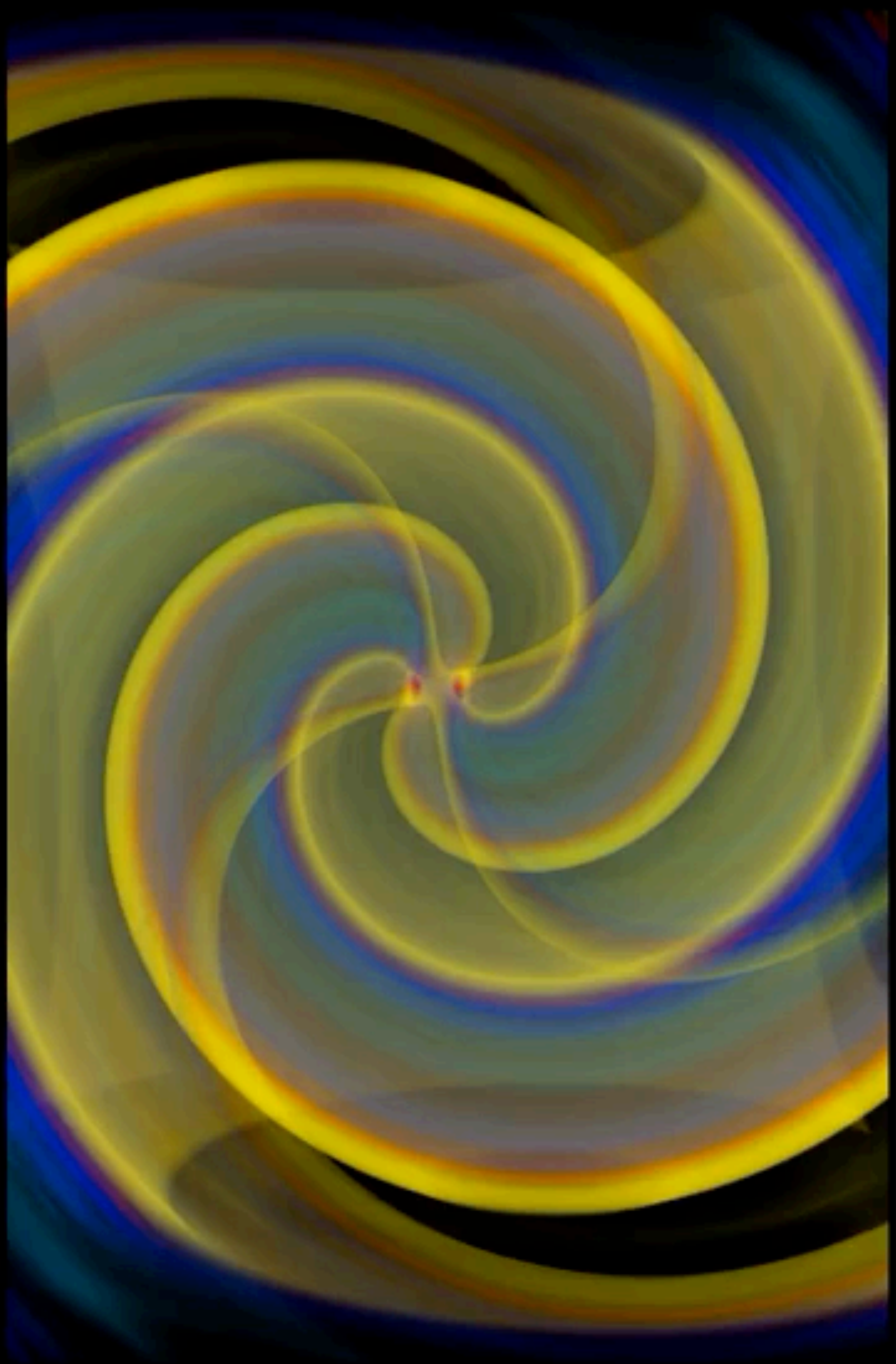
Black hole Mergers → Gravitational Waves

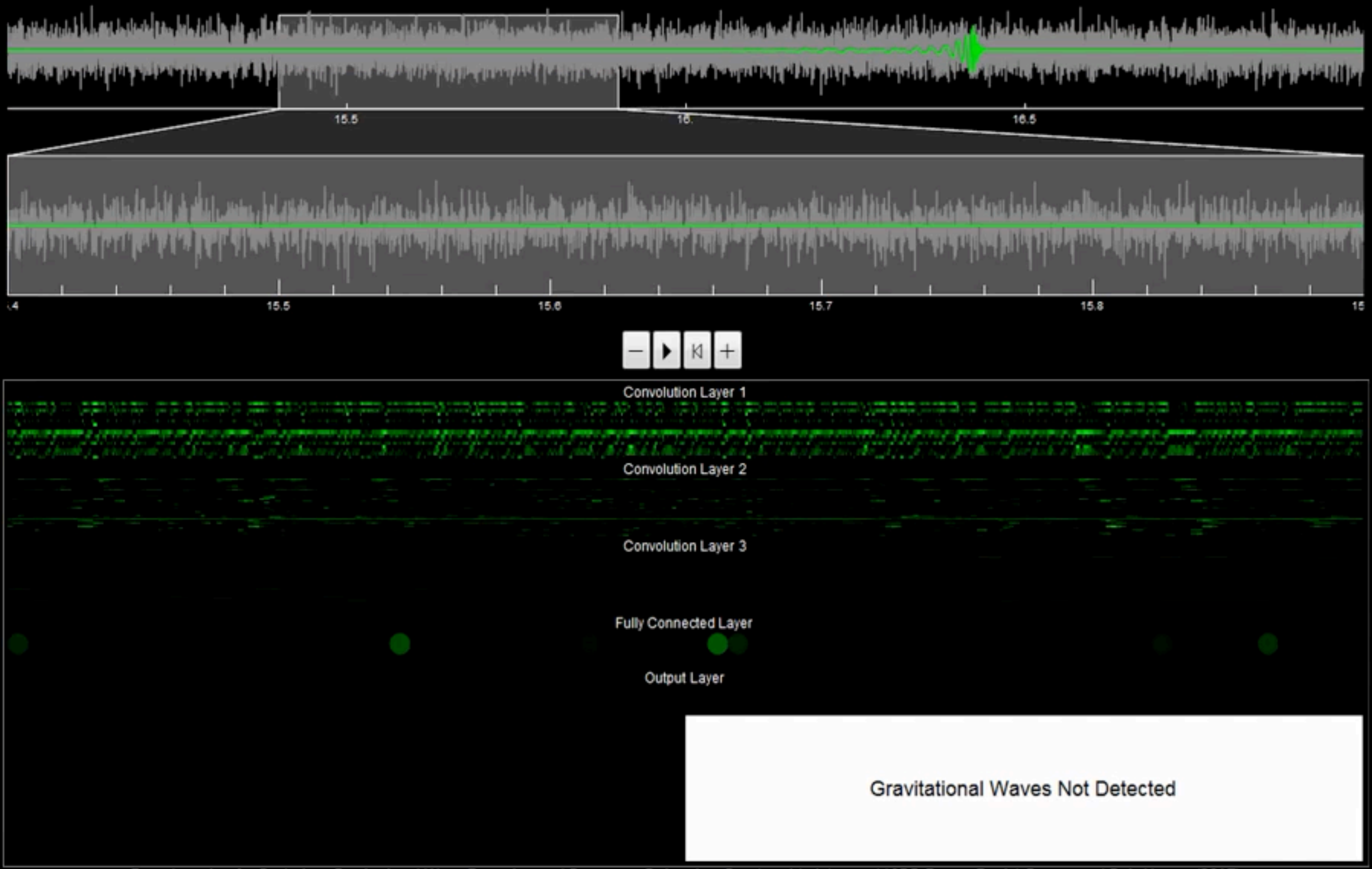


Credit: European Southern Observatory

Detecting Gravitational Waves in Real–Time with Deep Learning

Data from a LIGO Interferometer around the first event (GW150914)





Deep Learning for Real–time Gravitational Wave Detection and Parameter Estimation: Results with Advanced LIGO Data – Daniel George and E. A. Huerta (2017)

Let AI to learn the intuition

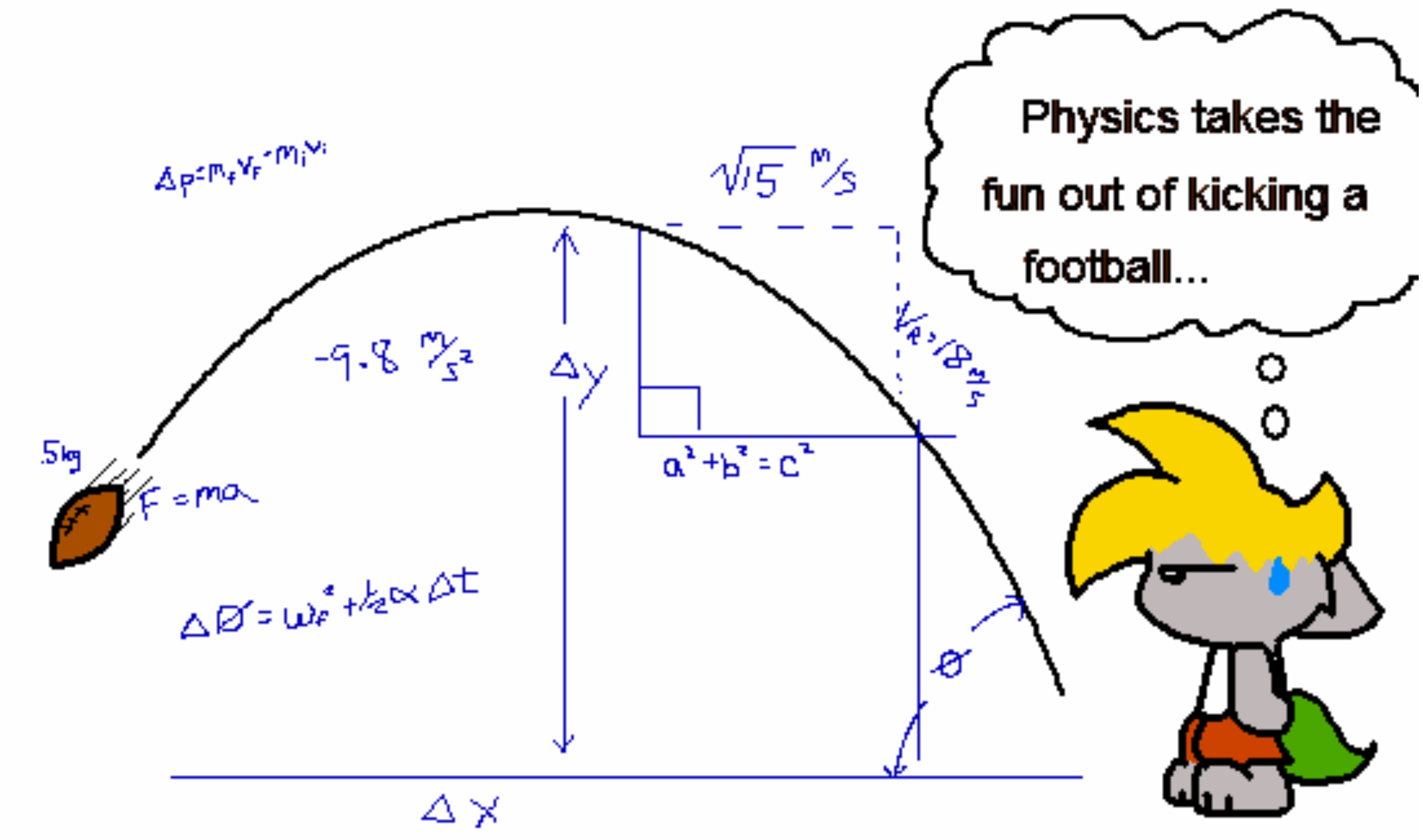
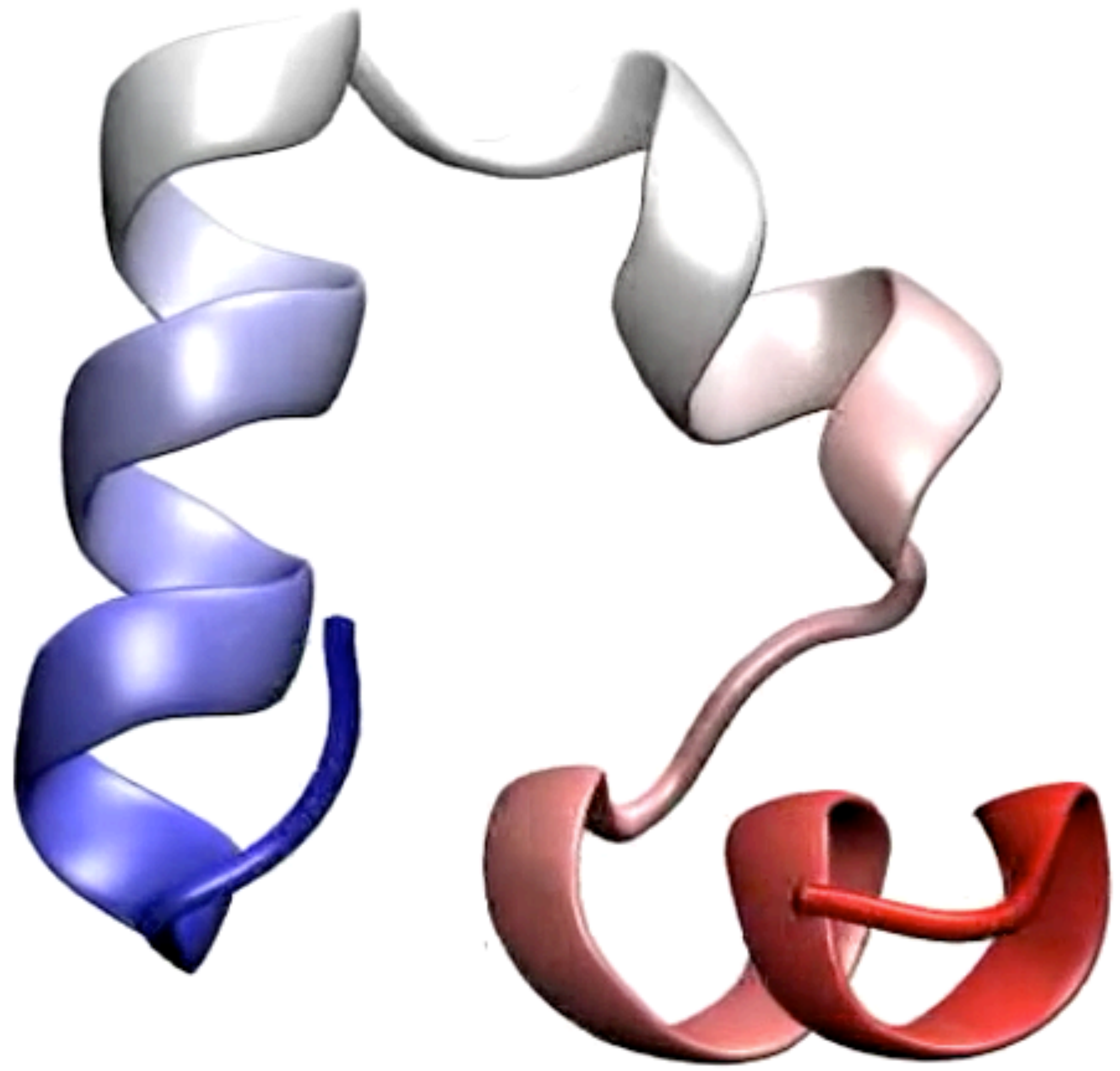
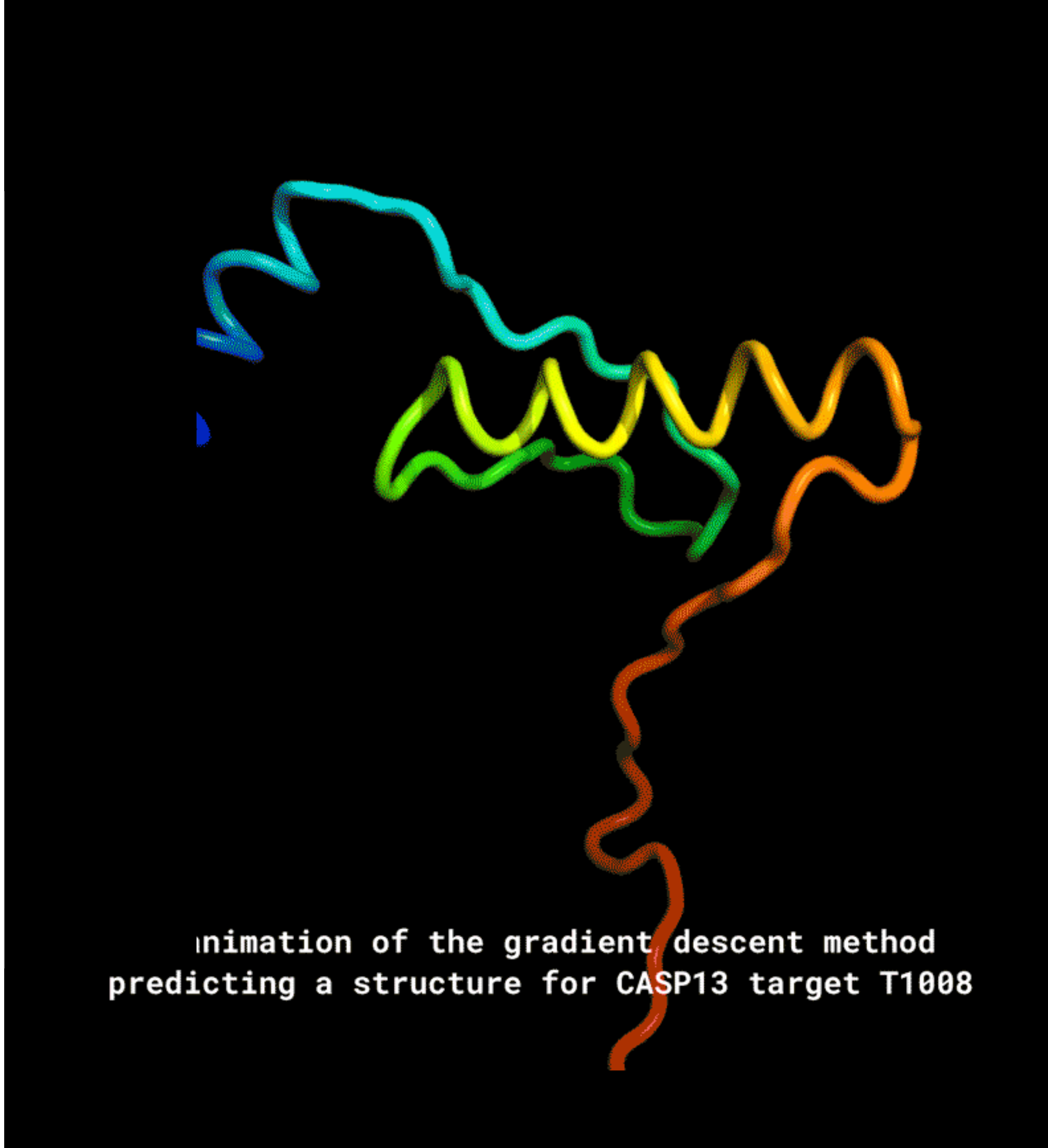


Image source: Deviant Art



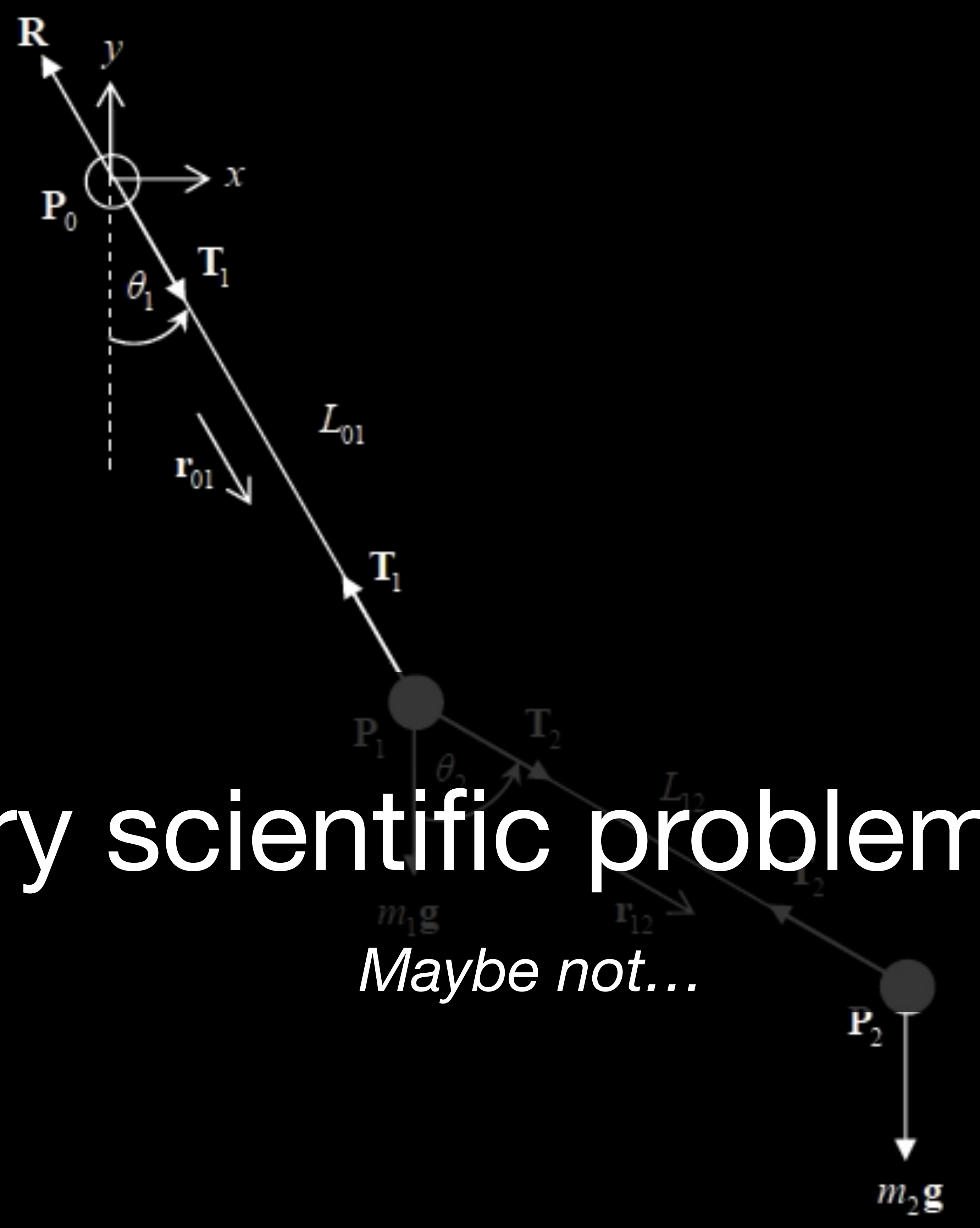

nimation of the gradient descent method
predicting a structure for CASP13 target T1008

The AI Challenges

Scientific research is non-trivial; using AI in scientific research is also non-trivial

Can we map every scientific problem to an AI problem?

Maybe not…



$$\dot{\theta}_1 = \omega_1$$
$$\dot{\theta}_2 = \omega_2$$

$$\dot{\omega}_1 = \frac{\frac{g}{L_{01}}\left(\sin\theta_2\cos(\theta_1-\theta_2) - \left(1+\frac{m_1}{m_2}\right)\sin\theta_1\right) - \omega_2^2\frac{L_{12}}{L_{01}}\sin(\theta_1-\theta_2) - \omega_1^2\sin(\theta_1-\theta_2)\cos(\theta_1-\theta_2)}{1+\frac{m_1}{m_2}-\cos^2(\theta_1-\theta_2)}$$

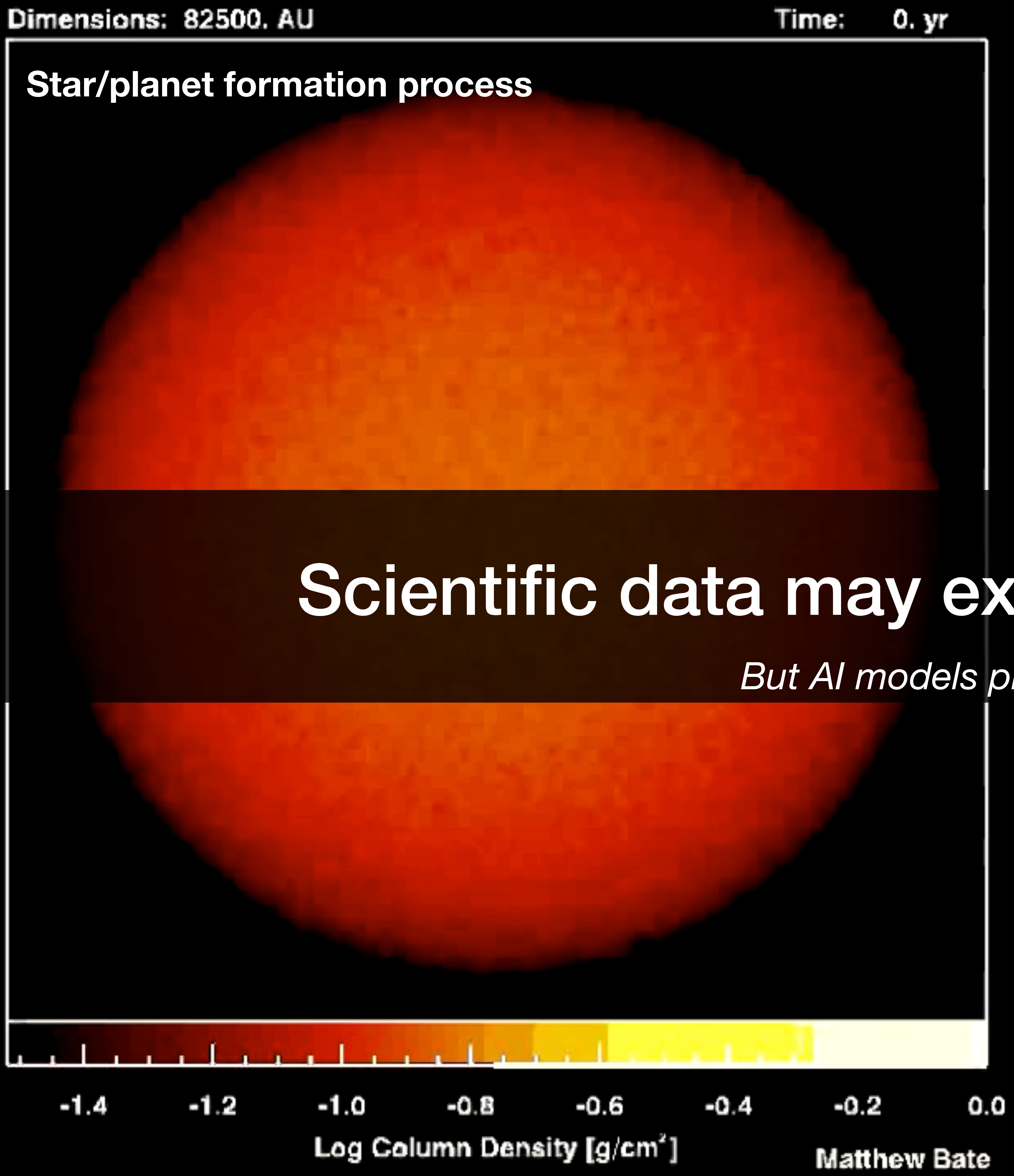
$$\dot{\omega}_2 = -\dot{\omega}_1\frac{L_{01}}{L_{12}}\cos(\theta_1-\theta_2)+\omega_1^2\frac{L_{01}}{L_{12}}\sin(\theta_1-\theta_2)-\frac{g}{L_{12}}\sin\theta_2$$

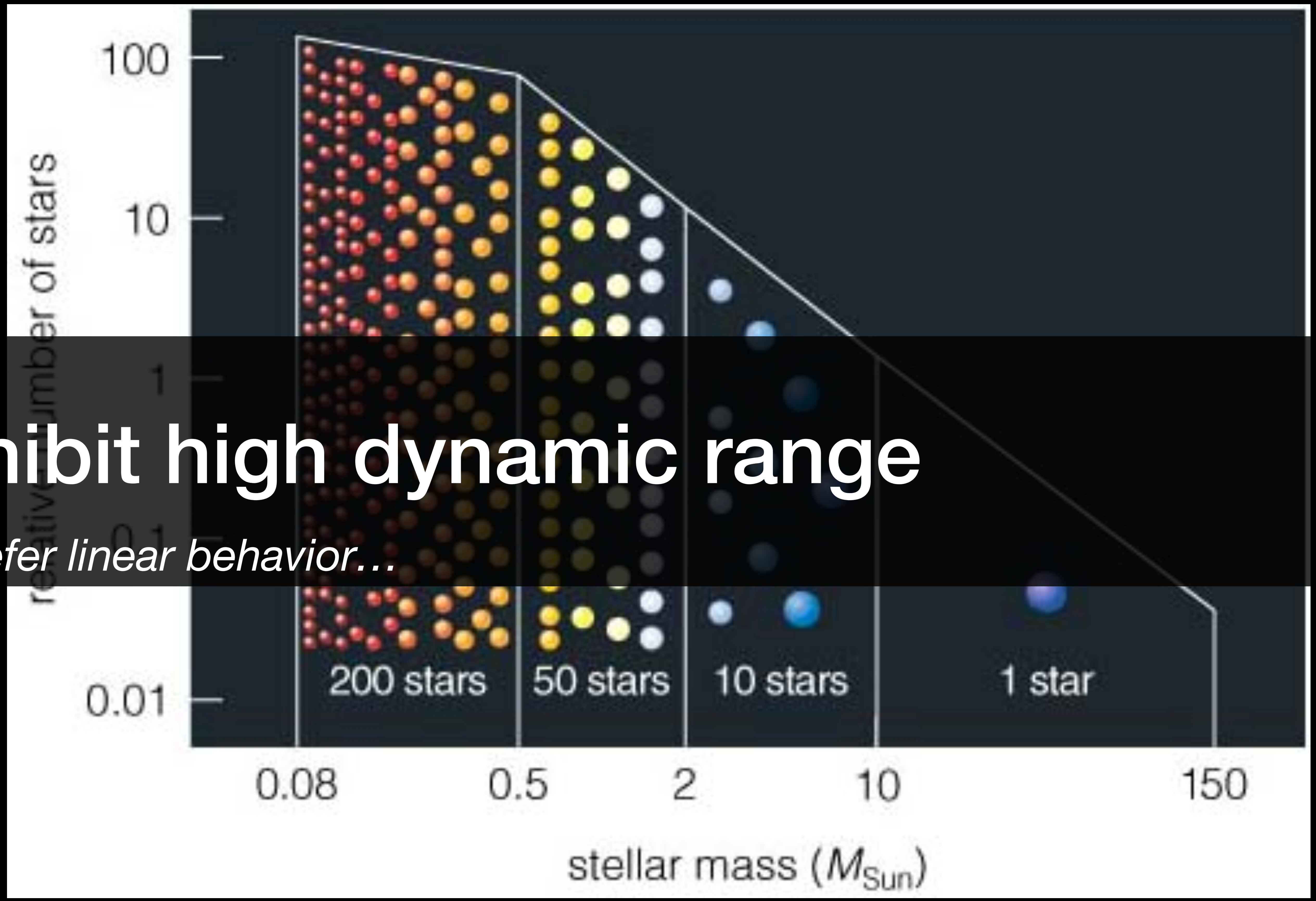
Mapping

Sometimes, outliners are more interesting!

But outliners are rare…

Data complexity





Initial stellar mass

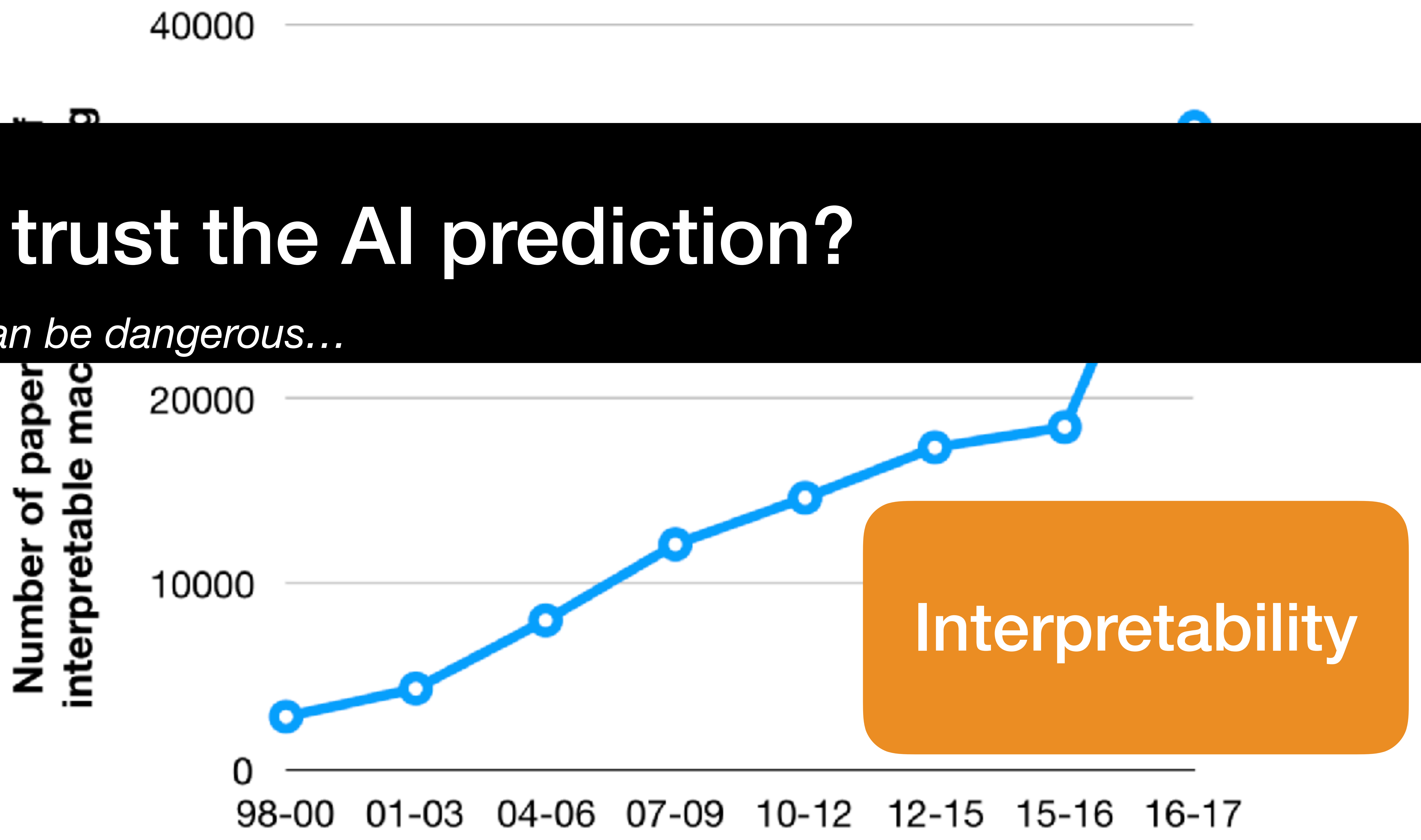
Scientific data may exhibit high dynamic range

But AI models prefer linear behavior…

Data complexity

Source: xkcd

"Please, explain the model!", says the scientist.



Interpretability

Can we simply trust the AI prediction?

Can be dangerous…

Image source: Been Kim (Google AI)

Opportunity

- 1 Data
- Compute 2
- Algorithms 2
- 1 Learning materials

Challenges

- 1 Mapping 2
- 1 Data complexity 2
- 1 Interpretability 2
- Community acceptance 1

1 **Domain scientists**

+

2 **AI experts**

Conclusions

➡ AI is a new programming paradigm

➡ AI is a modern version of empirical models

➡ AI is not a buzzword; it is a new tool available to the scientific community (after some adaption)

➡ AI is not foolproof

➡ Leveraging the potentials of AI in scientific research requires join efforts of domain scientists and AI experts