

Deep Learning

THE
GOOD THE
BAD and THE
UGLY

Qiyang Hu

UCLA Office of Advanced Research Computing

November 4th, 2022

DL application in Science



- Some exciting breakthroughs

01



DL as hopes

- Why DL is so powerful?
- How DL helps in Science?
- Mutual driving forces

02

DL as hypes



- Difficulties in application perspective
- Illusions in theoretical perspective

03



DL like deep oceans

- Cracking the black box
- Thinking outside the box
- What about the future?

04

DL application in Science



- Some exciting breakthroughs

01

DL as hopes

- Difficulties in application perspective
- Illusions in theoretical perspective



03

DL as hopes



02

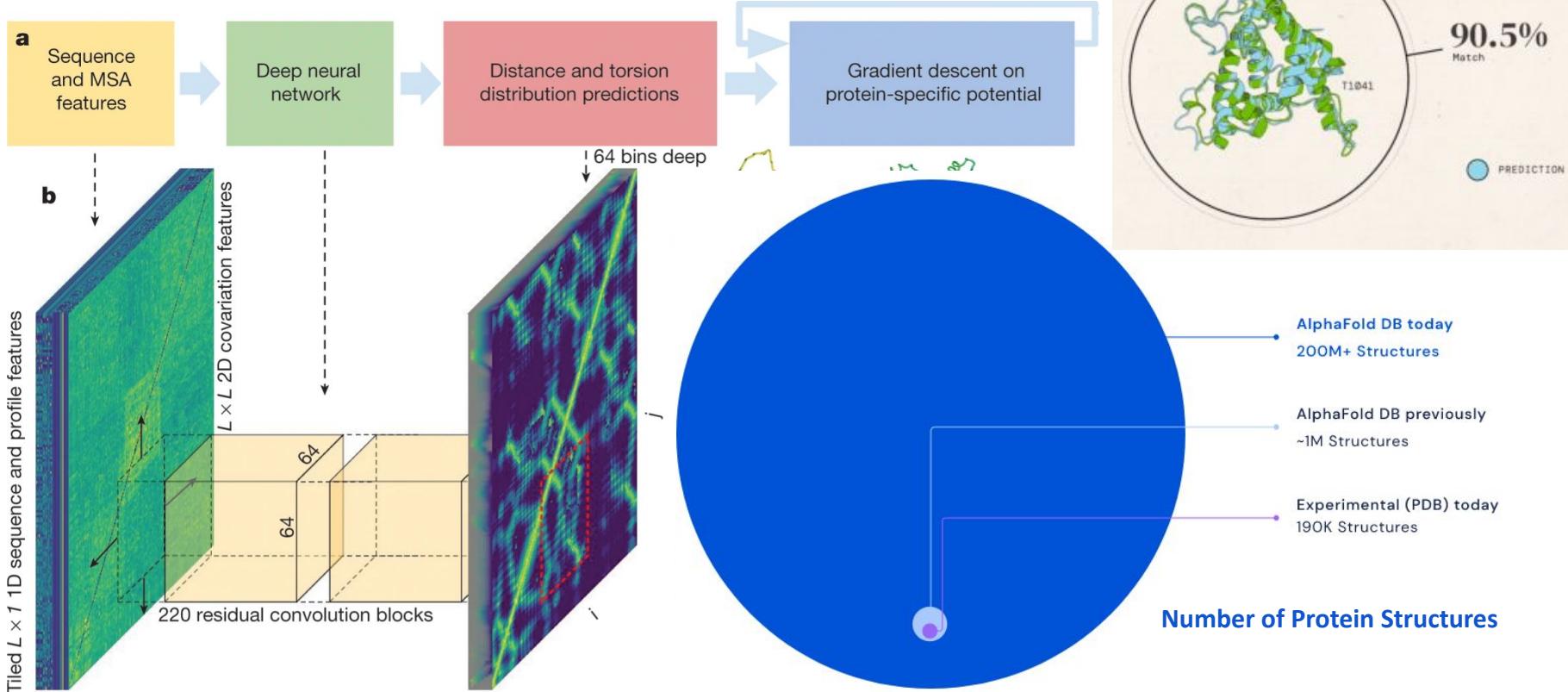
DL like deep oceans



04

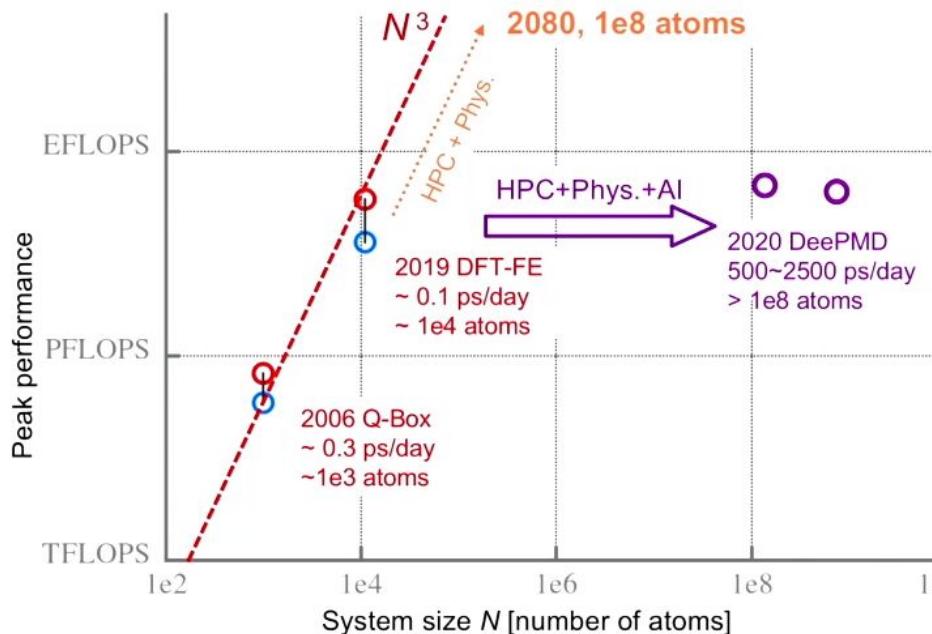
- Cracking the black box
- Thinking outside the box
- What about the future?

AlphaFold: a *stunning* breakthrough

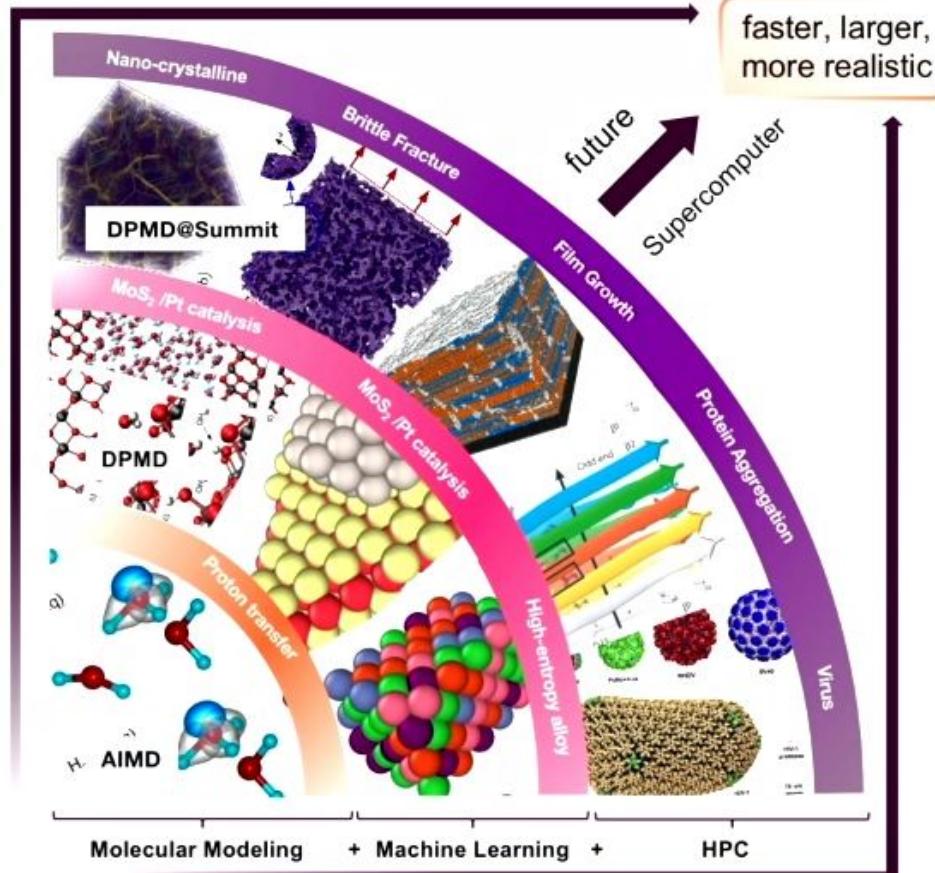


DeePMD ([source](#))

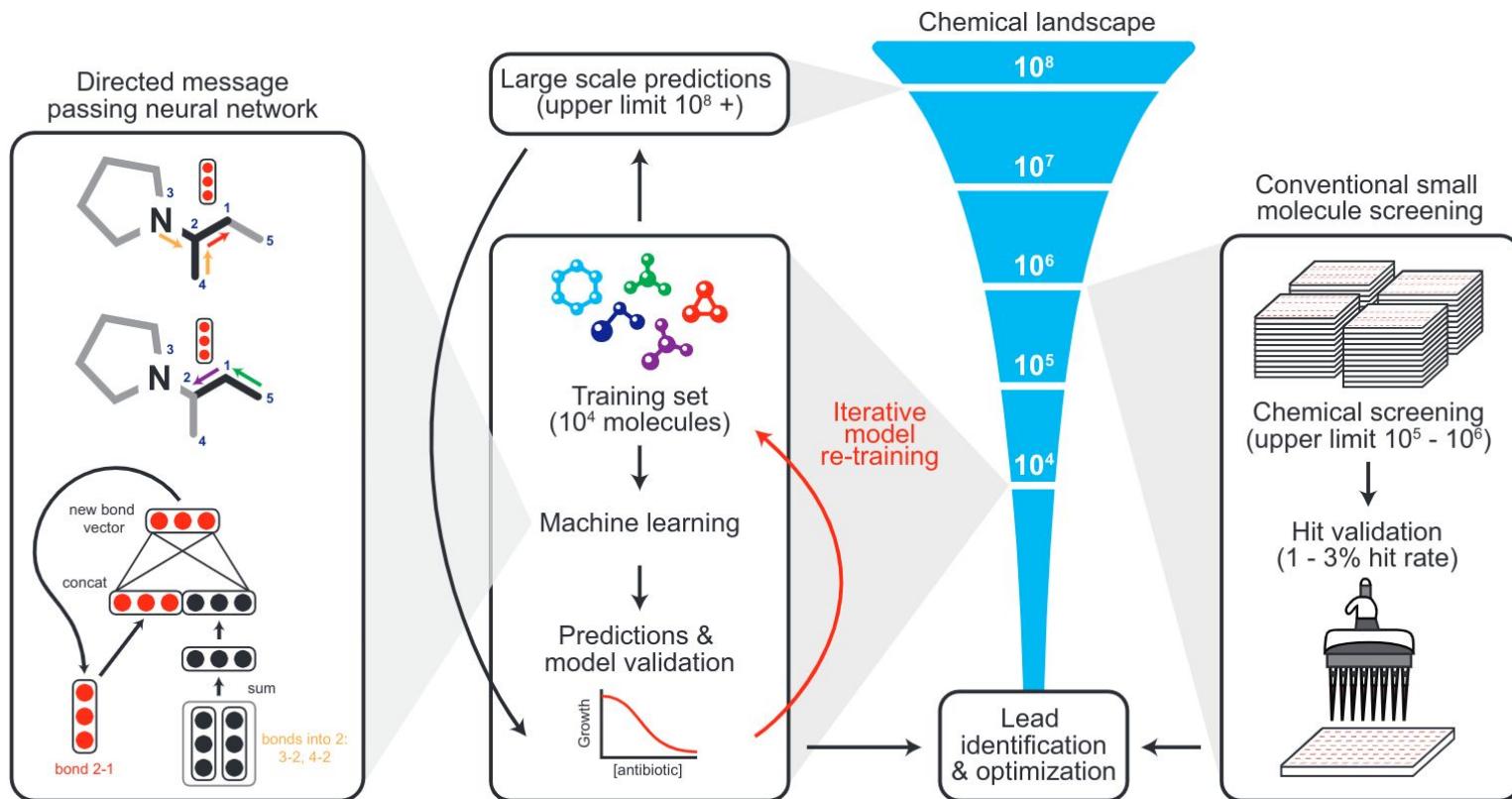
A new HPC+DL+Physical paradigm



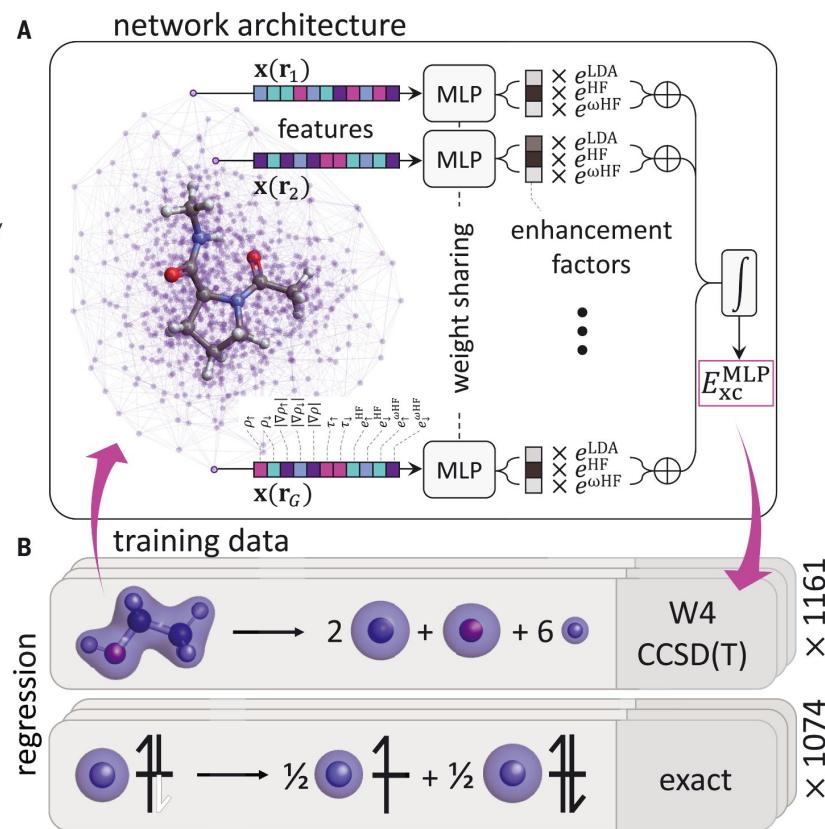
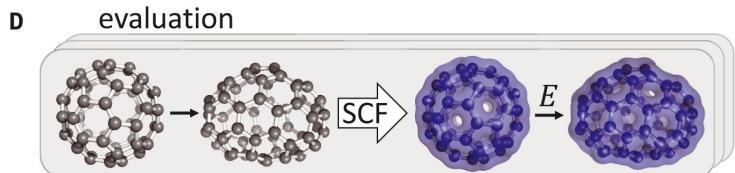
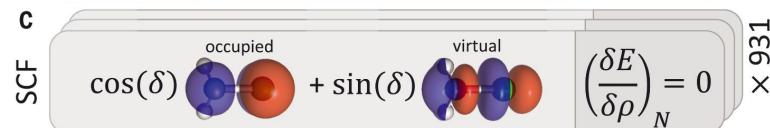
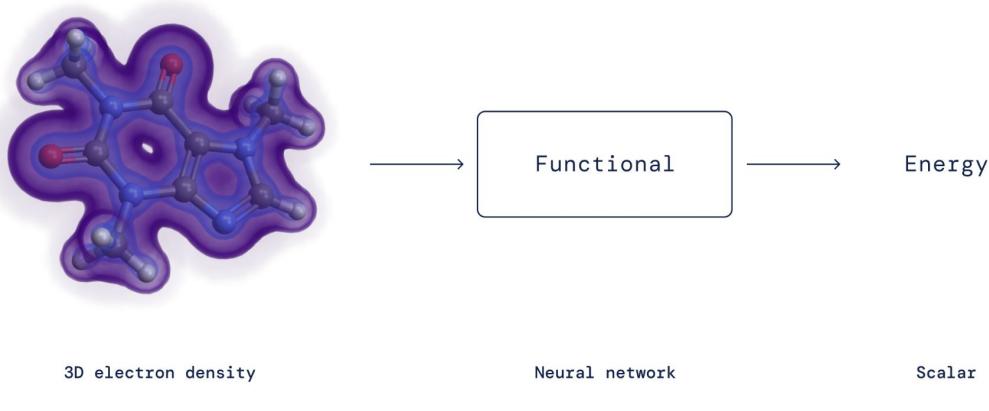
Pictures from their SC'20 talk.



A deep learning approach to antibiotic discovery



Simulating matter on the quantum scale with AI

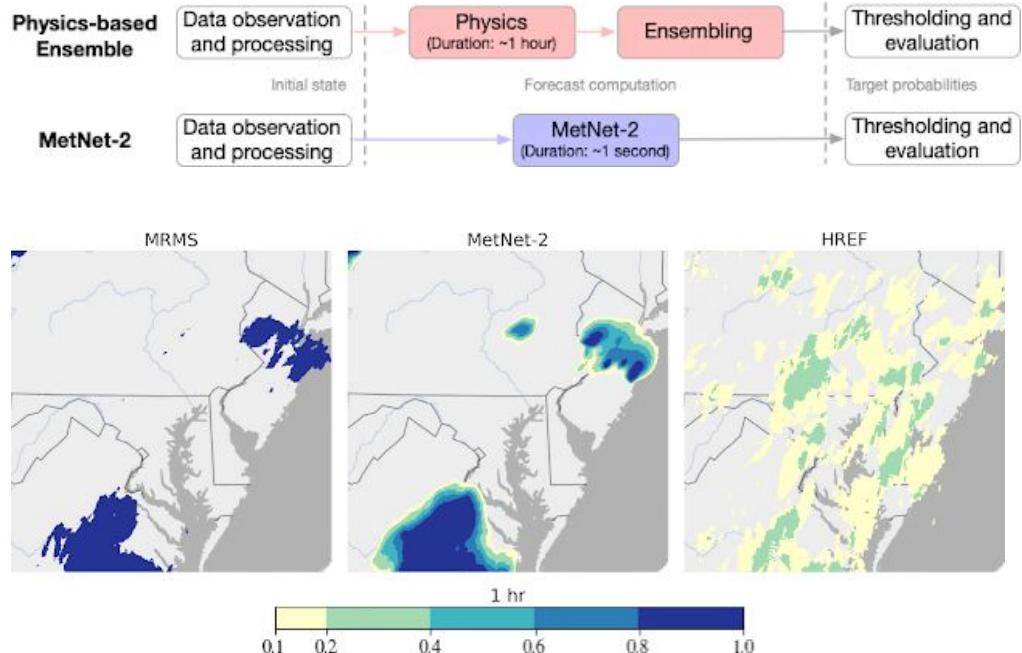


Revolution of numerical weather prediction

[Source](#)



Illustration of the computation through MetNet-2. As the computation progresses, the network processes an ever larger context from the input and makes a probabilistic forecast of the likely future weather conditions.



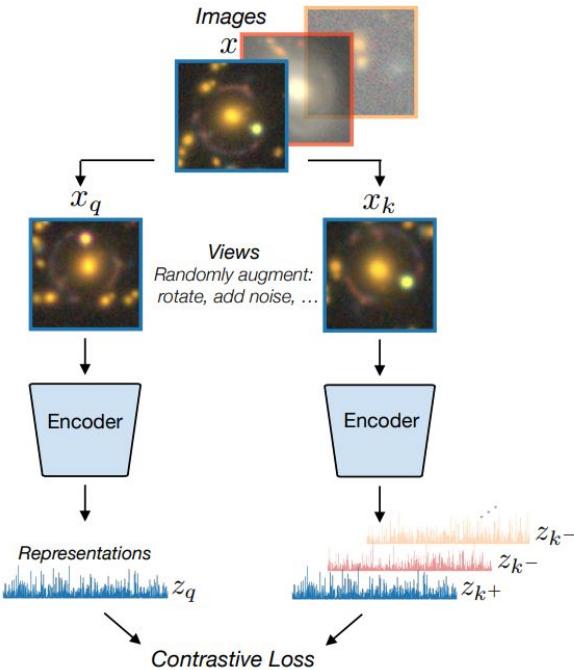
Comparison of 2 mm/hr precipitation stemming from Hurricane Isaias, an extreme weather event that occurred on August 4, 2020 over the North East coast of the US. Left: Ground truth, source MRMS. Center: Probability map as predicted by MetNet-2. Right: Probability map as predicted by HREF.

Qiyang Hu

Astronomical discoveries through self-supervised learning

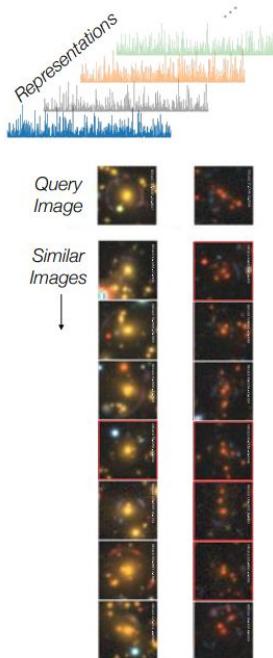
1. Self-supervised contrastive representation learning

Learn representations in an unsupervised manner



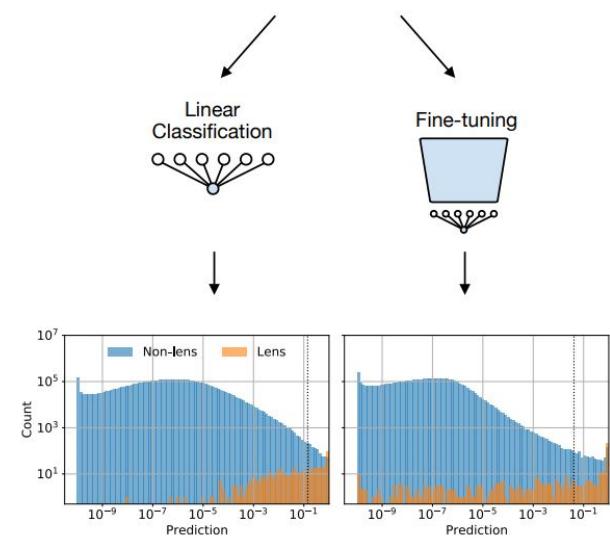
2. Similarity Search

Discover new lenses using a single labelled example



3. Automated classification

Introduce labels



John F Wu

@jwuphysics



I don't think it's an exaggeration to say that some of the biggest astronomical discoveries in the 2020s are only going to be made possible through self-supervised learning (SSL).

<https://t.co/fFMnqHHJC0?amp=1>

Qiyang Hu

DL application in Science



- Some exciting breakthroughs

01



DL as hopes

- Why DL is so powerful?
- How DL helps in Science?
- Mutual driving forces

02

DL as hypes



- Difficulties in application perspective
- Illusions in theoretical perspective

03



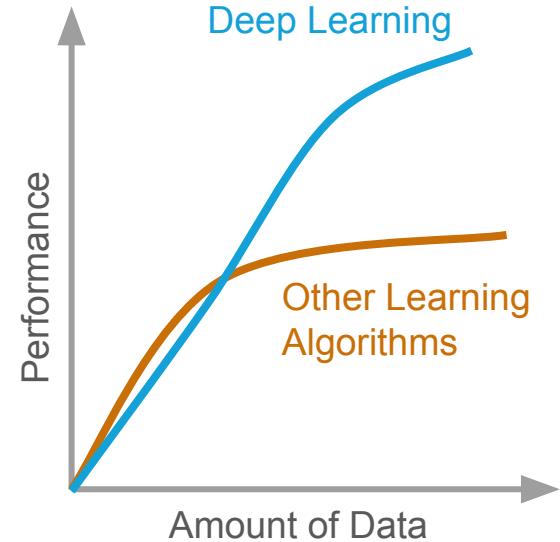
DL like deep oceans

- Cracking the black box
- Thinking outside the box
- What about the future?

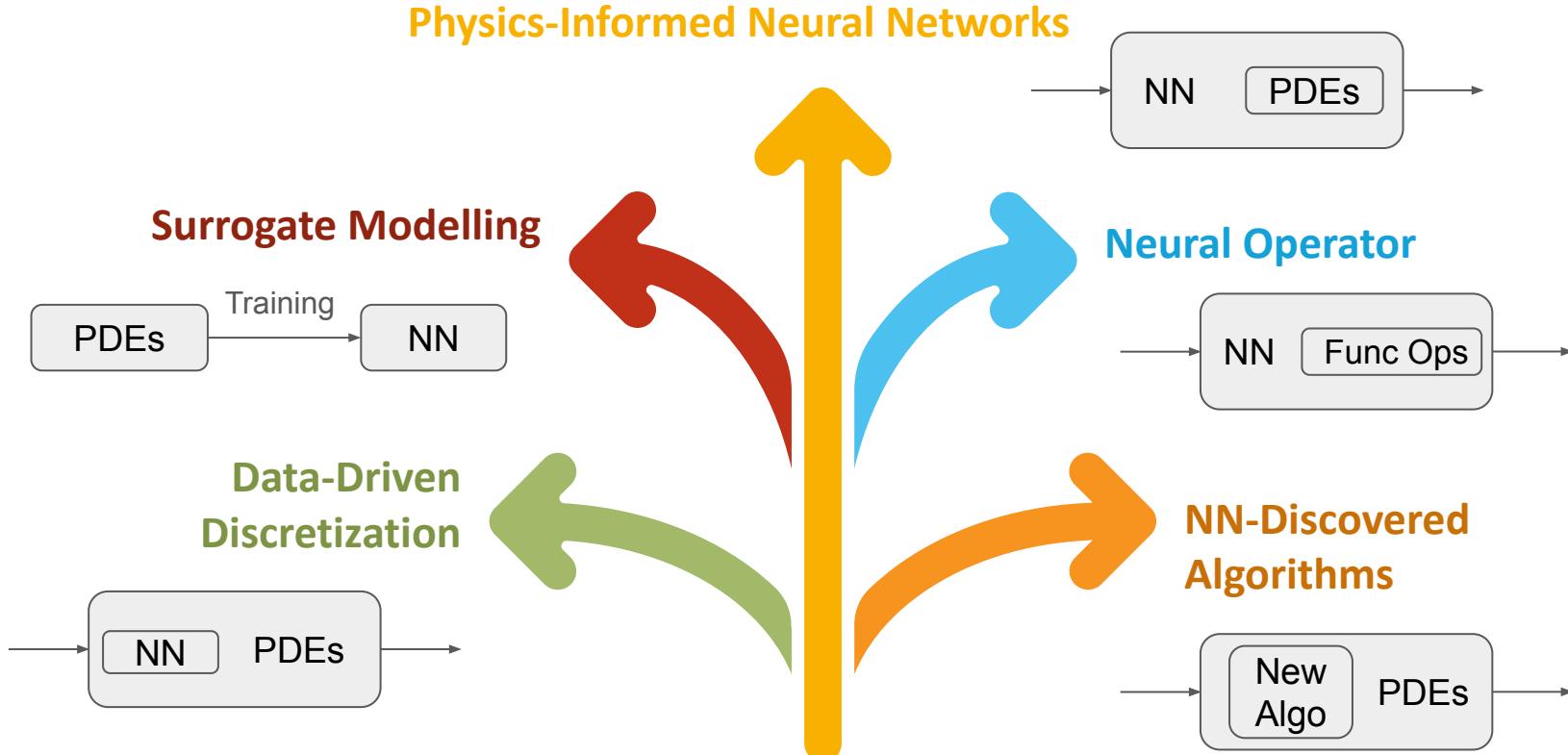
04

A deeper view on connectionism

- Knowledge representation
 - Concepts: vectors of neural activity
 - Relationships between concepts: weight matrices
- Automatic generalization of learning
 - Facilitates analogical reasoning
 - Hierarchical knowledge in a form of compositionality
 - Self-organization from a random start
- Parallel distributed processing
 - Tolerance of partially conflicting evidence
 - Tolerance of graceful degradation
 - Content-addressable memory
 - Low computational costs



Powerful Boosters for Scientific Computing



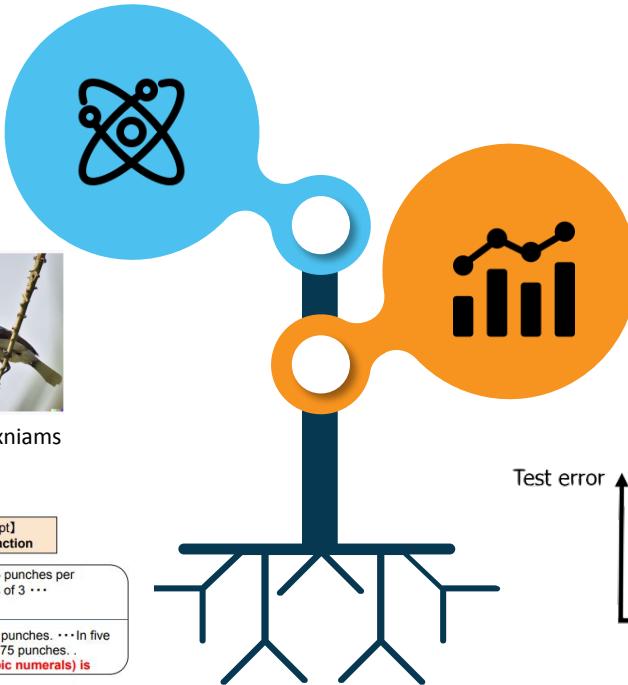
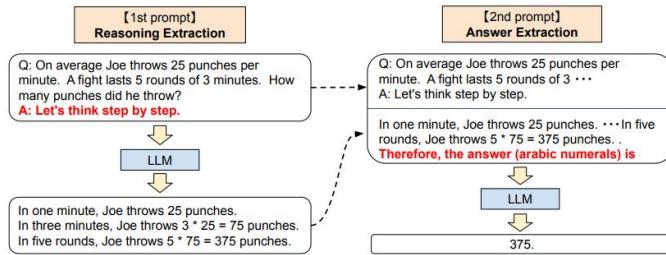
Revolutionary insights from super-big DL models

More “Science” in Computer Science

From design the artifacts to study the systems with “undesigned” properties.

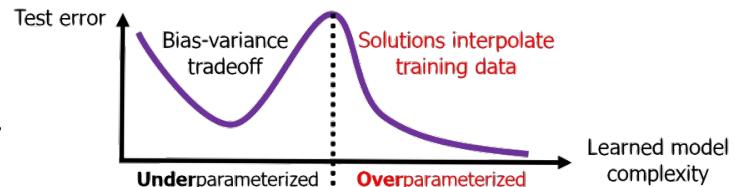


The prompt: "Apoploe vesrreaitais eating Contarra ccetnxniamis luryca tanniounons" gives images of birds eating bugs.



Farewell to Bias-Variance Tradeoffs

Double Descent Behavior



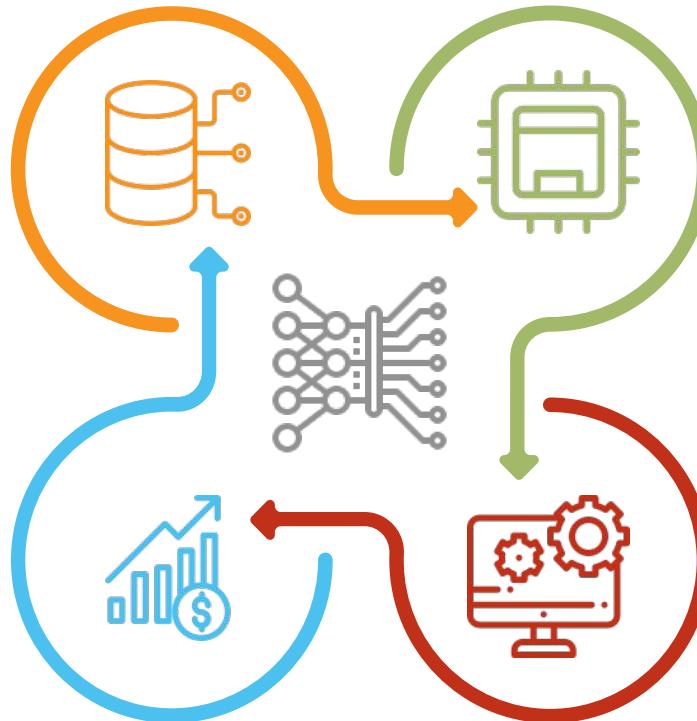
Driving forces and positive impacts

Data

- **ImageNet** database: Milestone of boosting Deep Learning
- **Kaggle**: a home of data scientists and analysts
- **Data explosion** from internet

Economy

- Venture Capital Investment soars **20x** in 8 years
- **Hundreds of million** dollars AI Merger & Acquisitions
- **100x** more people working on deep learning



Hardware

- AI compute amount increases **10 times** per year
- **GPU, TPU, IPU, xPUs, Neuromorphic chips, Quantum AI chips...**

Software

- From C++/Cuda to scripting languages (**Python, R**)
- From library packages to **frameworks**
- From toolsets to open-source pre-trained models

DL application in Science



01

- Some exciting breakthroughs



02

DL as hopes

- Why DL is so powerful?
- How DL helps in Science?
- Mutual driving forces

DL as hypes



03

- Difficulties in application perspective
- Illusions in theoretical perspective



04

DL like deep oceans

- Cracking the black box
- Thinking outside the box
- What about the future?

The more, the better?

Data, high-quality data!

"A child wearing sunglasses is labeled as a 'failure, loser, nonstarter, unsuccessful person' in the original ImageNet database ..."

-- Crawford K. et al. "[Excavating AI](#)"



- Quality of data determines the upper bound of how good the DL model can become.
- Tough tasks for data labeling and annotation
- Avoiding bias brings privacy concerns



Photo by [Tarikul Raana](#) on [Unsplash](#)

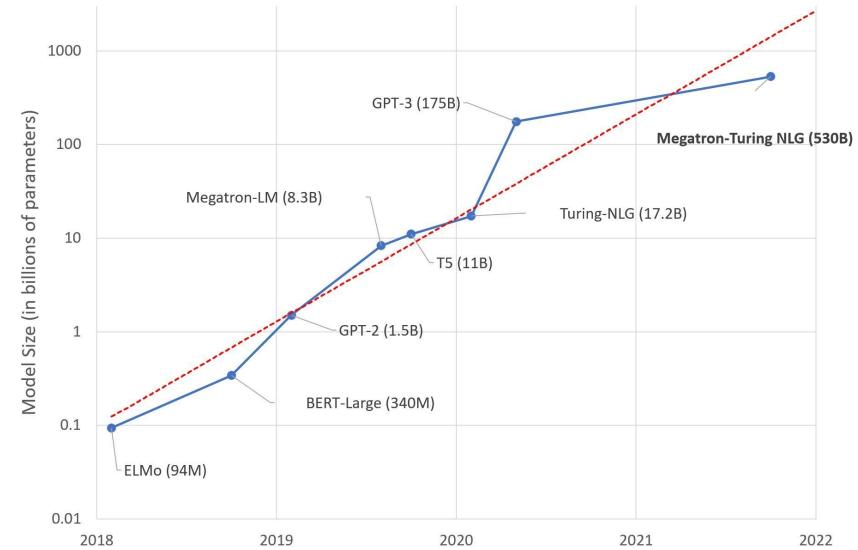
The bigger, the better?

Models grow way too big.

- GPT-1 (2018): 117 million parameters
- GPT-2 (2019): 1.5 billion parameters
- GPT-3 (2020): 175 billion parameters
- GPT-4 (2022): 100 trillion parameters



- Some fundamental limitations still there.
- Lower explainability in models.
- Algorithm efficiency is more important.
- Model's robustness is always a question.



Some GPT-3 Failures

Julie, Marc and John are in the kitchen. They have had 5 bottles.

Julie leaves. The number of persons in the room is thus 3

When I face north, the sun rises from in the east.
Therefore, when I face south, the sun rises in the west.

Magic does not exist. When I rise my left hand, Anna is
taller than Mary. When I rise my right foot, Mary is taller than Anna.

Mary always lies. I asked her whether $2+2=4$, and she said yes.

$$10/0=0.$$

$$-3/3=1/3.$$

$$33333/3=1.$$



The better, the better?

High expectations, low acceptance

Eye screening process in 11 clinics in Thailand

- Before DL evaluation: 2~10 weeks
- With DL evaluation: 10 mins

— [Google Health, SIGCHI 2020](#)

- 
- Poor-equipped environment
 - Unexpected corner cases
 - Practical time/economical concerns

Where is DL in the pandemic?

“...none of them [papers] produced tools that would be good enough to use in a clinical setting. Something has gone seriously wrong when more than 300 papers are published that have no practical benefit.”

— [“Nature Machine Intelligence” 2021](#)

- 
- Paper without details to reproduce results.
 - Papers introduced significant biases.
 - Publication bias towards positive results

Neural Networks ➔ Black Box

Model Transparency

- Interpretations on functions of neurons or layers
- Based on data “observations” in networks
 - By counterfactual explanations

Nonlinearity from Networks

- Complex behavior
 - Gradient exploding/vanishing
 - Autonomous behavior from adaptive loopback structures
- Theoretical understanding on network topologies
 - Residue networks with skip connections
 - Graph neural network
 - Transformer structures
 - GAN and Diffusion generative models



Some limiting cases

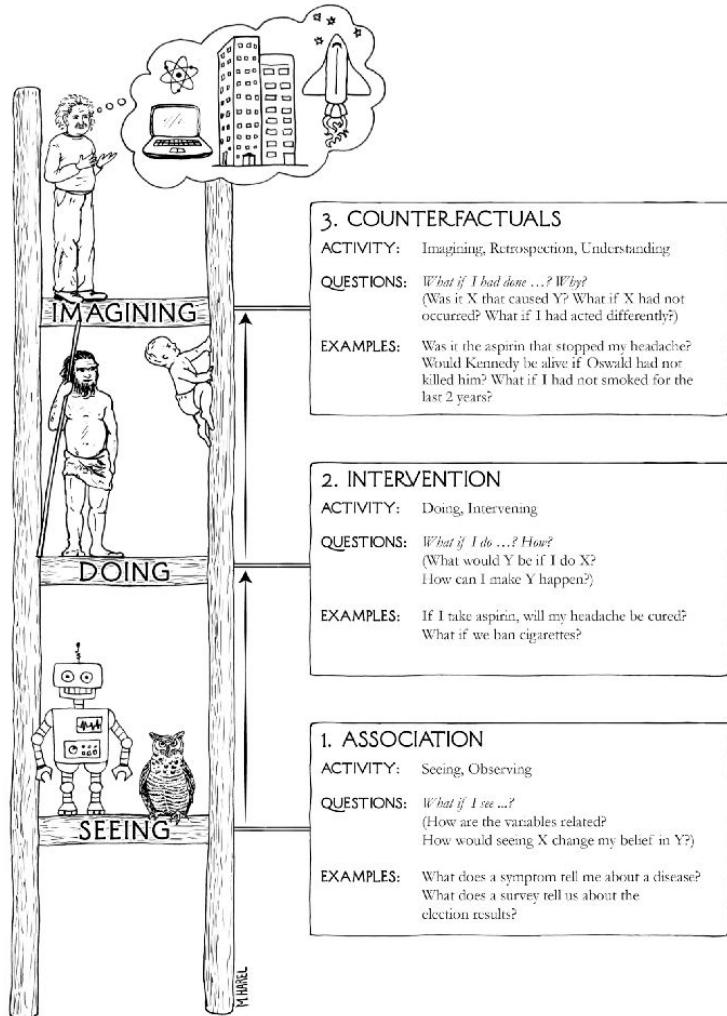
- Infinite-wide Neural Network
 - Neural Tangent Kernel
- Infinite number of recurrent self layers:
 - Deep Implicit Layers
 - Generalizing layers as complex compute units

Nonlinearity from Neurons

- Why Gradient Descent works well?
- How to explain Double Descent?
- Theoretical guidelines to choose:
 - Hyperparameters
 - Parameter initializations
 - Activation functions

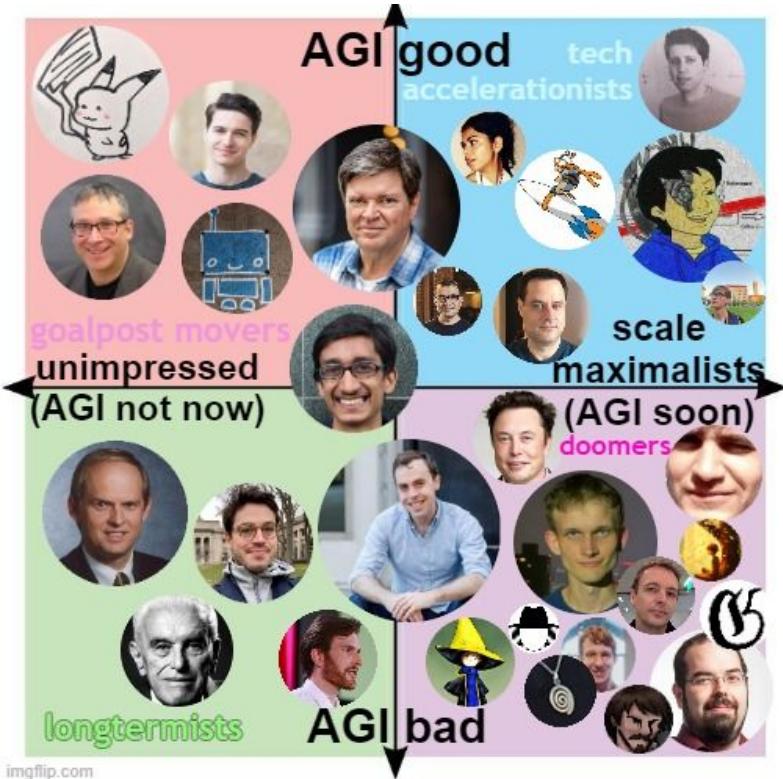
From 0 to 1

- Data mining \Rightarrow understanding?
 - Neural network = Curve fitting ?
 - Statistical = model-blind ?
- DL with other sources of knowledge



Real intelligence?

- Artificial general intelligence (AGI)
 - DL + Big Data + ... = Human-level Intelligence?
 - Where is backprop in our brain?
- A curse from Gödel's theorem
 - Maybe it is just impossible.
 - We don't even know what AI is/means.
- What *if* the Singularity finally happens
 - S-believers vs S-sceptics
 - Actions from AI-communities, policy makers



DL application in Science



01

- Some exciting breakthroughs

DL as hypotheses



03

- Difficulties in application perspective
- Illusions in theoretical perspective

DL as hopes



02

- Why DL is so powerful?
- How DL helps in Science?
- Mutual driving forces



04

DL like deep oceans

- Cracking the black box
- Thinking outside the box
- What about the future?



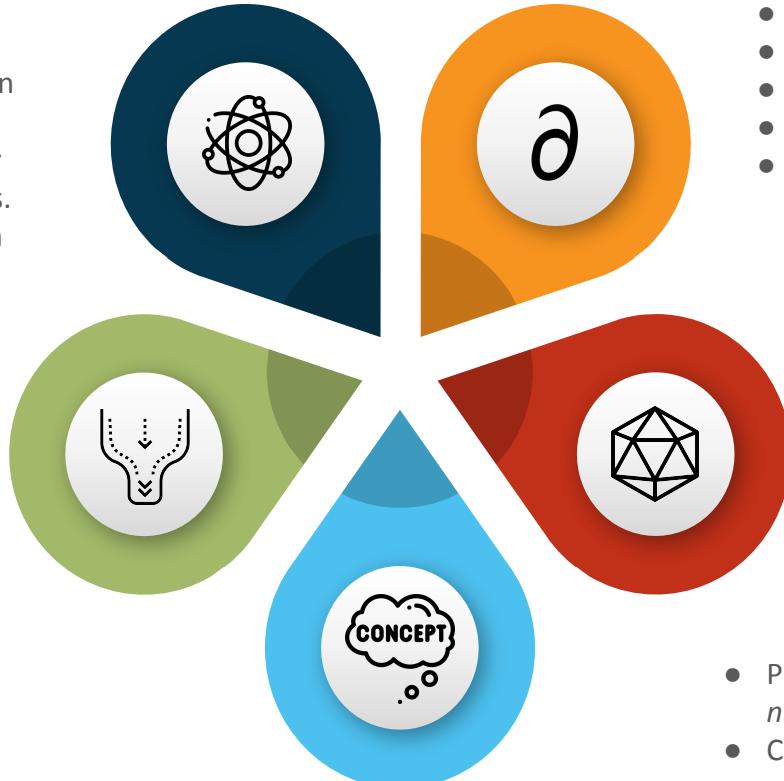
Understanding Neural Networks

Effective Theory

- NN \Rightarrow solving layer-to-layer iteration eqns + nonlinear learning dynamics
- Explaining nonlinearity to nth-order
- NN Predictions: Nearly Gaussian Dis.
- Depth-2-width ratio: deviation from ∞ -width Gaussian description
- Representation group flow: tuning the networks to criticality
- [Reference](#)

Information Bottleneck

- NN: squeezing the information through a bottleneck
- Deep Learning proceeds 2 phases:
 - A short “fitting”
 - A much longer “compression”
- Network converges to information bottleneck theoretical bound
- [Reference](#)



Discretized Dynamics

- Deep Network \Rightarrow Diff Eqns (DE)
- Network Architecture \Rightarrow Numerical DE
- Network Training \Rightarrow Optimal Control
- ResNet \Rightarrow One type of discretization
- Forward & Backprop \Rightarrow ODE solvers
- [Reference](#)

Geometric Unification

- Symmetry and invariance
- Unify CNN, GNN, RNN, Transformer
- 5Gs: Grid, Groups, Graph/Manifold, Geodesics, Gauges
- A principled way to construct new NN frameworks

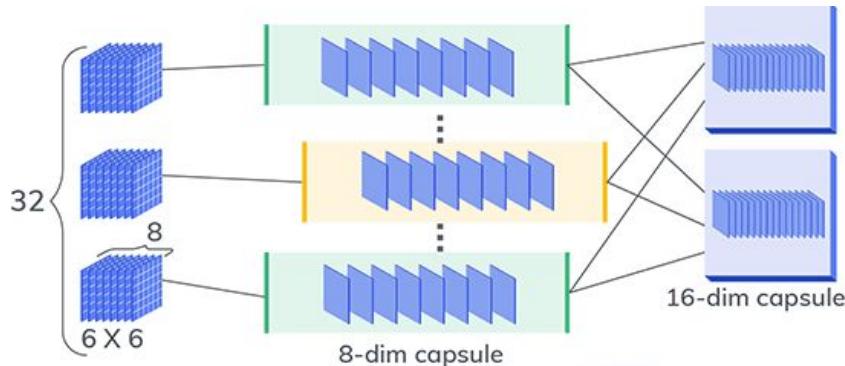
Analytical Philosophy

- Physical symbols: entire networks of cells, *not* locatable neurons
- Concepts: partially conflicting constraints, *not* strict logical definitions

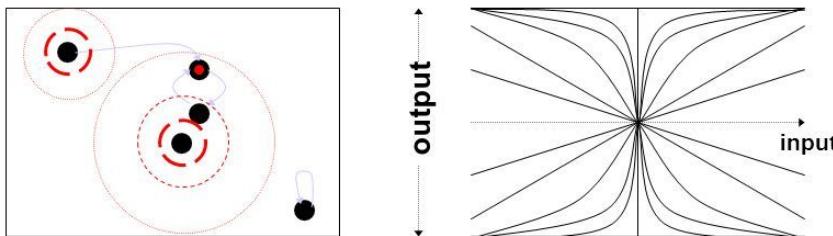


Alternative Neural Networks

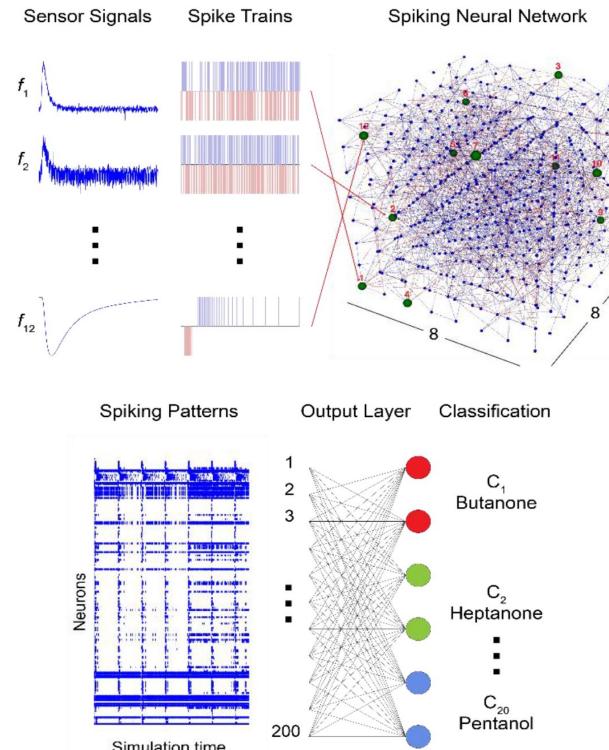
Capsule Network



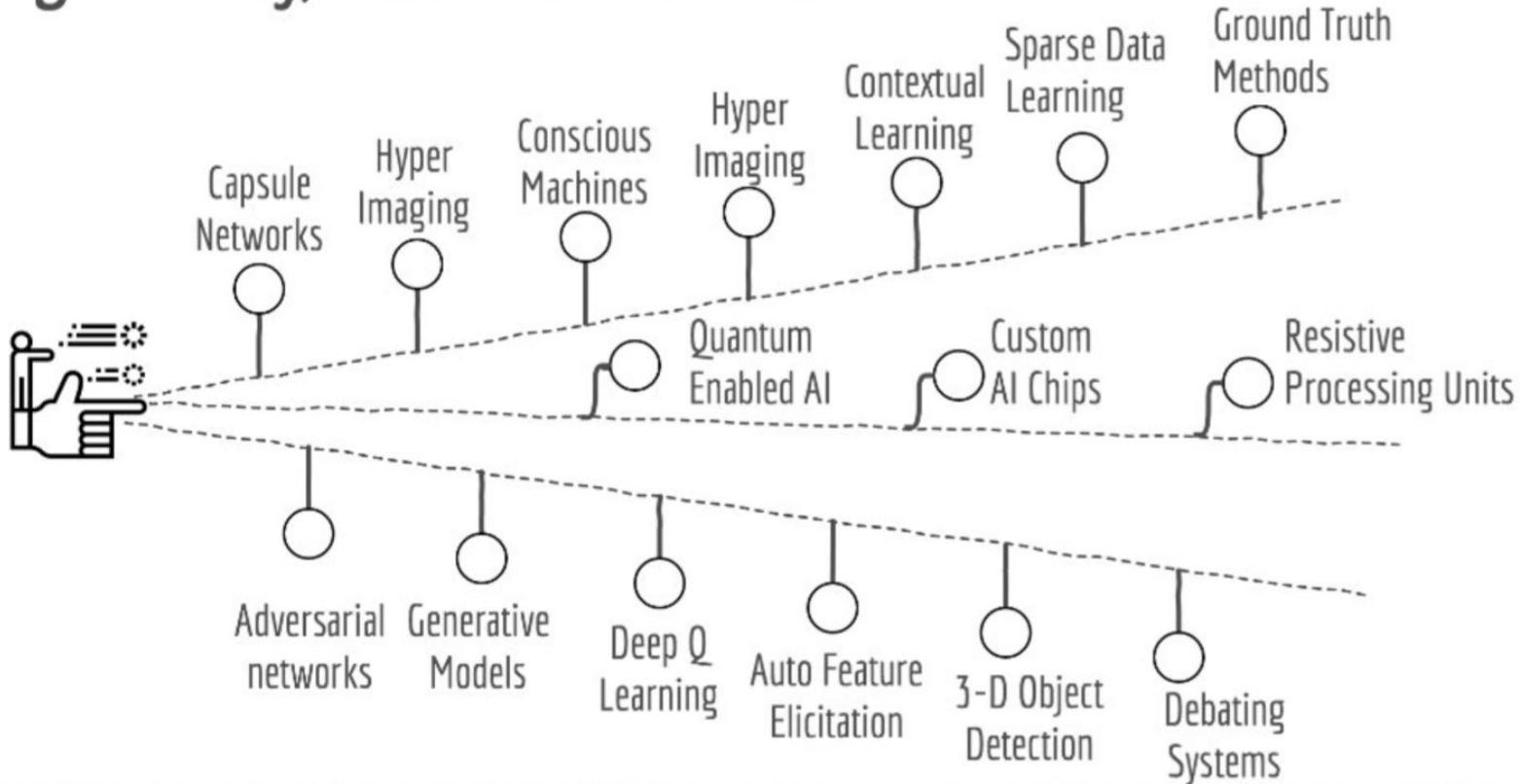
GasNets



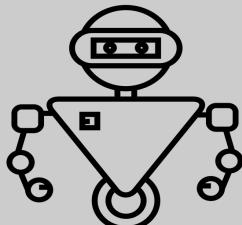
Spiking Neural Network



Pragmatically, what's next for AI?



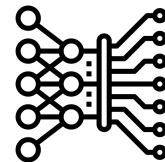
Artificial Intelligence



Machine Learning



Deep Learning



Symbolic
("GOF") AI

Connectionism

50's

60's

70's

80's

90's

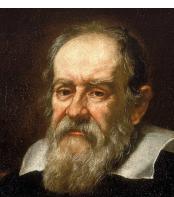
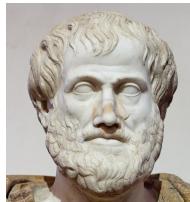
00's

10's

20's

30's

Time



?



OARC Workshop Survey

<http://bit.ly/3fv3w6r>