Deep Learning

Qiyang Hu

UCLA Office of Advanced Research Computing

November 8th, 2021

What makes our Deep Learning workshops special?

- Different from other learning resources
  - **Broad**: high-level descriptive review
  - **Intuitive**: avoiding math to explain the fundamentals
  - **Practical**: jupyter notebook examples on basic techniques

- Different from my previous LDL series
  - More **insights**: esp. from AI etc.
  - More **critical** perspectives
  - More research in **Science**

- Plans:
  - This quarter: General discussions on DL, Learning Mech, PyTorch
  - Next quarter: Specific topics on CNNs, GANs, RNN...

Neutral

General Intro

- What's ANN?
- DL, ML and AI

01

Good

DL as hopes

- Why DL is so powerful?
- Exciting achievements
- Mutual driving forces

02

Bad

DL as hypes

- Difficulties in application perspective
- Illusions in theoretical perspective

03

"Ugly"

DL like deep oceans

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

04

Neutral

General Intro

- What's ANN?
- DL, ML and AI

01

Good

DL as hopes

- Why DL is so powerful?
- Exciting achievements
- Mutual driving forces

02

Bad

DL as hypes

- Difficulties in application perspective
- Illusions in theoretical perspective

03

"Ugly"

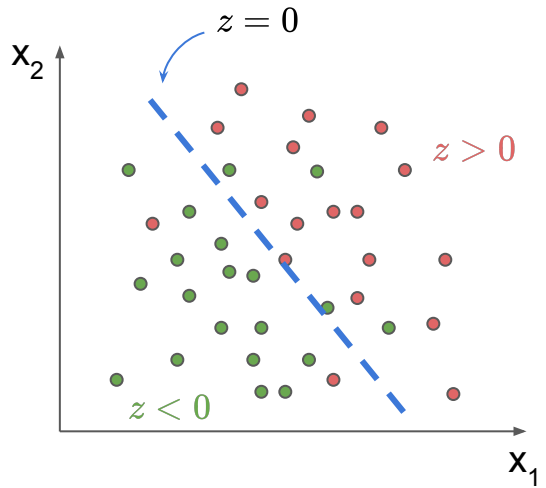
DL like deep oceans

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

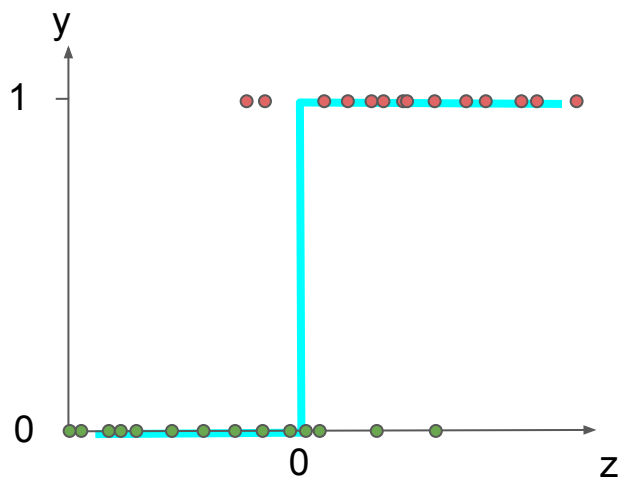
04

What is Neural Network?

- Recap for simple linear classification problem



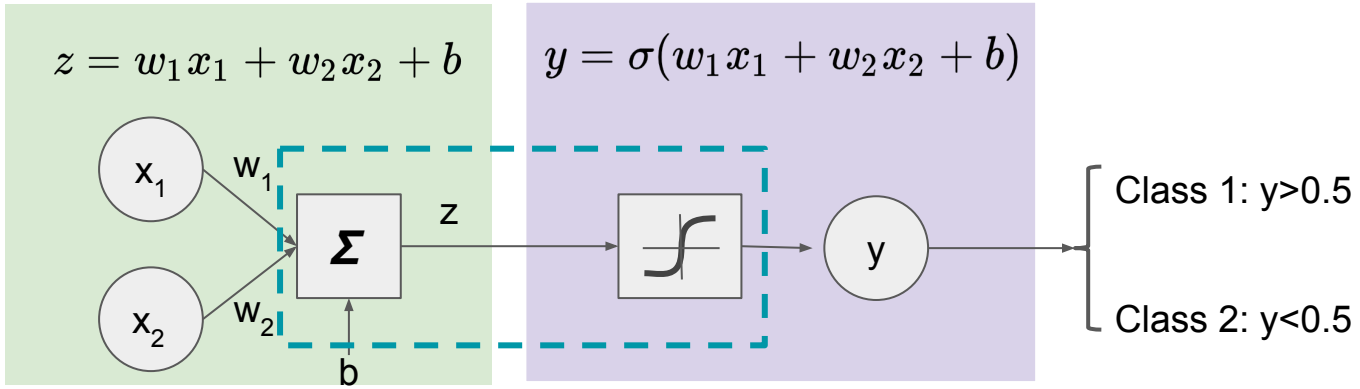
$$z = w_1 x_1 + w_2 x_2 + b$$

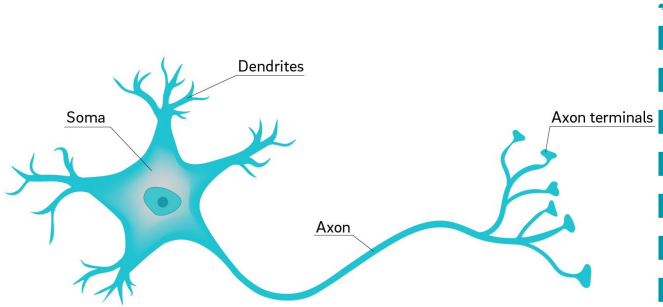


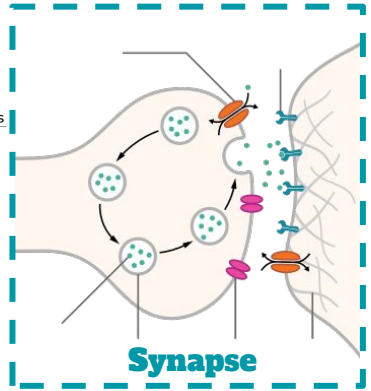
$$y = \sigma(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z < 0 \end{cases}$$

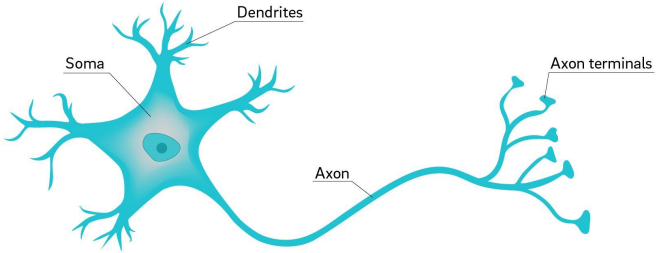
Artificial Neuron and Biological Neuron

McCulloch-Pitts (MCP) neuron model

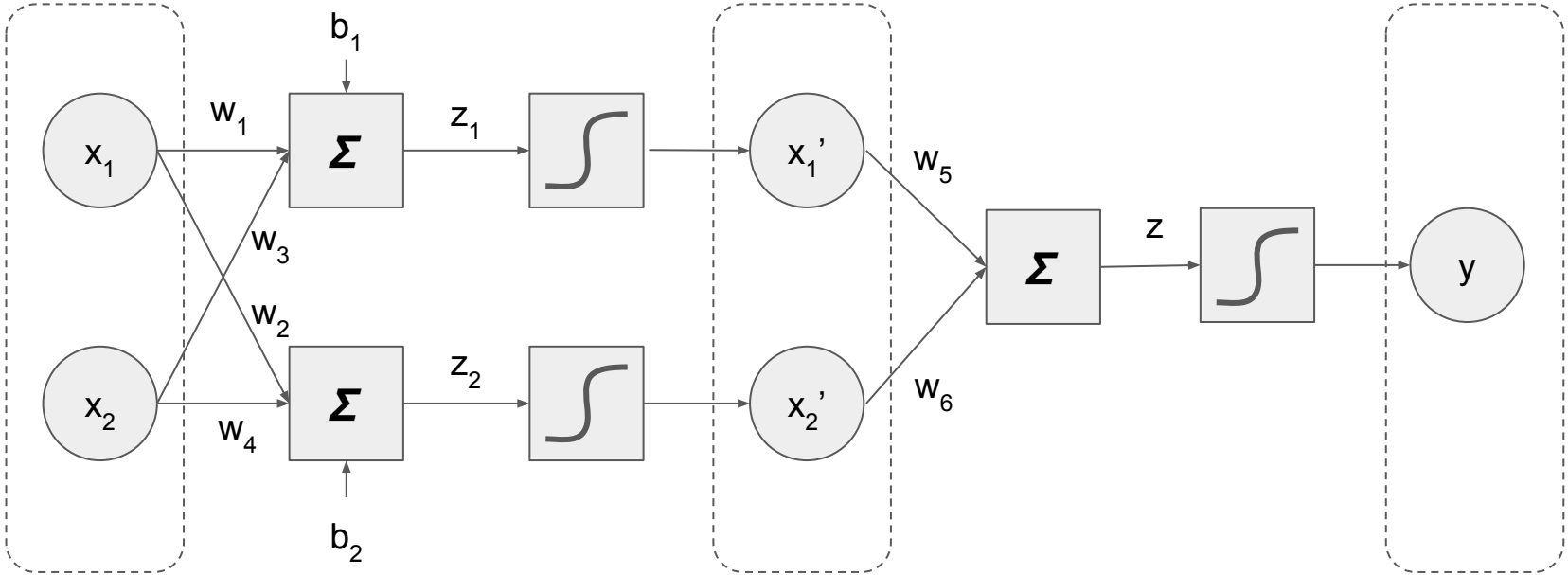








Neural Networks ~ piling/stacking logistic-regression classifiers

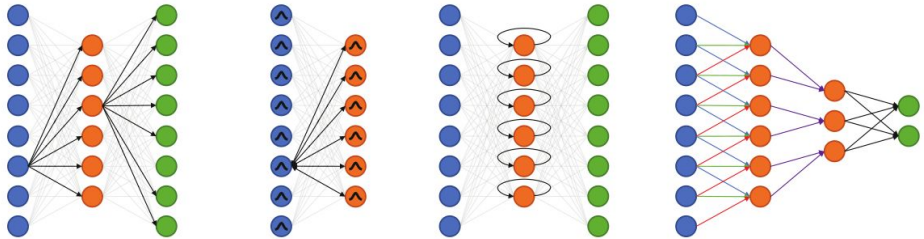


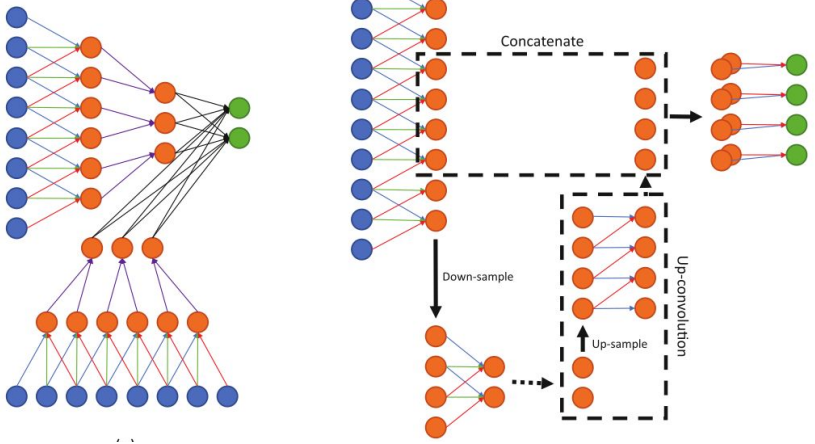
Deep neural networks

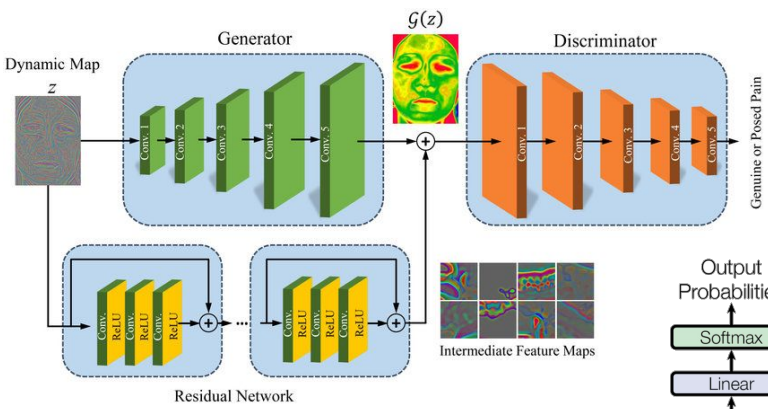
- LeNet-5 (1998)

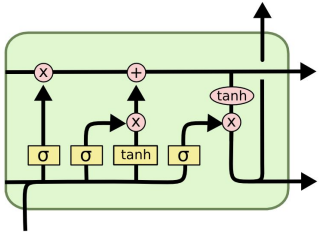
Year	CNN	Developed by	Place	Top-5 error rate	No. of parameters
1998	LeNet(8)	Yann LeCun et al			60 thousand
2012	AlexNet(7)	Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever	1st	15.3%	60 million
2013	ZFNet()	Matthew Zeiler and Rob Fergus	1st	14.8%	
2014	GoogLeNet(19)	Google	1st	6.67%	4 million
2014	VGG Net(16)	Simonyan, Zisserman	2nd	7.3%	138 million
2015	ResNet(152)	Kaiming He	1st	3.6%	

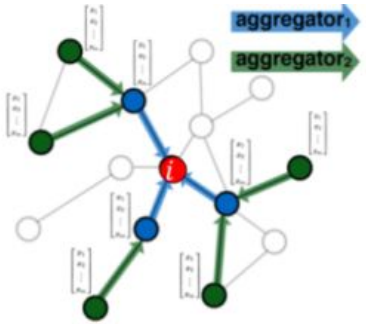
Deep neural ***networks***

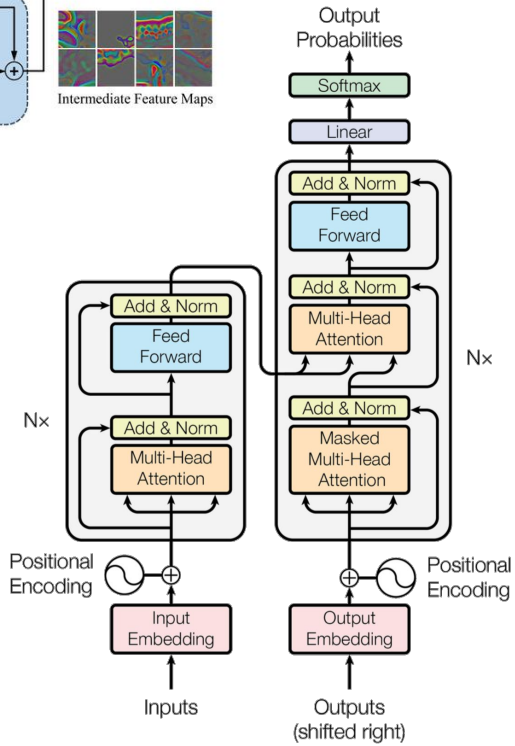




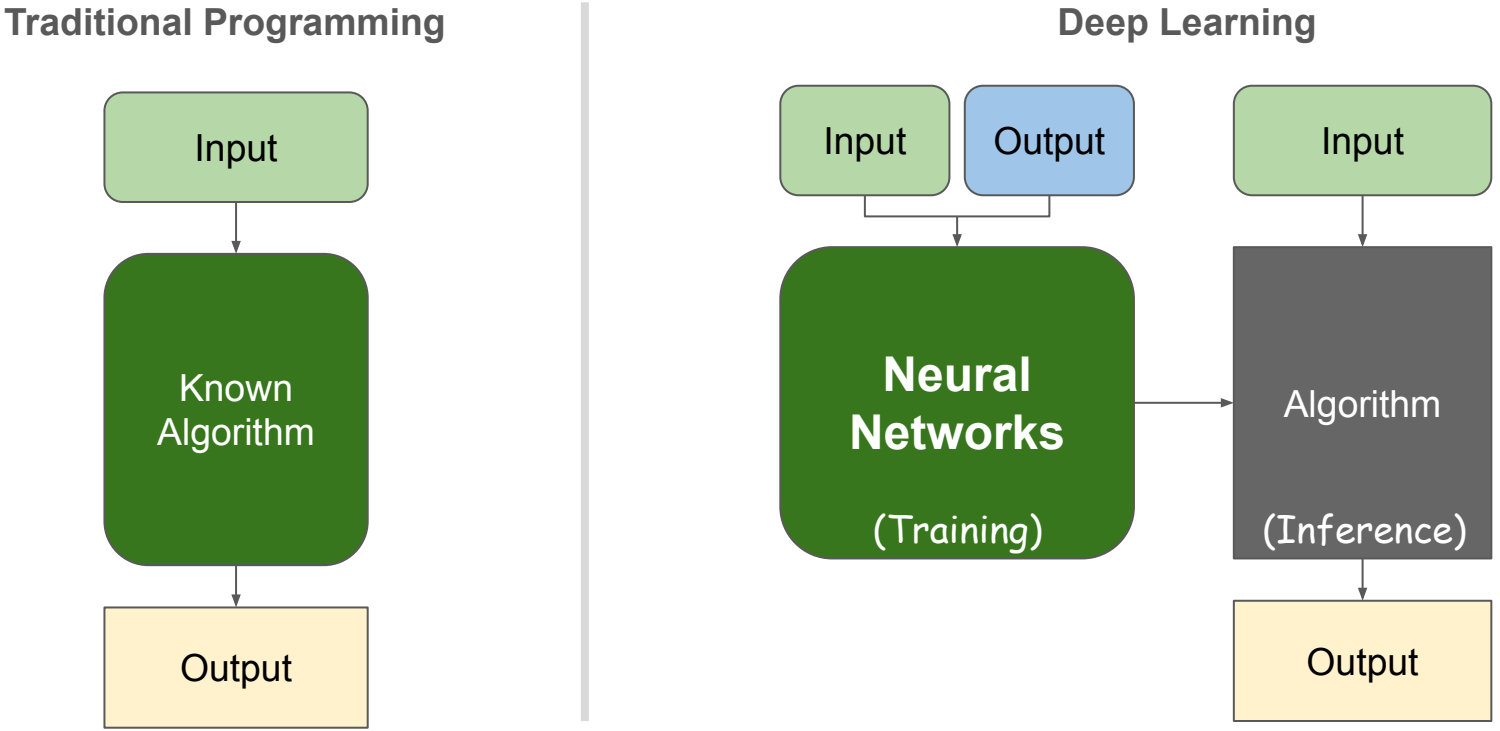




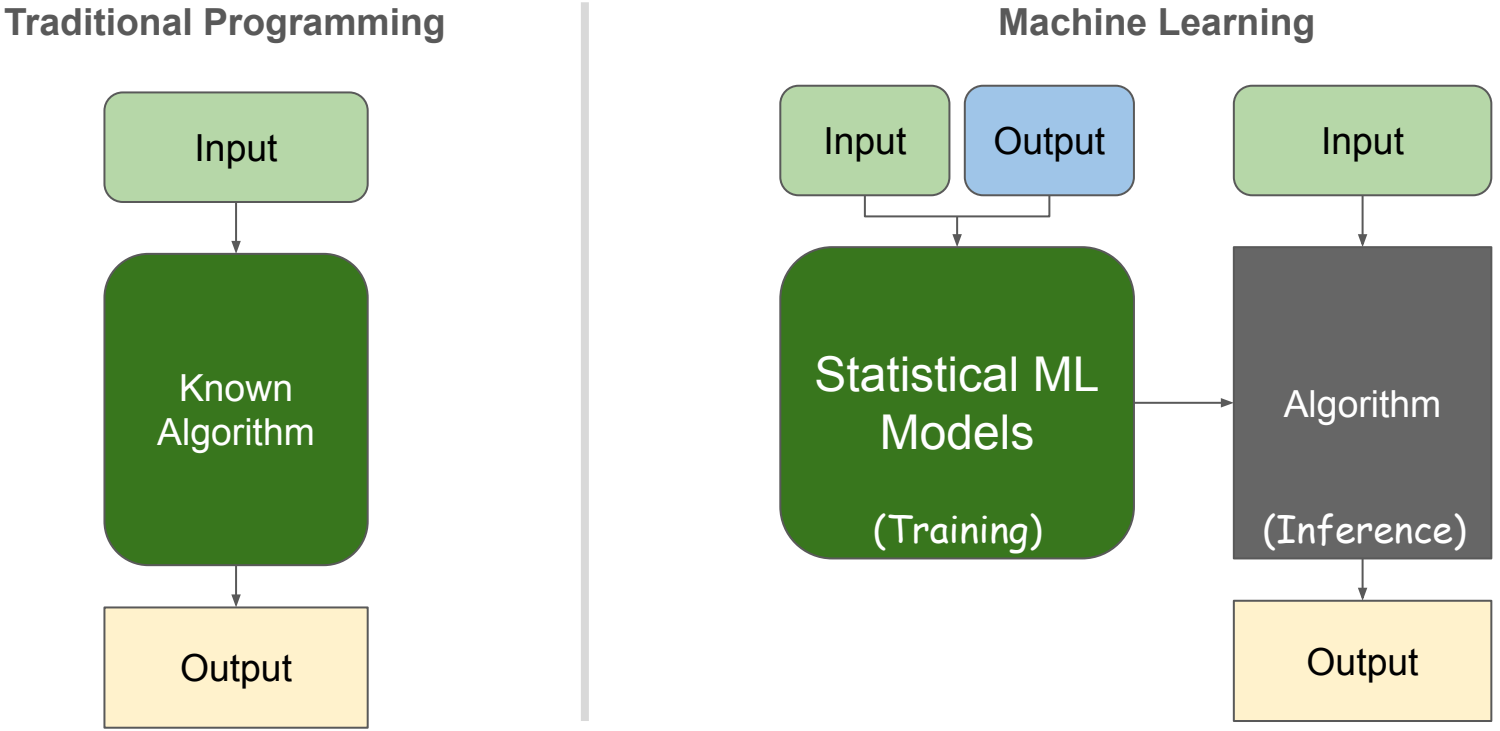


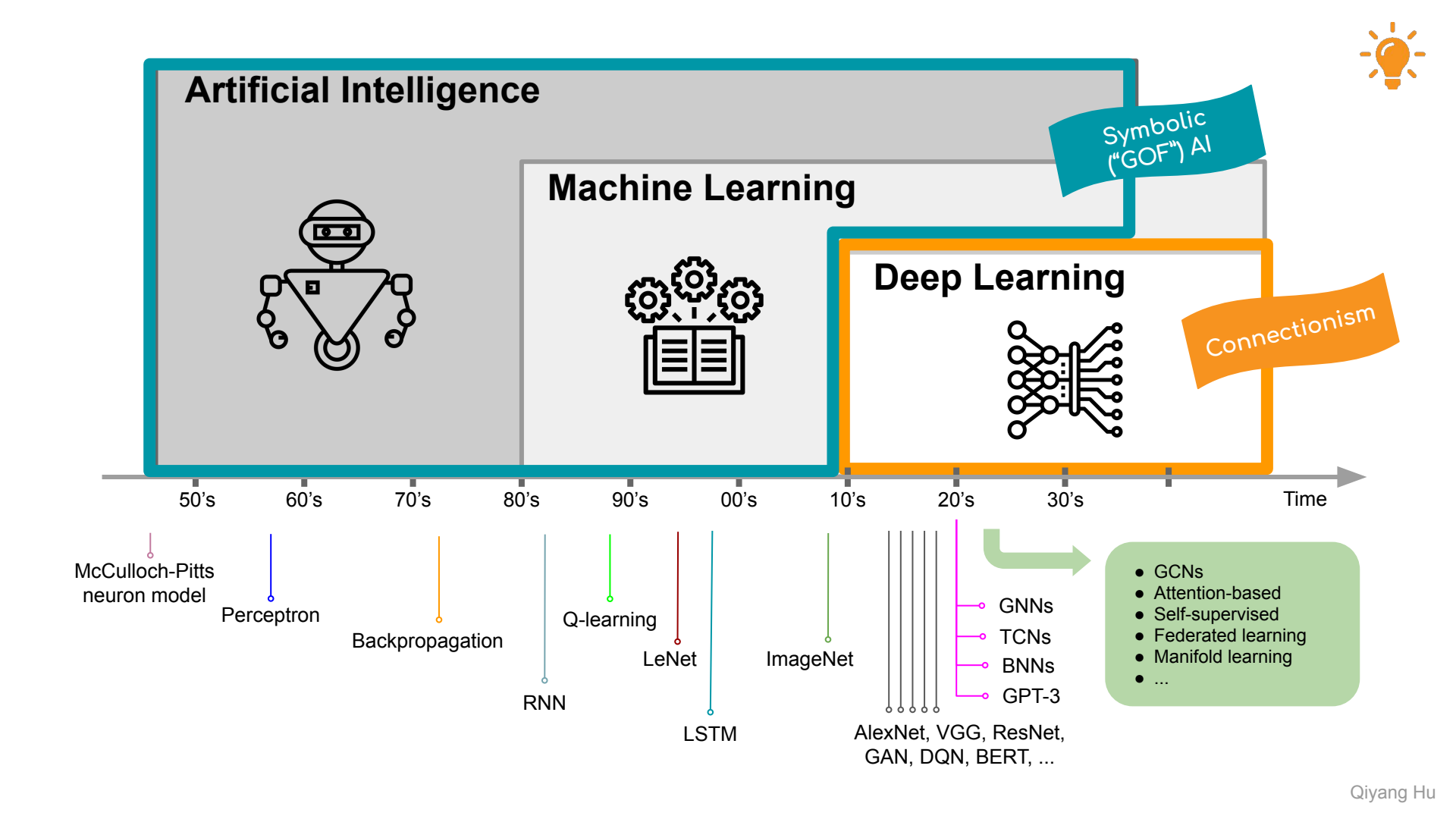


What is Deep Learning?



What is Machine Learning?




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
McCulloch-Pitts neuron model
Perceptron
Backpropagation
RNN
Q-learning
LeNet
LSTM
ImageNet
AlexNet, VGG, ResNet, GAN, DQN, BERT, ...
GNNs
TCNs
BNNs
GPT-3
GCNs
Attention-based
Self-supervised
Federated learning
Manifold learning
...

Neutral

General Intro

- What's ANN?
- DL, ML and AI

01

Good

DL as hopes

- Why DL is so powerful?
- Exciting achievements
- Mutual driving forces

02

Bad

DL as hypes

- Difficulties in application perspective
- Illusions in theoretical perspective

03

"Ugly"

DL like deep oceans

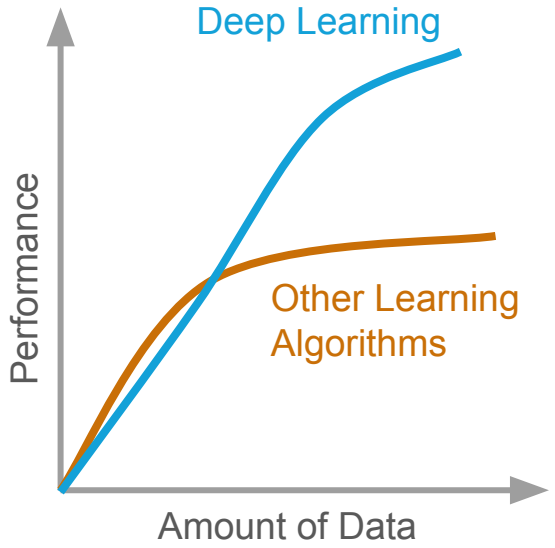
- Cracking the black box
- Thinking outside the box
- What about the future?

04

DATA CENTER & INTERNET	RESEARCH HIGHER EDUCATION	HEALTHCARE	FINANCE	AUTOMOTIVE
SEARCH RECOMMENDATION	SIMULATION LANGUAGE	DRUG DISCOVERY GENOMICS	QUANT FINANCE FRAUD DETECTION	IMAGE DETECTION, CLASSIFICATION
ML DOMAIN				

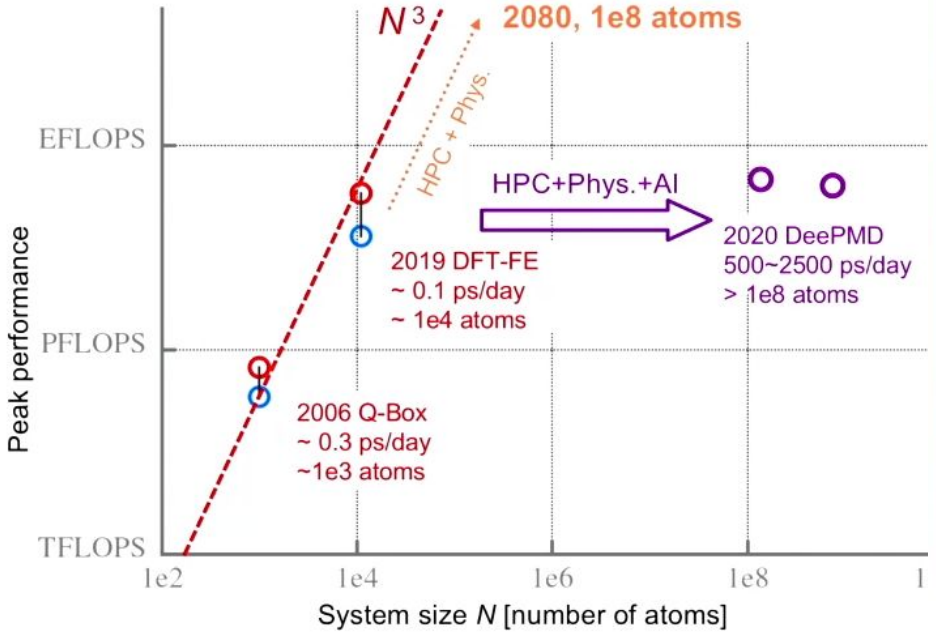
A deeper view on connectionism

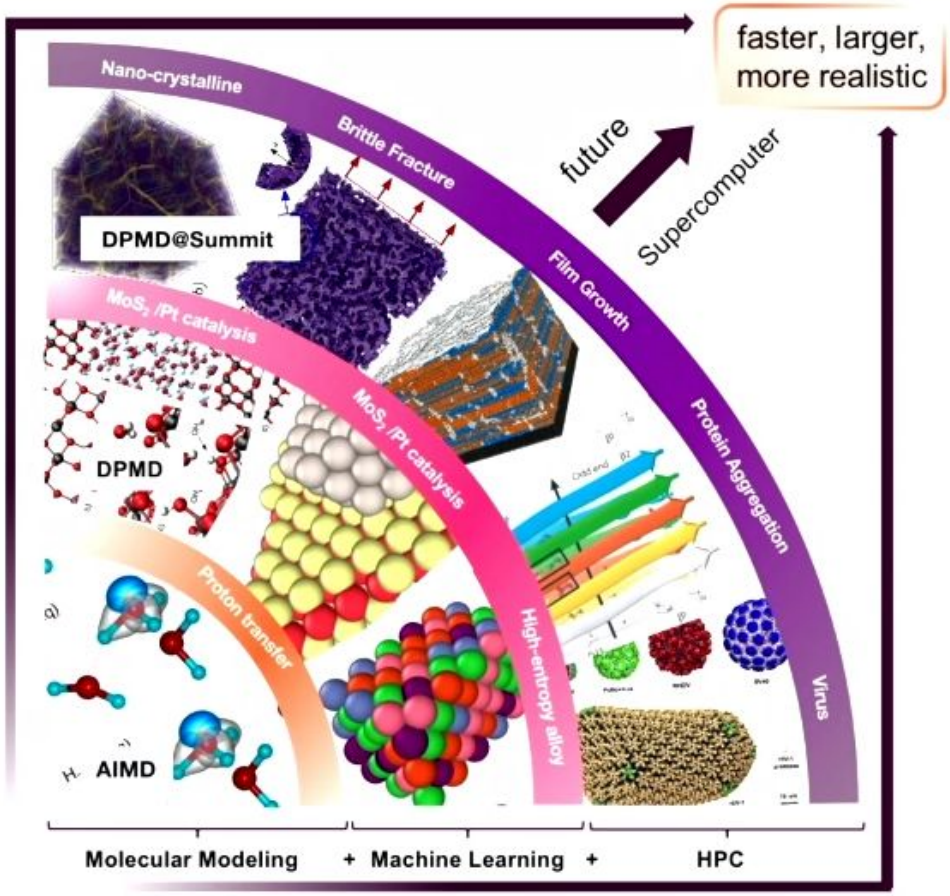
- Knowledge representation
  - Concepts: vectors of neural activity
  - Relationships between concepts: weight matrices
- Automatic generalization of learning
  - Facilitates analogical reasoning
  - Hierarchical knowledge in a form of compositionality
  - Self-organization from a random start
- Parallel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



DeePMD (source)

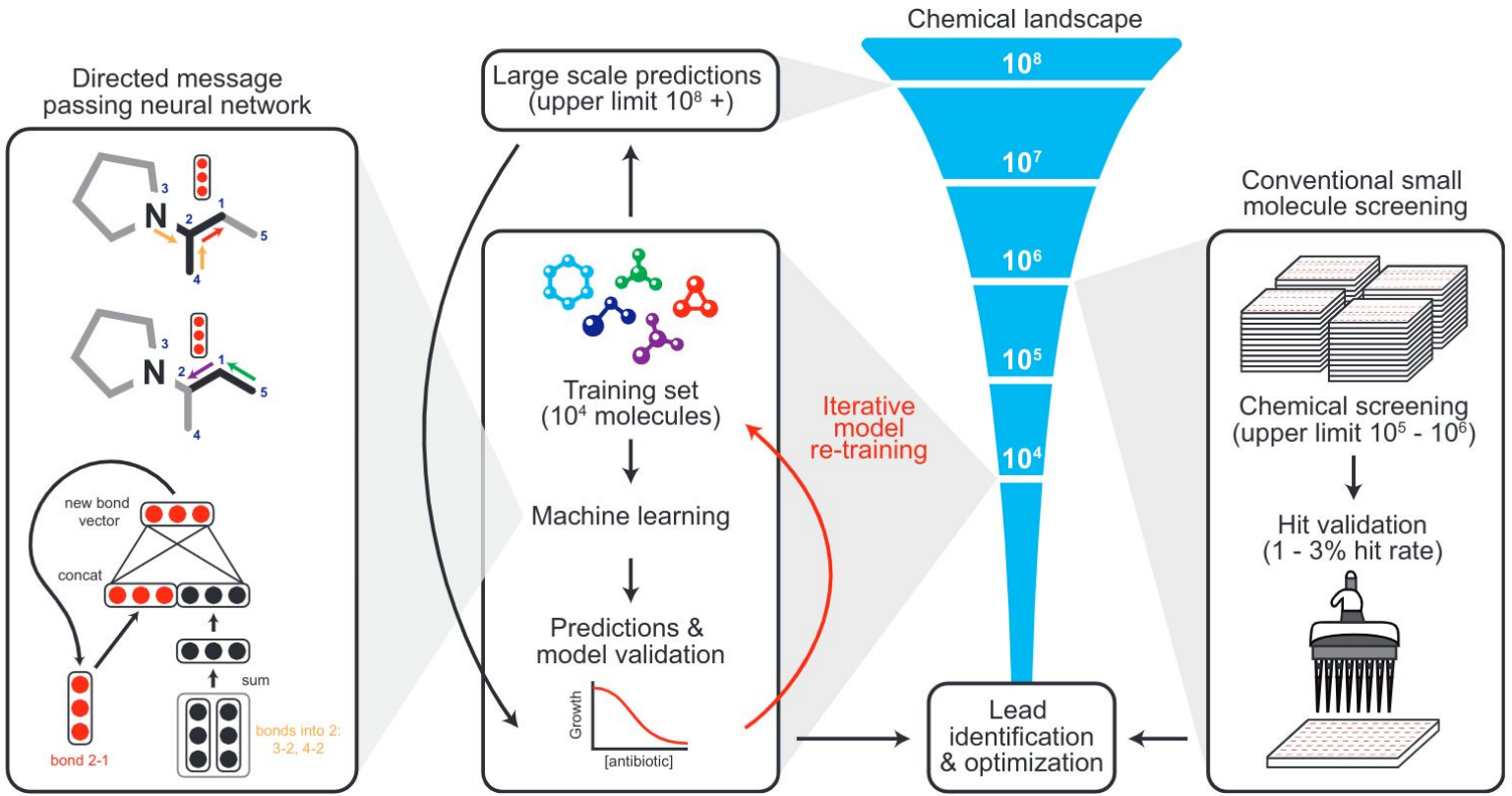
A new HPC+DL+Physical paradigm





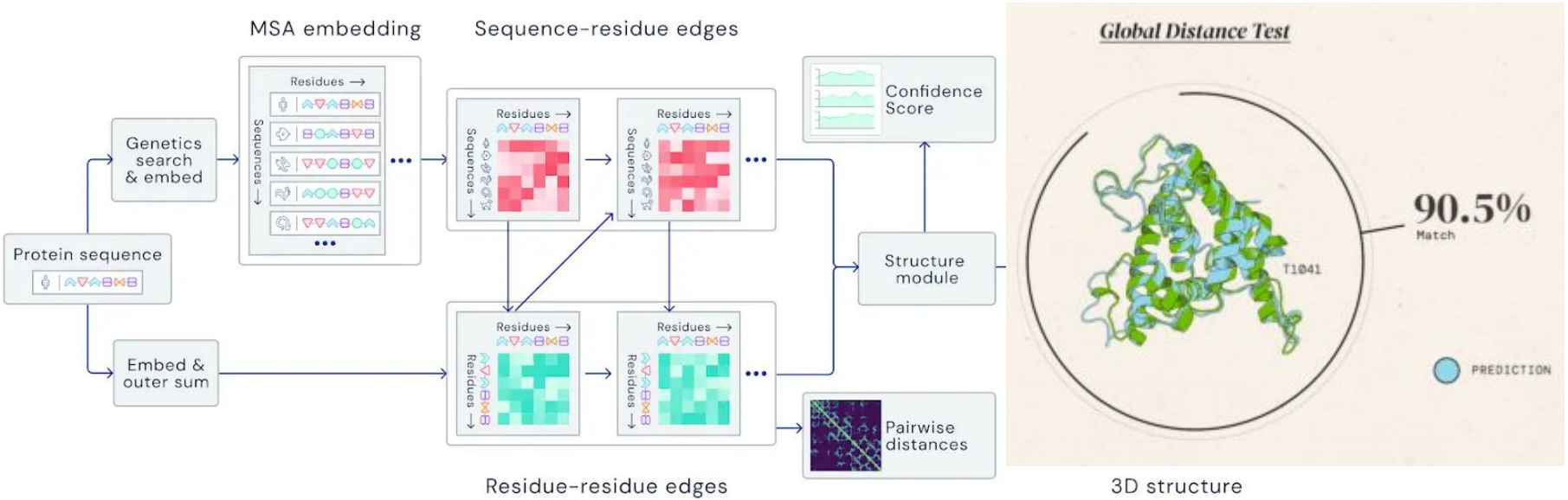
Pictures were from their SC'20 talk.

A deep learning approach to antibiotic discovery

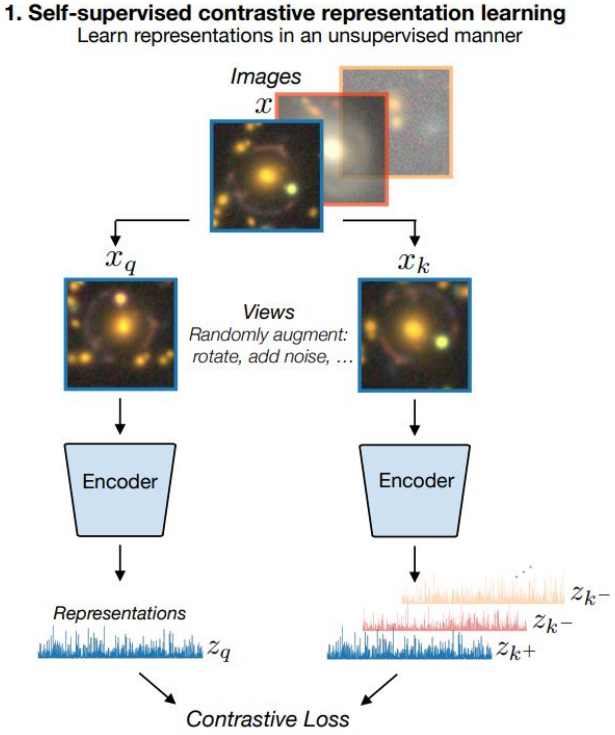


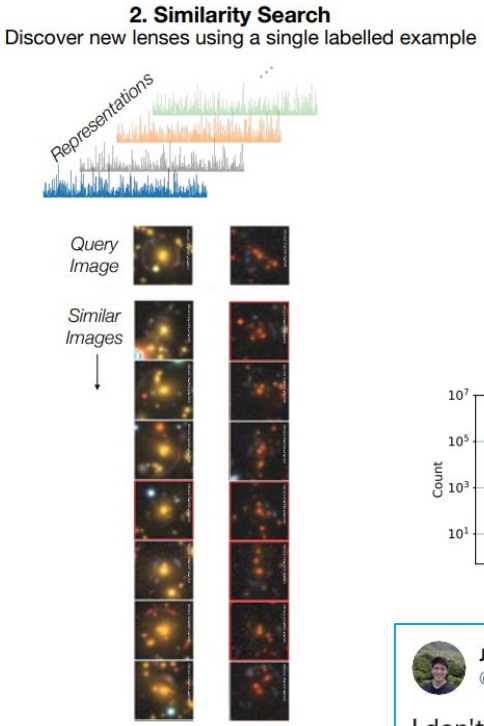
Qiyang Hu

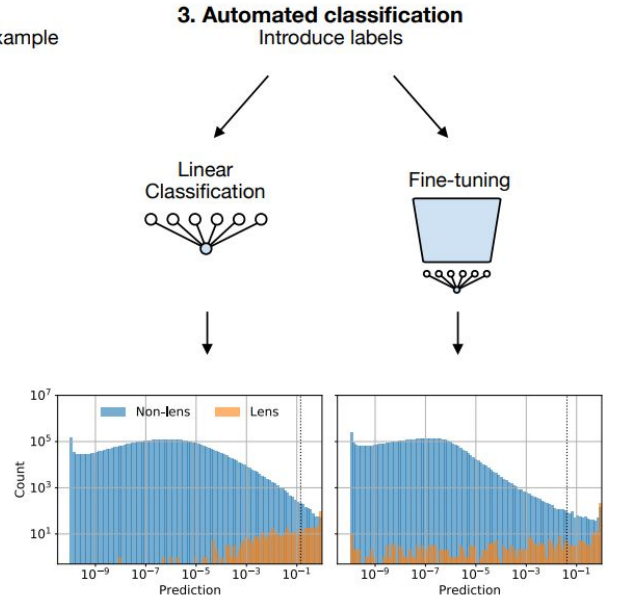
AlphaFold: a significant breakthrough of AI on Science



Astronomical discoveries through self-supervised learning







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

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).

Qiyang Hu

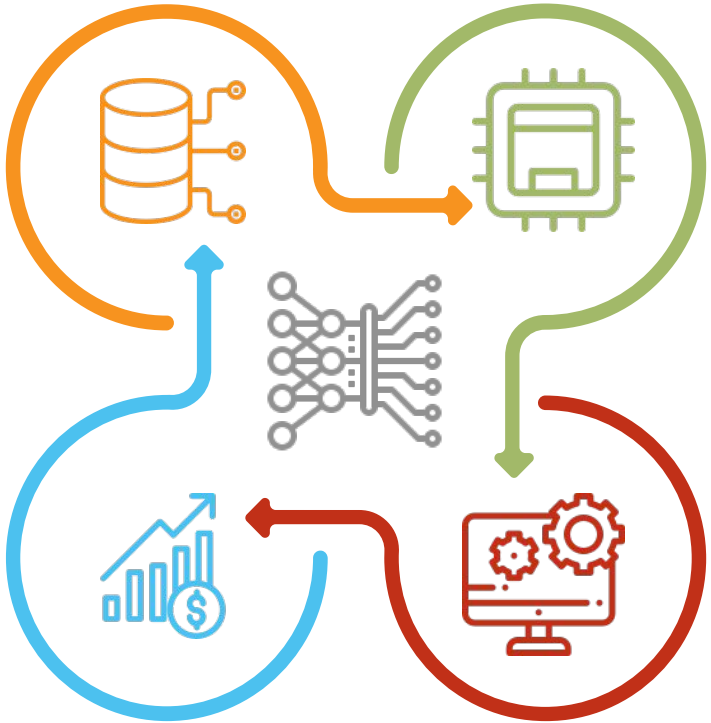
Driving forces and positive impacts

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, Neuromophic 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

Neutral

General Intro

- What's ANN?
- DL, ML and AI

01

Good

DL as hopes

- Why DL is so powerful?
- Exciting achievements
- Mutual driving forces

02

Bad

DL as hypes

- Difficulties in application perspective
- Illusions in theoretical perspective

03

"Ugly"

DL like deep oceans

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

04

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

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.

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?

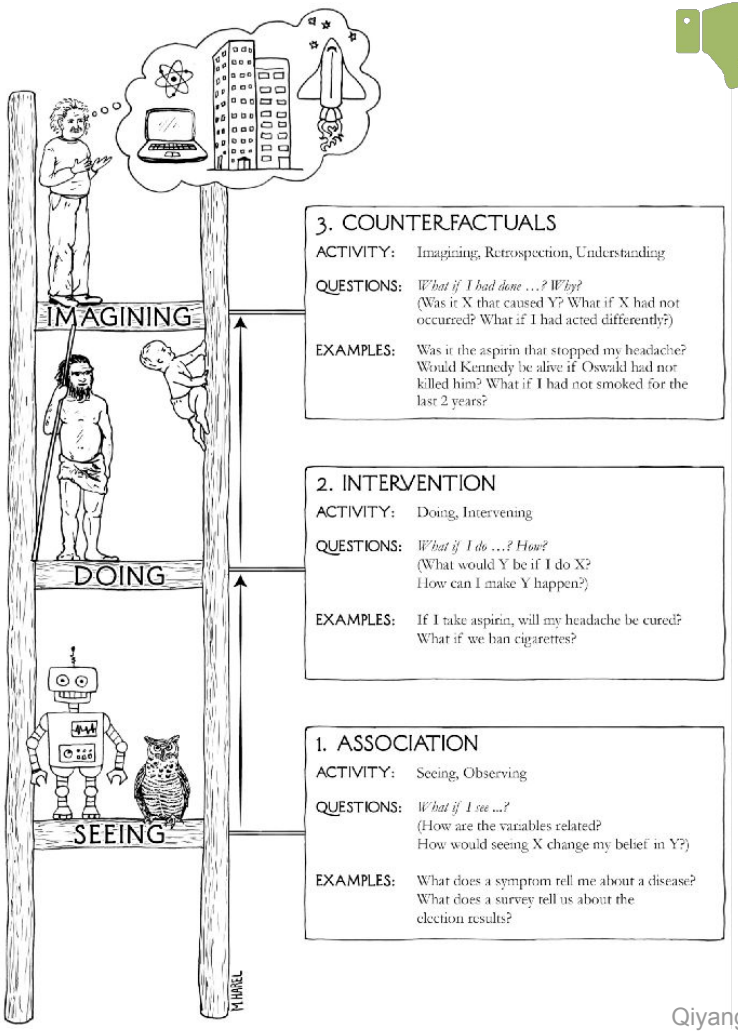
"...none of them [papers] produced tools that would be good enough to use in a clinical setting. Something has gone seriously wrong when more than 300 papers are published that have no practical benefit."

—— "Nature Machine Intelligence" 2021

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

Is it the real intelligence?

- Data mining ⇒ understanding?
  - Neural network = Curve fitting
  - Statistical = model-blind
- DL need something else to work together.
  - Causal reasoning: inductive -> deductive ?
  - Hybrid system with GOFAI? with Physics?
- 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



Neutral

General Intro

- What's ANN?
- DL, ML and AI

01

Good

DL as hopes

- Why DL is so powerful?
- Exciting achievements
- Mutual driving forces

02

Bad

DL as hypes

- Difficulties in application perspective
- Illusions in theoretical perspective

03

"Ugly"

DL like deep oceans

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

04

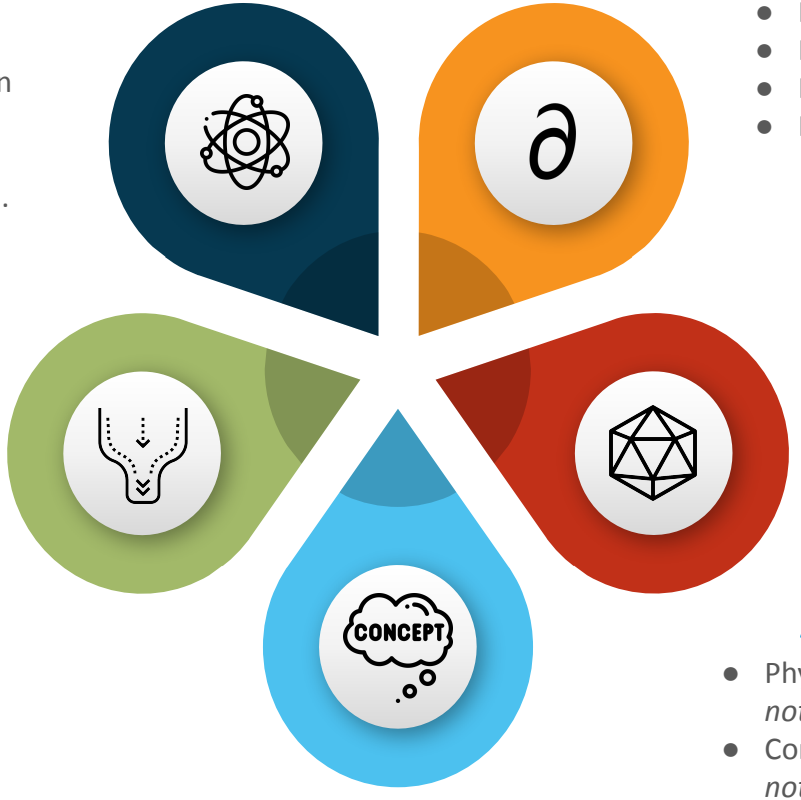
Understanding Neural Networks

Effective Theory

- NN ⇒ 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

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



Discretized Dynamics

- Deep Network ⇒ Diff Eqns (DE)
- Network Architecture ⇒ Numerical DE
- Network Training ⇒ Optimal Control
- ResNet ⇒ One type of discretization
- Forward & Backprop ⇒ ODE solvers

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

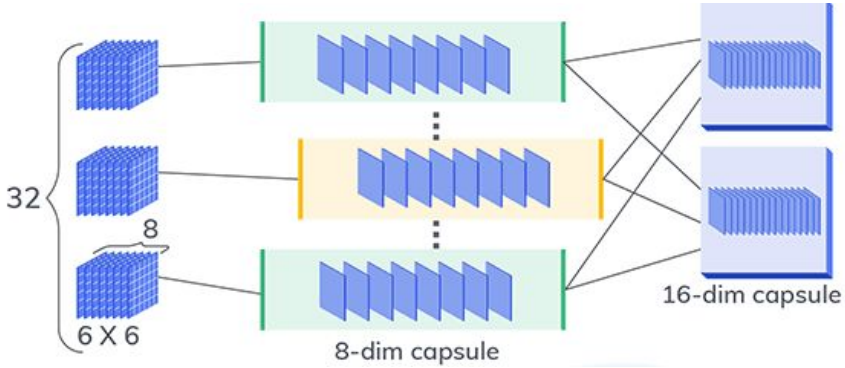
Analytical Philosophy

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

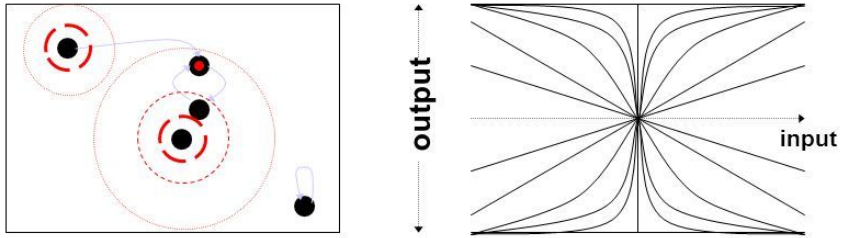
Qiyang Hu

Alternative Neural Networks

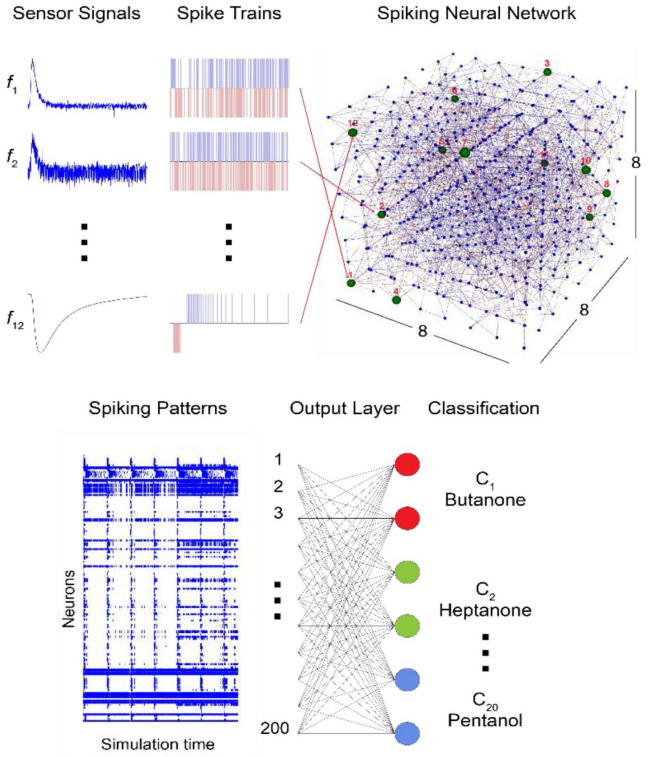
Capsule Network

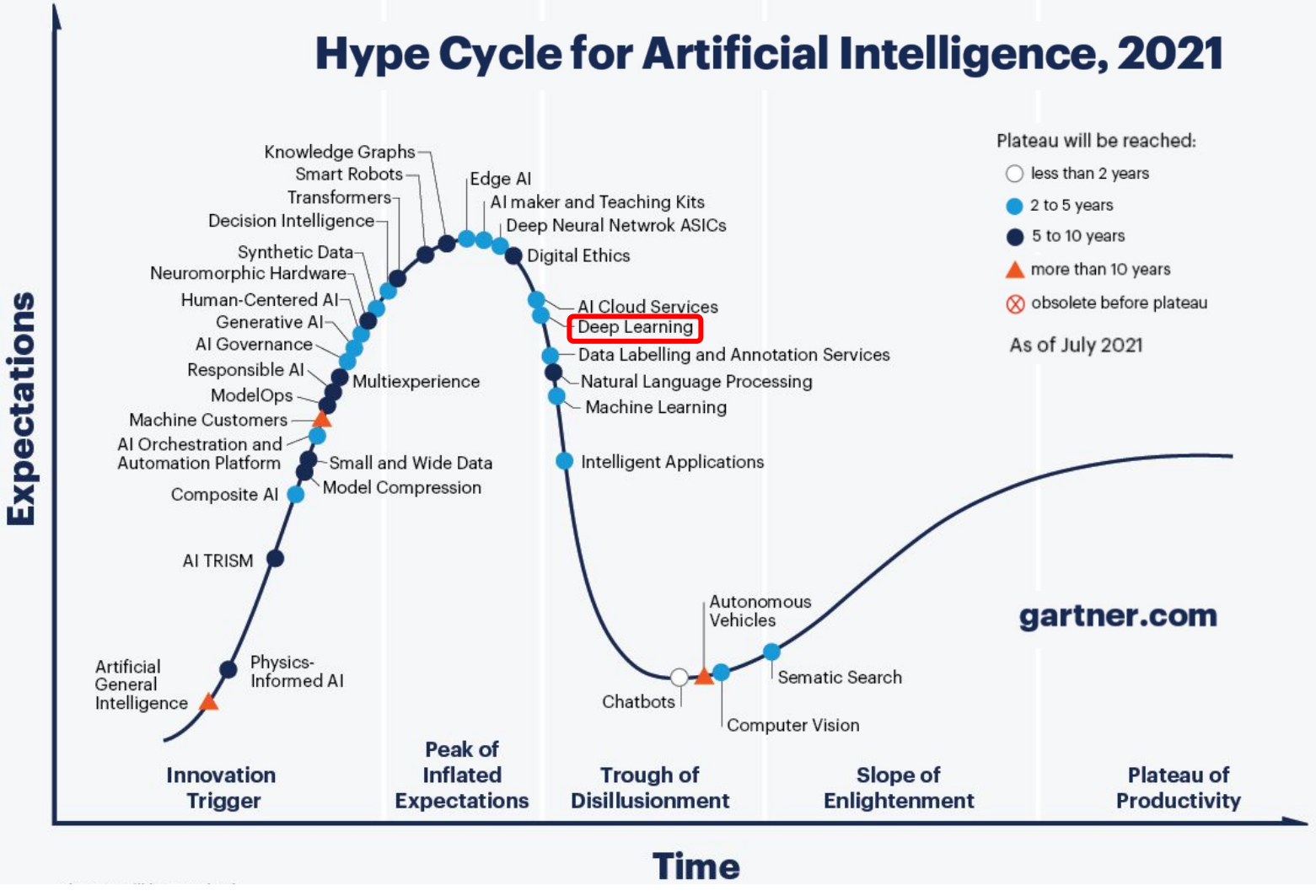


GasNets



Spiking Neural Network




Hype Cycle for Artificial Intelligence, 2021
Expectations
Time
Knowledge Graphs
Smart Robots
Transformers
Decision Intelligence
Synthetic Data
Neuromorphic Hardware
Human-Centered AI
Generative AI
AI Governance
Responsible AI
ModelOps
Machine Customers
AI Orchestration and Automation Platform
Composite AI
AI TRISM
Artificial General Intelligence
Physics-Informed AI
Multiexperience
Small and Wide Data
Model Compression
Edge AI
AI maker and Teaching Kits
Deep Neural Netwrok ASICs
Digital Ethics
AI Cloud Services
Deep Learning
Data Labelling and Annotation Services
Natural Language Processing
Machine Learning
Intelligent Applications
Chatbots
Autonomous Vehicles
Computer Vision
Sematic Search
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
Plateau will be reached:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau
As of July 2021
gartner.com

Qiyang Hu

OARC Workshop Survey

https://forms.gle/nbWgNP45qCwZhLRh9