# Deep Learning for Particle Physicists

Lewis Tunstall | AEC Graduate Seminar | April 29th 2022

# 🧹 Housekeeping

📅 **Timetable**
14:15-16:00 every Friday in Room 119 from April 29 to June 3 (+1 extra slot to make up for first lecture)

🚀 **Course website**
https://lewtun.github.io/dl4phys/intro.html

✉️ **Contact details**
lewis.c.tunstall@gmail.com

🤗 **Hugging Face Hub organisation**
https://huggingface.co/dl4phys

💻 **Free GPU platforms:**

- Google Colab (OK, but small GPUs): https://colab.research.google.com/
- Kaggle notebooks (P100 GPUs 💪): https://www.kaggle.com/code
- Paperspace notebooks: https://gradient.run/notebooks
- SageMaker Studio Lab: https://studiolab.sagemaker.aws/

# About me

PhD @ University of Adelaide

↓

Postdoc @ ITP Bern

↓

Stream processing @ SPOUD

↓

Geometric ML @ L2F

↓

NLP @ Hugging Face

# About me

PhD @ University of Adelaide

↓

Postdoc @ ITP Bern

↓

Stream processing @ SPOUD

↓

Geometric ML @ L2F

↓

NLP @ Hugging Face

ChPT, CFT,
Kaon physics

# About me

PhD @ University of Adelaide

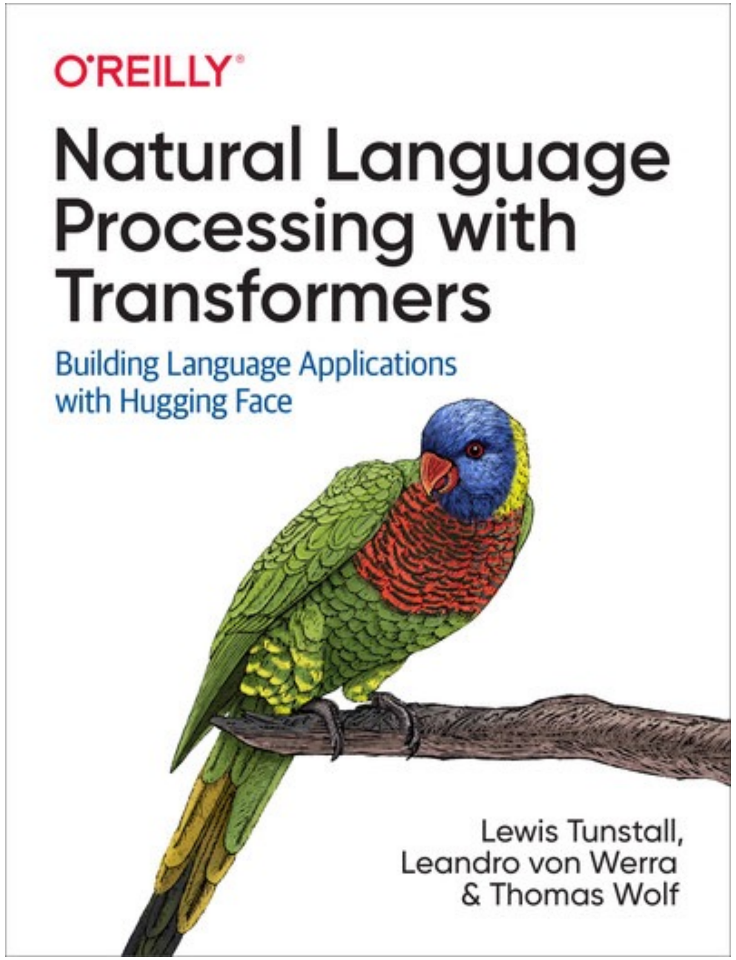
Postdoc @ ITP Bern

Stream processing @ SPOUD

Geometric ML @ L2F

NLP @ Hugging Face

ChPT, CFT,
Kaon physics




transformersbook.com

Jet tagging with DNNs

DNN deep dive

Classifying jet images with CNNs

CNN deep dive

👷 foundations

Transformers

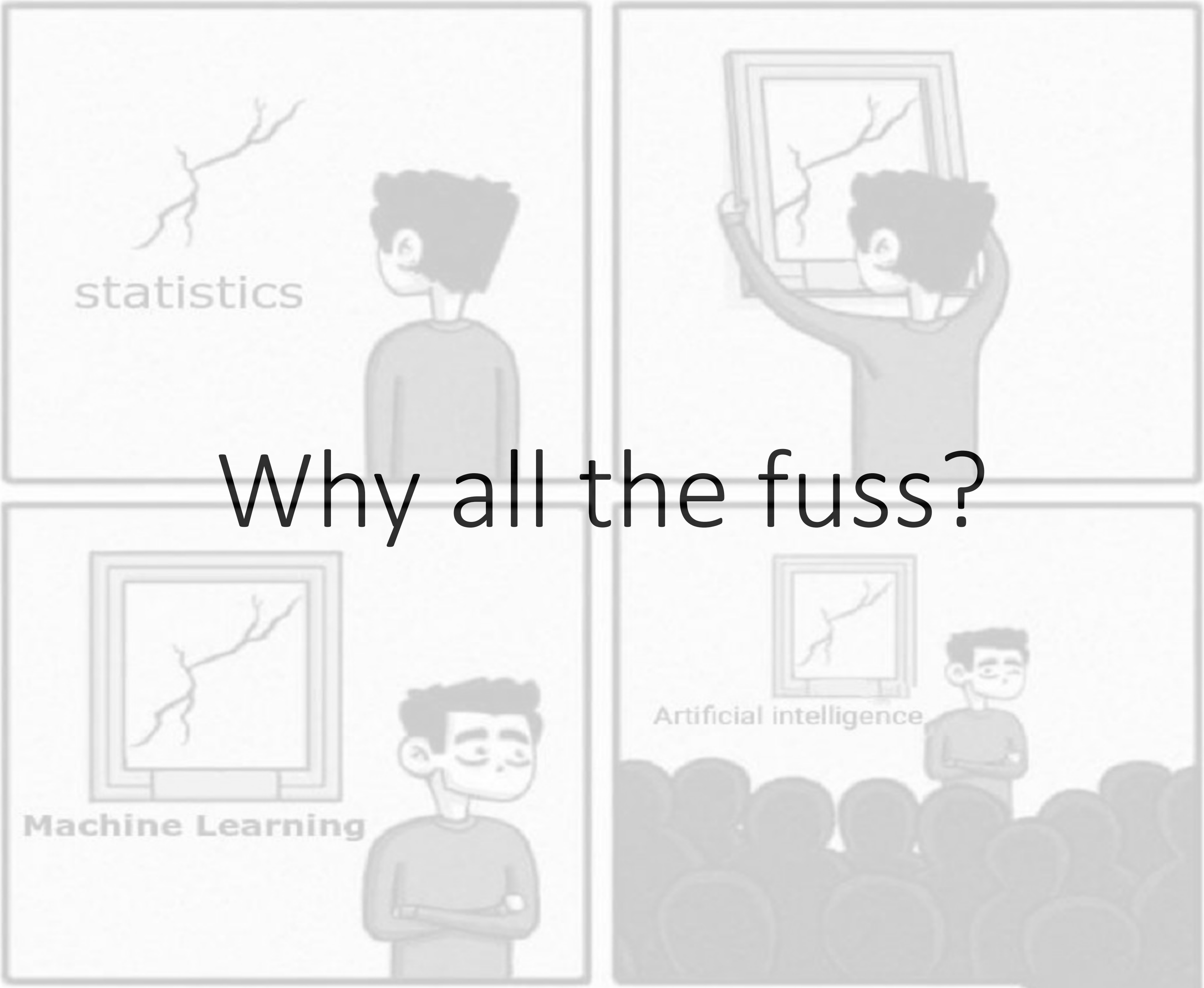
Normalising flows

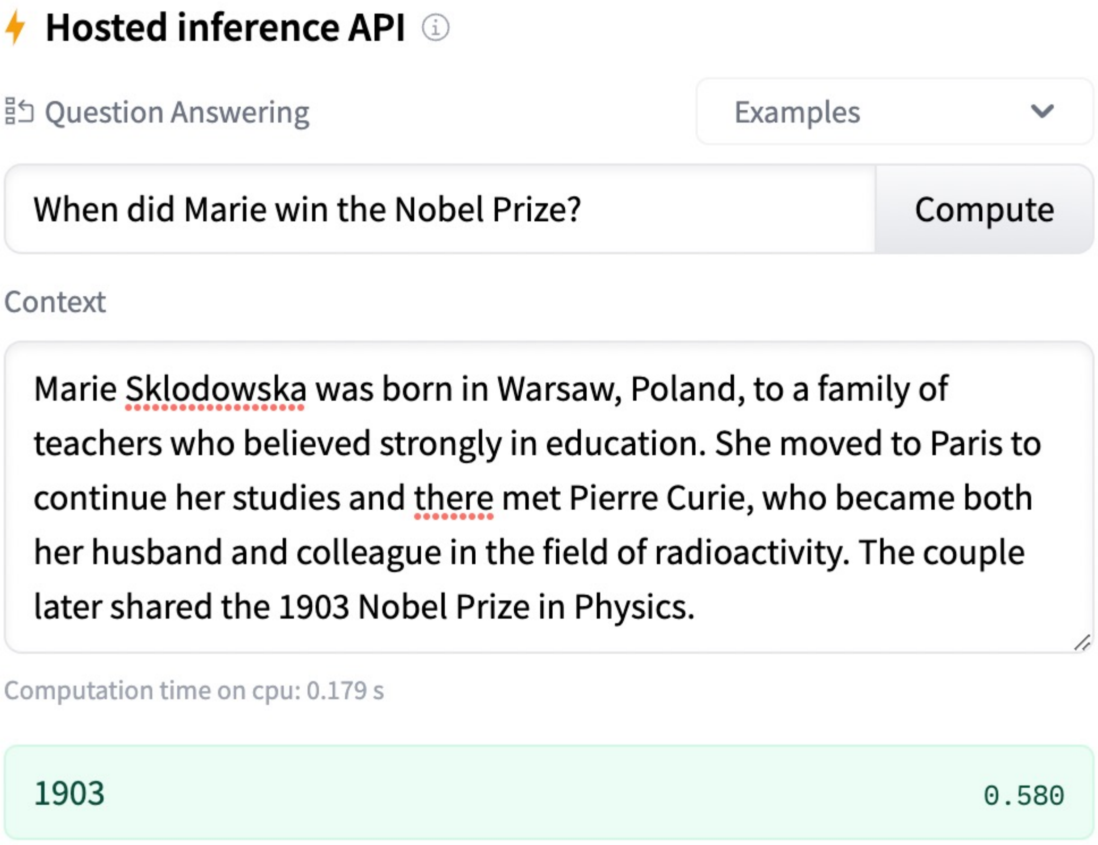
Symmetries & Lagrangian NNs

😎 fancy stuff

# What is this course about?

# Why all the fuss?

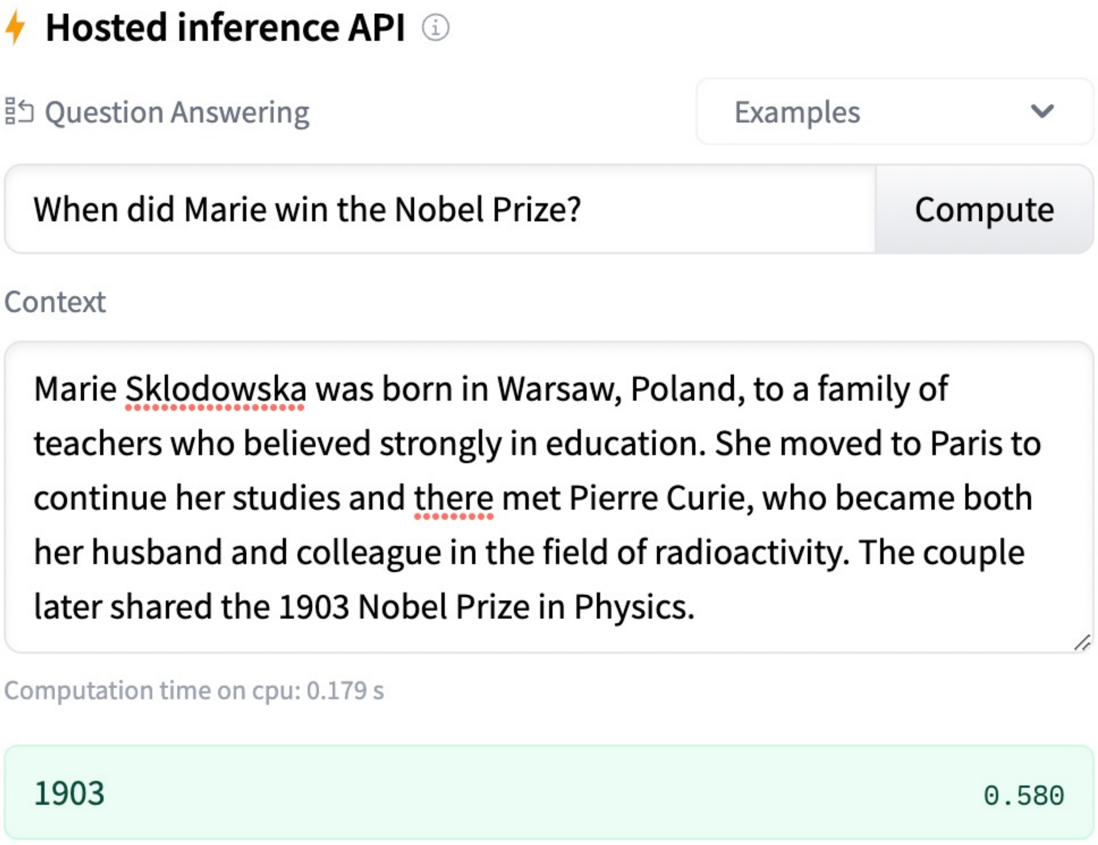




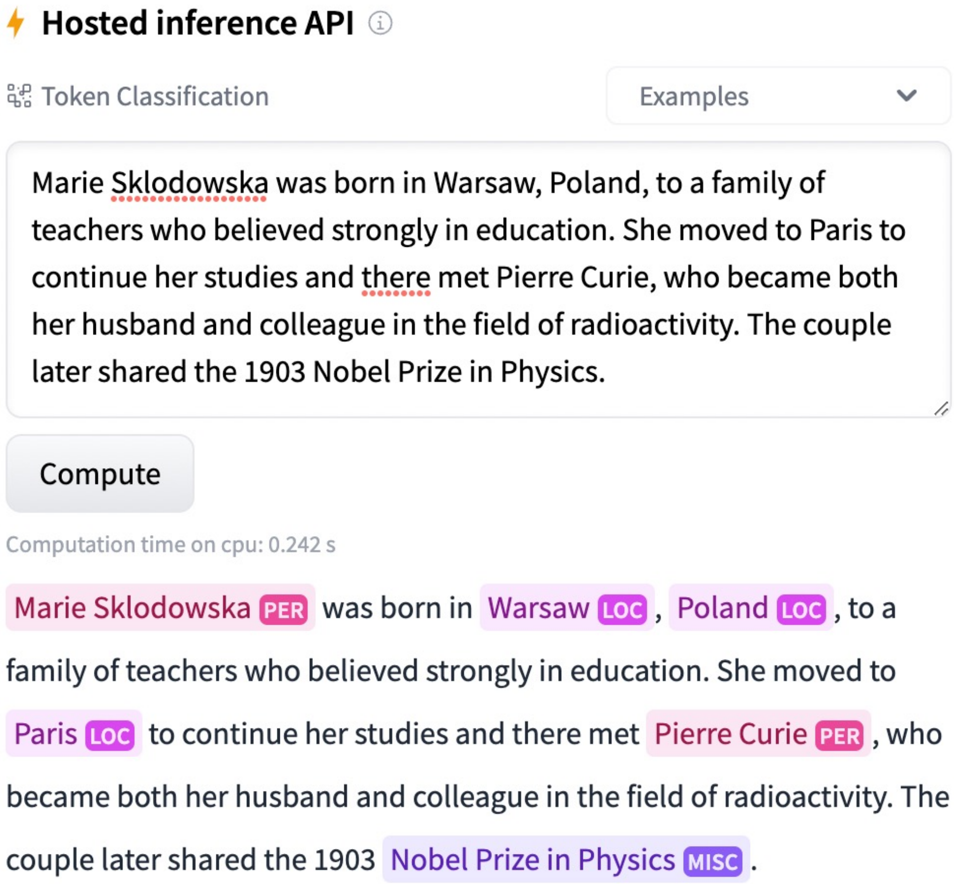

hf.co/deepset/roberta-base-squad2

hf.co/xlm-roberta-large-finetuned-conll03-english

# Deep learning is currently the best technique for analysing text




hf.co/deepset/roberta-base-squad2

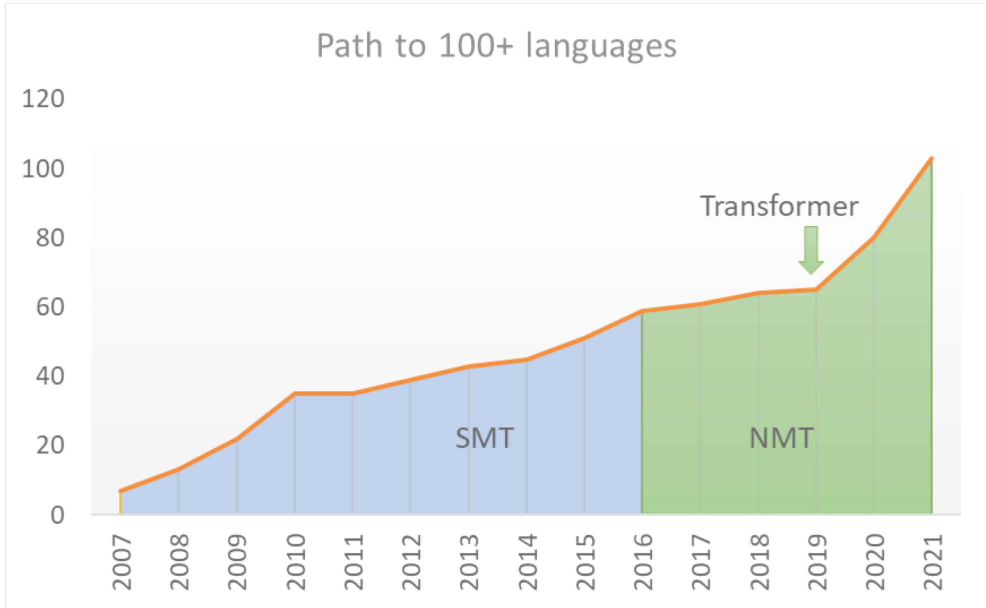



hf.co/xlm-roberta-large-finetuned-conll03-english

# Deep learning is currently the best technique for analysing text

## Microsoft Translator now works across 103 languages

Microsoft adds 12 languages to its Microsoft Translate app that can help 84.6 million people.

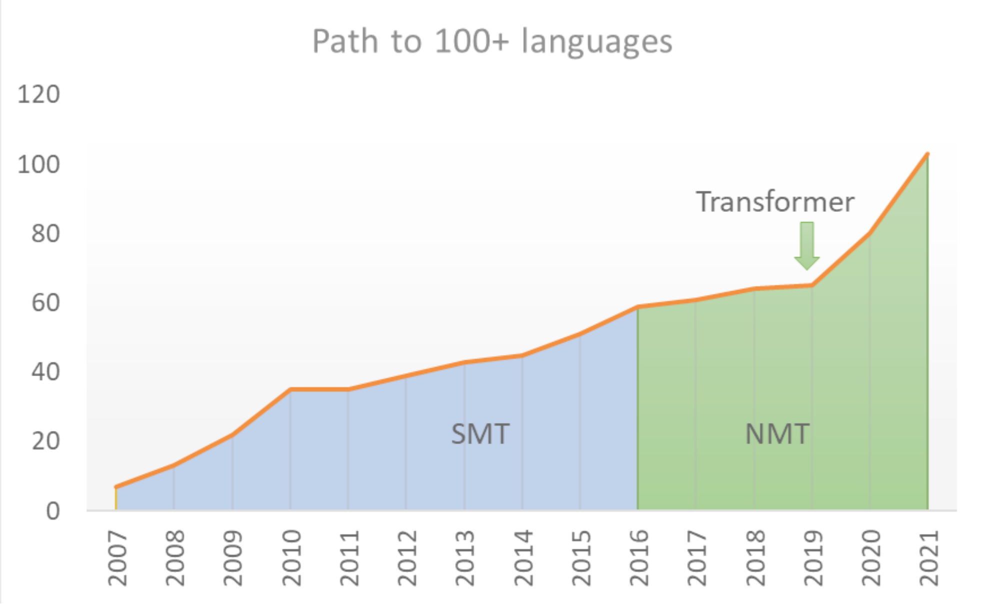



Deep learning is currently the best technique for analysing text

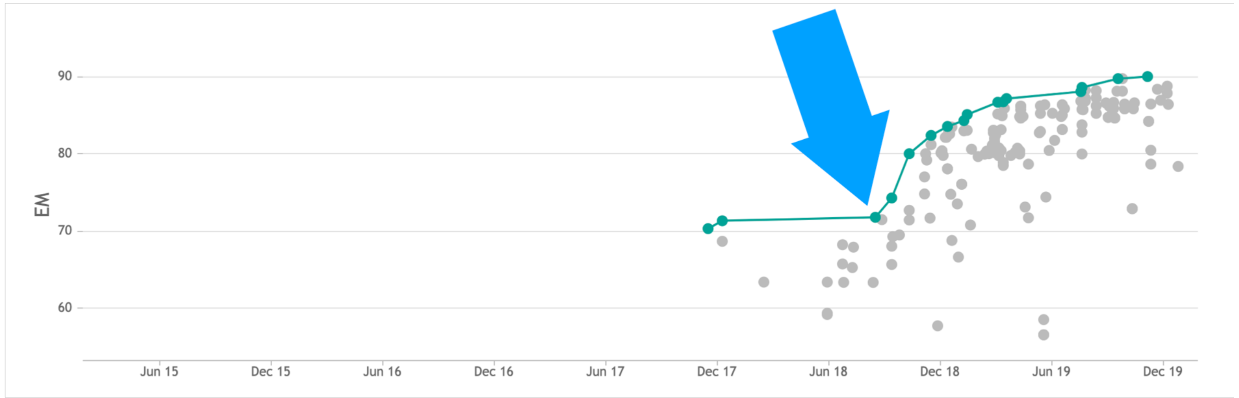
Microsoft

## Microsoft Translator now works across 103 languages

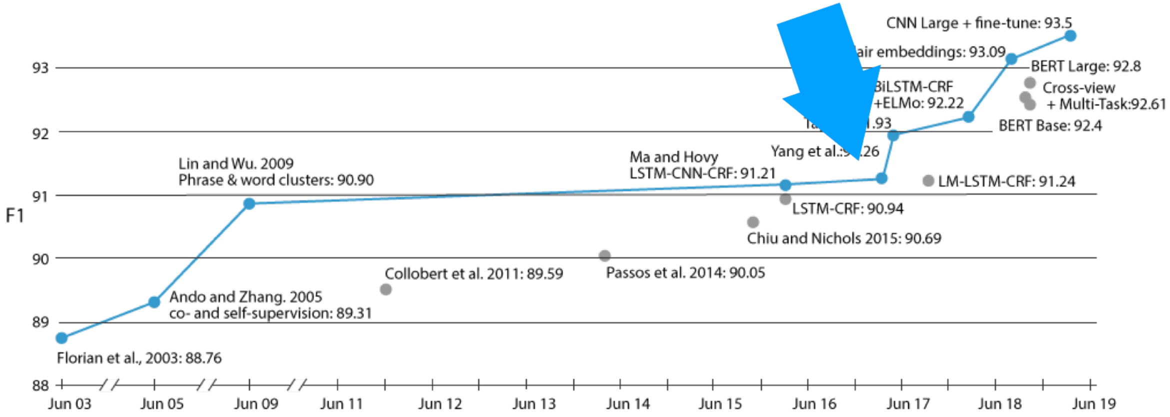
Microsoft adds 12 languages to its Microsoft Translate app that can help 84.6 million people.




Performance on Question Answering benchmark (SQUAD 2.0)




Performance on Named Entity Recognition benchmark (CoNLL)




# Deep learning is currently the best technique for analysing text

**Hosted inference API**

Image Classification

Computation time on cpu: 0.199 s

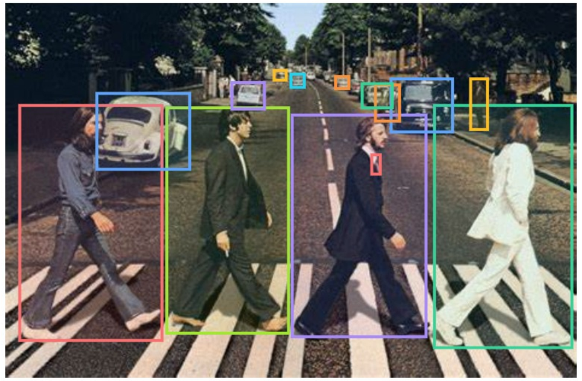
| | |
|---|---|
| Eskimo dog, husky | 0.403 |
| Siberian husky | 0.316 |
| Norwegian elkhound, elkhound | 0.057 |
| dingo, warrigal, warragal, Canis dingo | 0.052 |
| malamute, malemute, Alaskan malamute | 0.009 |

hf.co/microsoft/beit-base-patch16-224

**Hosted inference API**

Object Detection

Examples

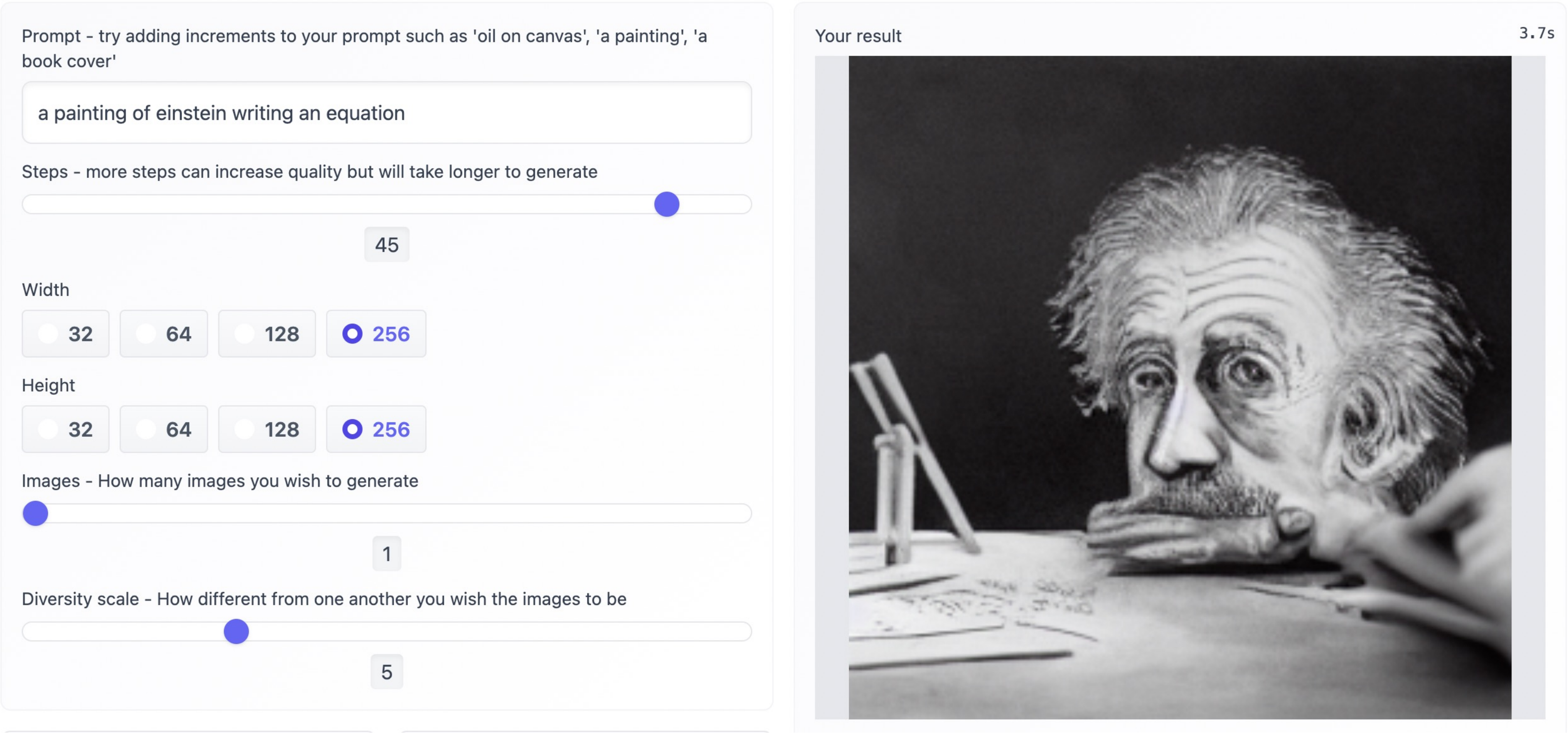


Computation time on cpu: 1.295 s

| | |
|---|---|
| tie | 0.955 |
| person | 0.999 |
| person | 0.971 |
| car | 0.998 |
| car | 0.904 |

hf.co/facebook/detr-resnet-50

# Deep learning is currently the best technique for analysing images




hf.co/spaces/multimodalart/latentdiffusion

# Deep learning is currently the best technique for analysing text and images

TEXT DESCRIPTION

An astronaut Teddy bears A bowl of soup

that is a portal to another dimension that looks like a monster as a planet in the universe

knitted out of wool spray-painted on a wall made out of plasticine

→

DALL·E 2

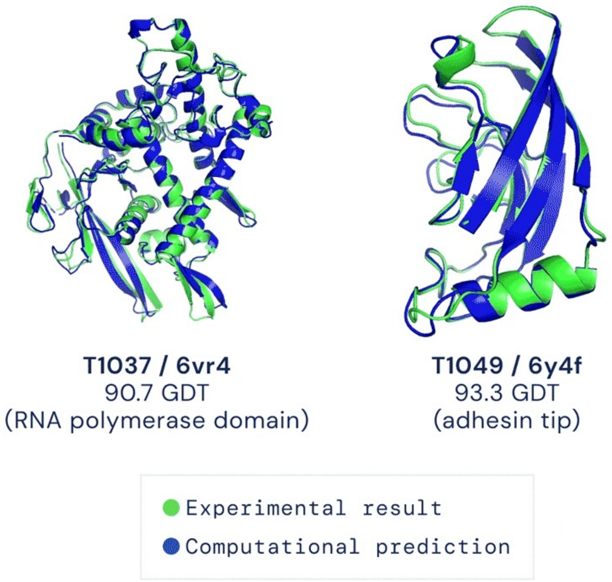
openai.com/dall-e-2/

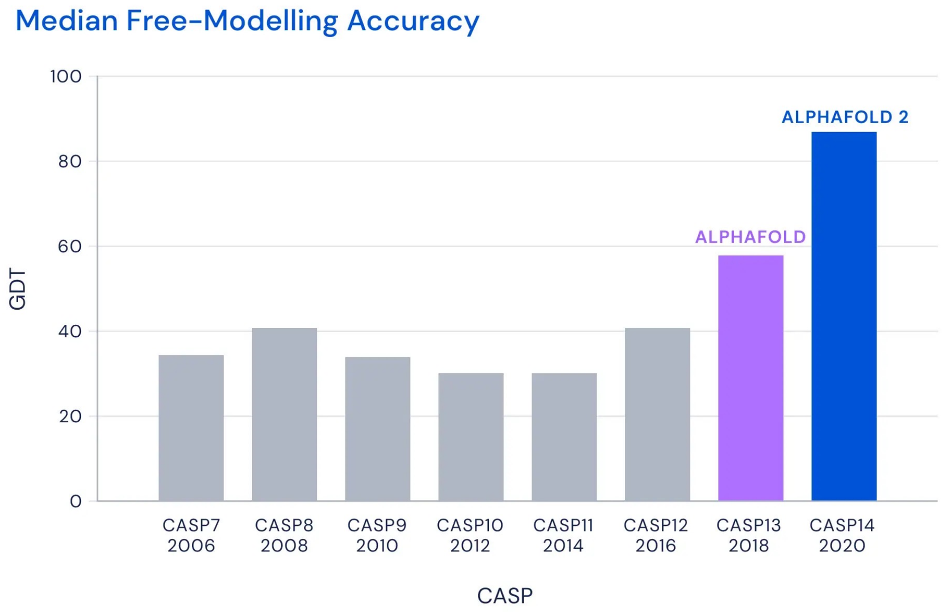
# Deep learning is currently the best technique for analysing text and images

NEWS | 22 July 2021

# DeepMind's AI predicts structures for a vast trove of proteins

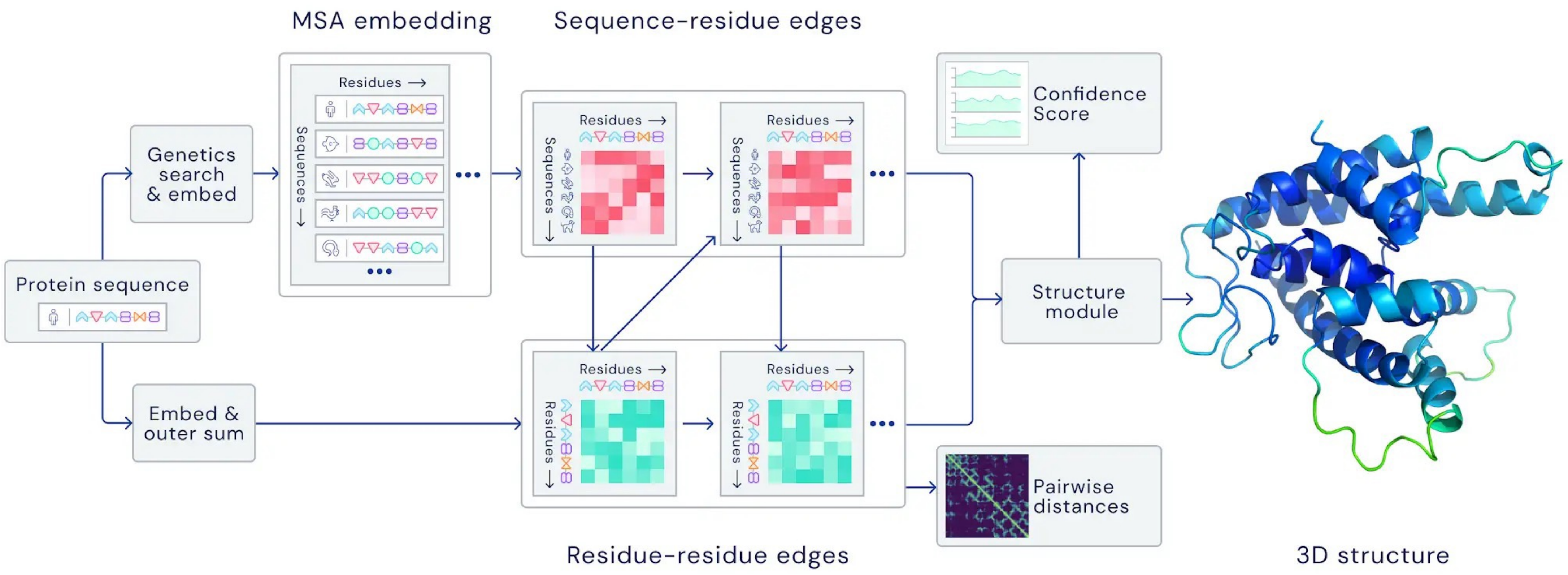
AlphaFold neural network produced a 'totally transformative' database of more than 350,000 structures from *Homo sapiens* and 20 model organisms.




https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

# And it's not just text and images where deep learning shines

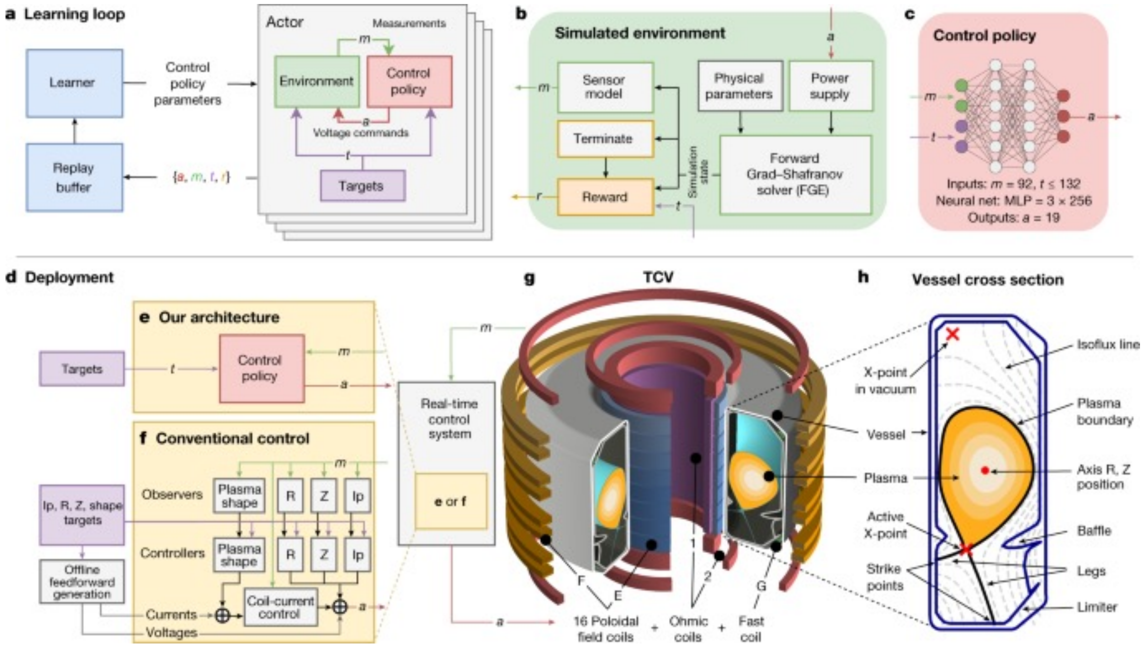


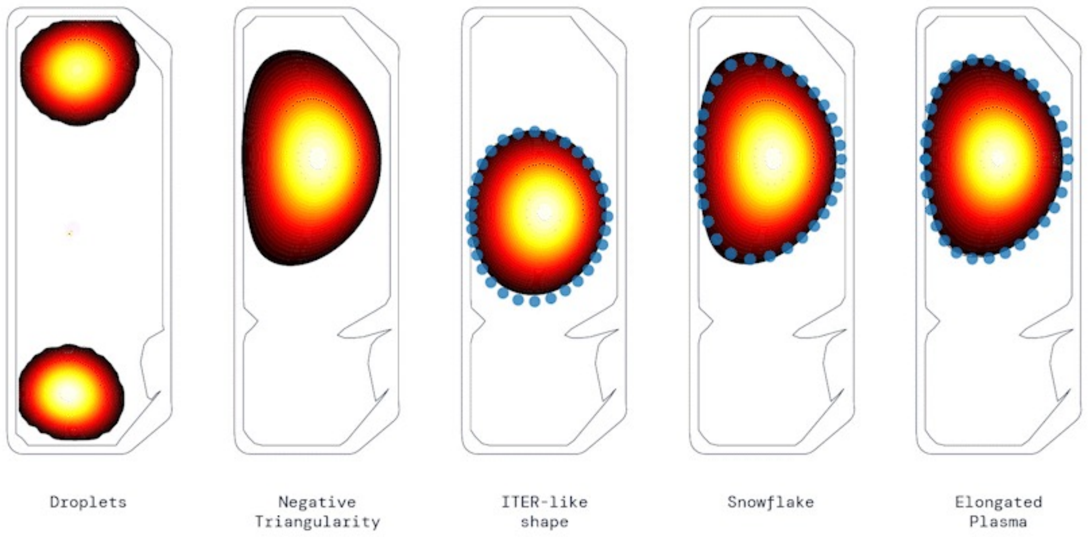

https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

# And it’s not just text and images where deep learning shines

# DeepMind Has Trained an AI to Control Nuclear Fusion

**The Google-backed firm taught a reinforcement learning algorithm to control the fiery plasma inside a tokamak nuclear fusion reactor.**




www.nature.com/articles/s41586-021-04301-9

# And it’s not just text and images where deep learning shines

Deep
neural networks

Data availability
(text, images, etc)

Computer hardware
(GPU, TPU, etc)

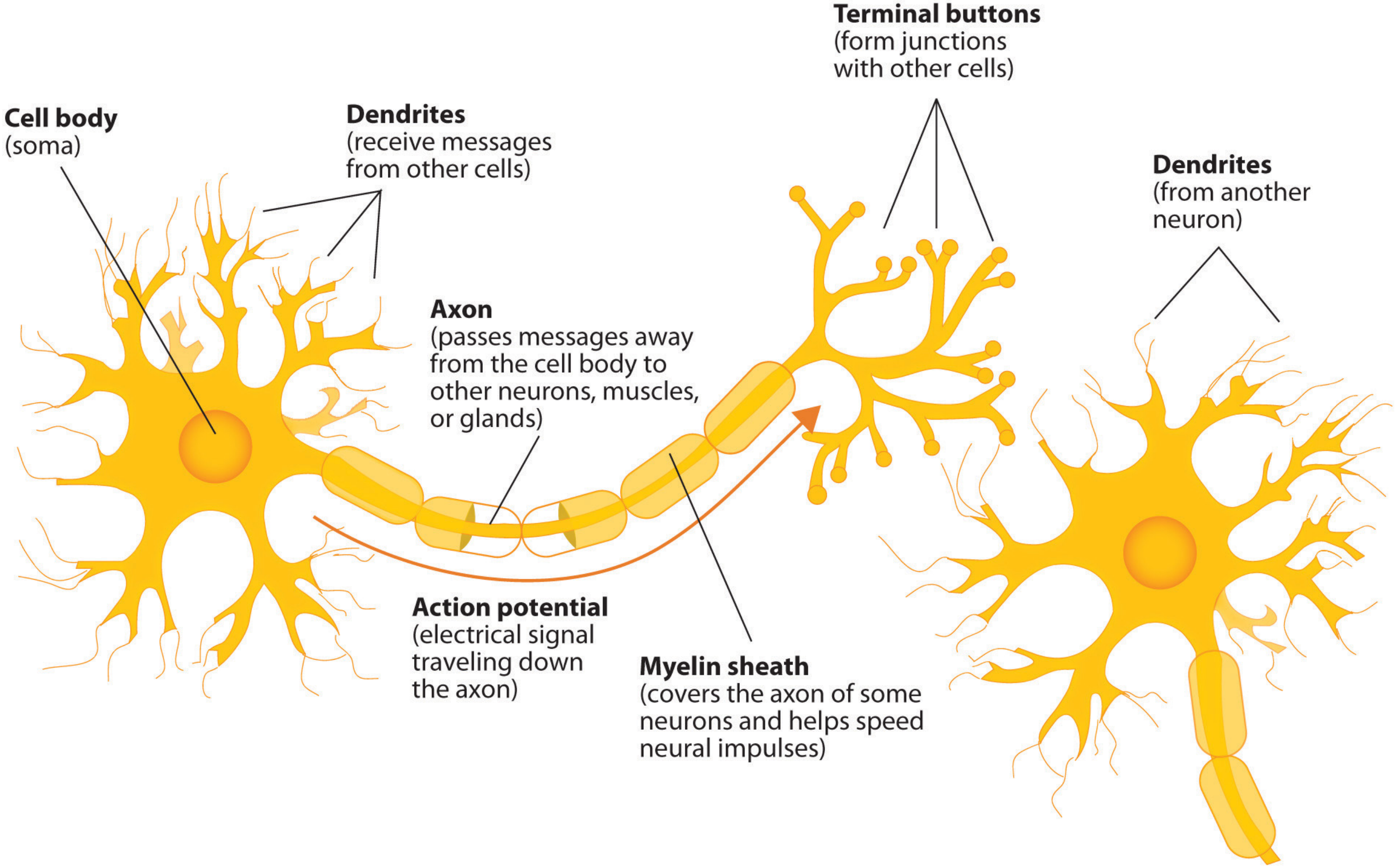
# Main ingredients

Deep
neural networks

Data availability
(text, images, etc)

Computer hardware
(GPU, TPU, etc)

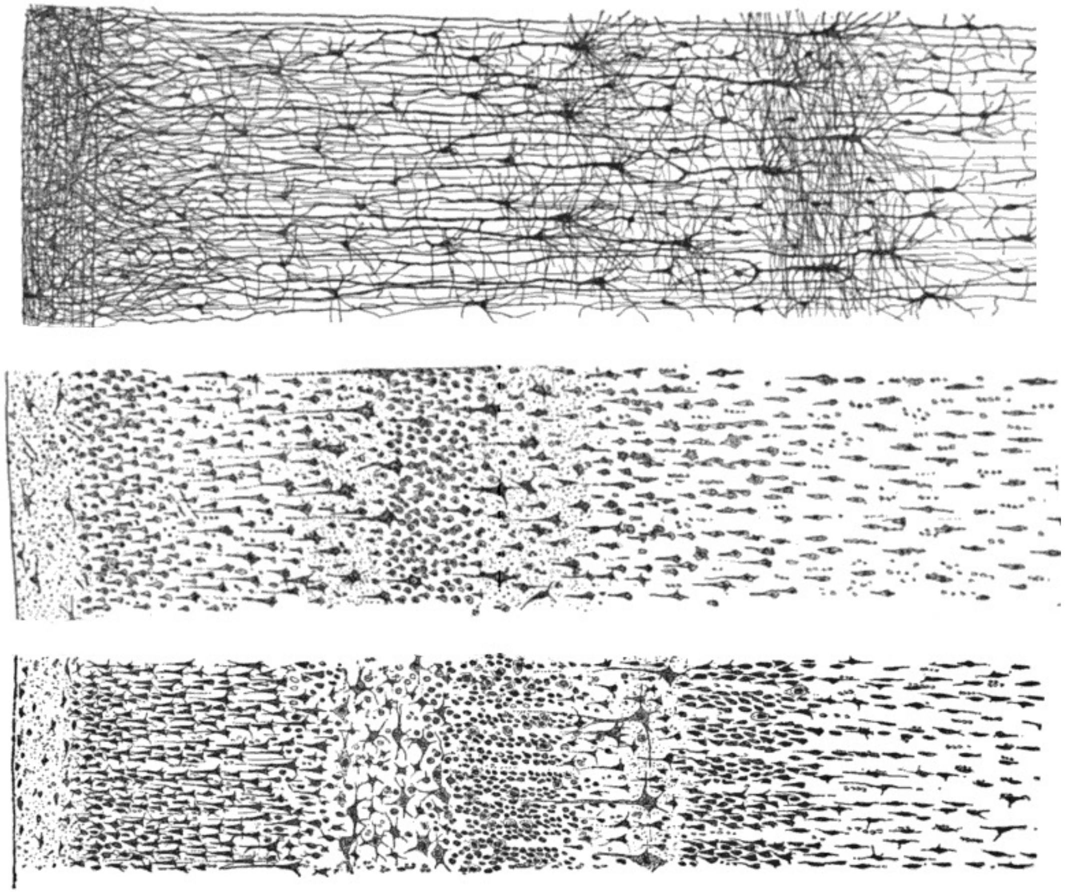
# Main ingredients




en.wikipedia.org/wiki/Neuron

# From biological to artificial neurons

neurons often organized in consecutive *layers*

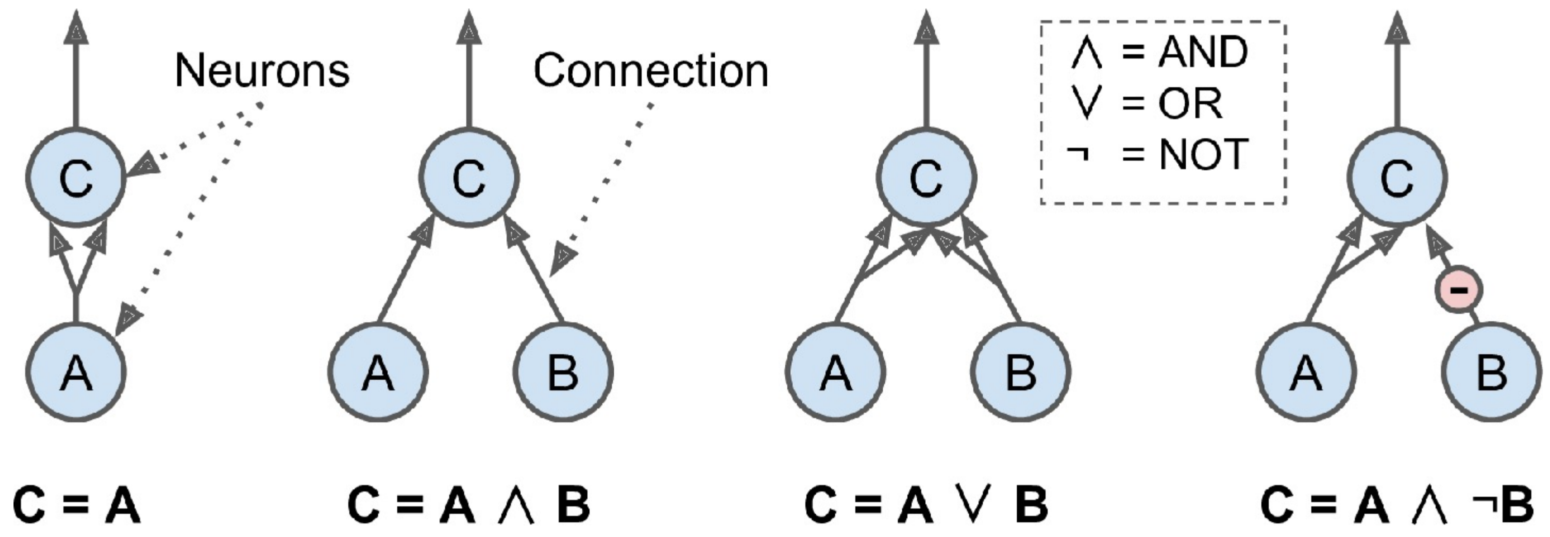


en.wikipedia.org/wiki/Cerebral_cortex

# From biological to artificial neurons

# McCulloch-Pitts model (1943)

*binary* output




*binary* inputs

Hands-On Machine Learning with Scikit-Learn & TensorFlow, A. Geron

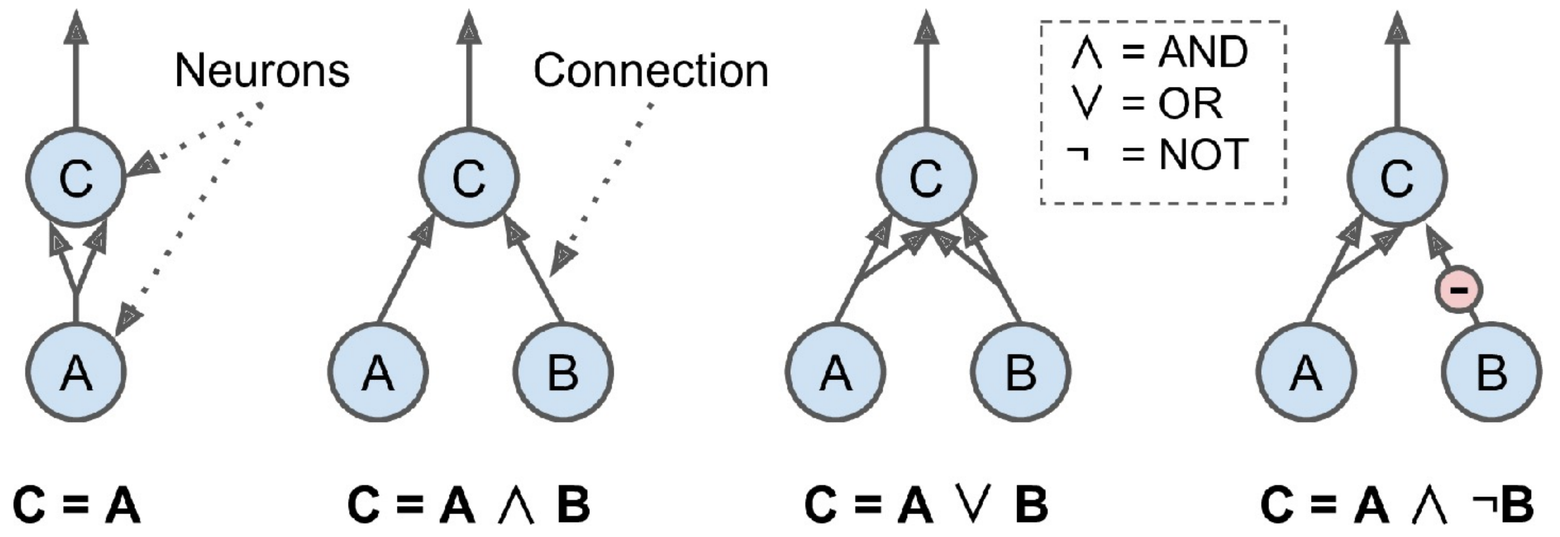
# From biological to artificial neurons

# McCulloch-Pitts model (1943)

*binary* output

neuron *activated* when more than some number of inputs are active

*binary* inputs

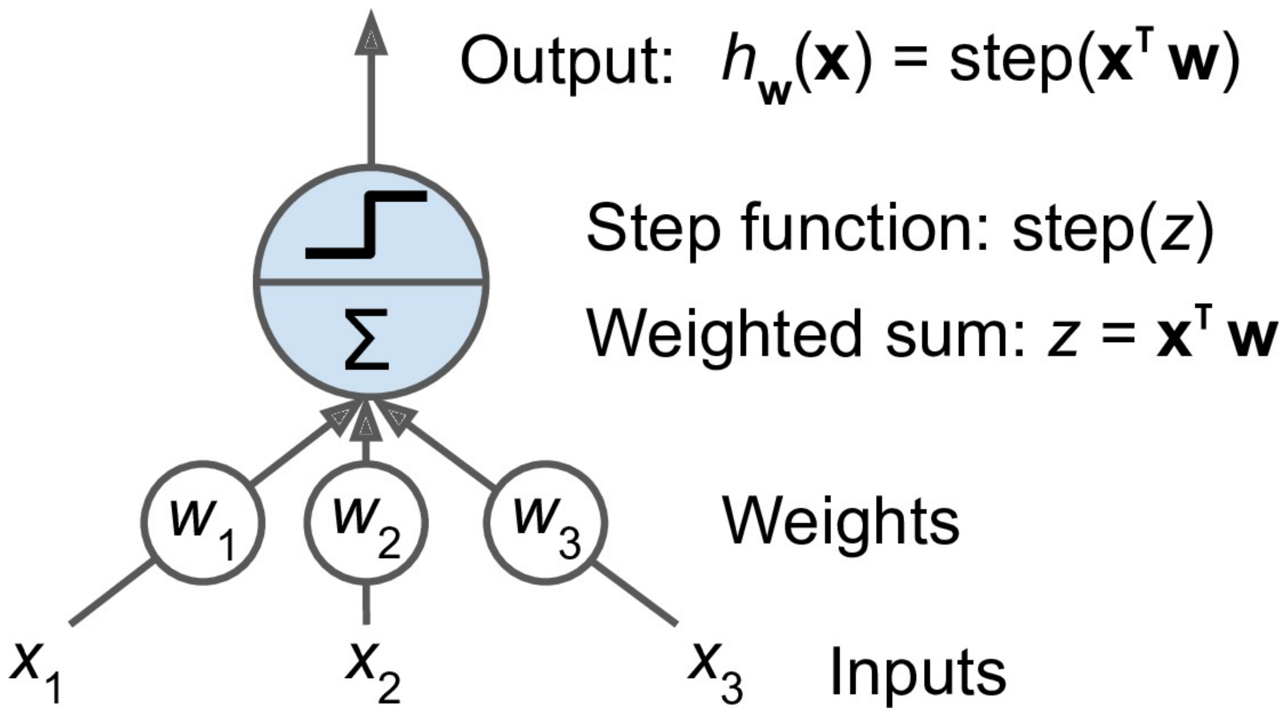



Hands-On Machine Learning with Scikit-Learn & TensorFlow, A. Geron

# From biological to artificial neurons

# The Perceptron (1957)

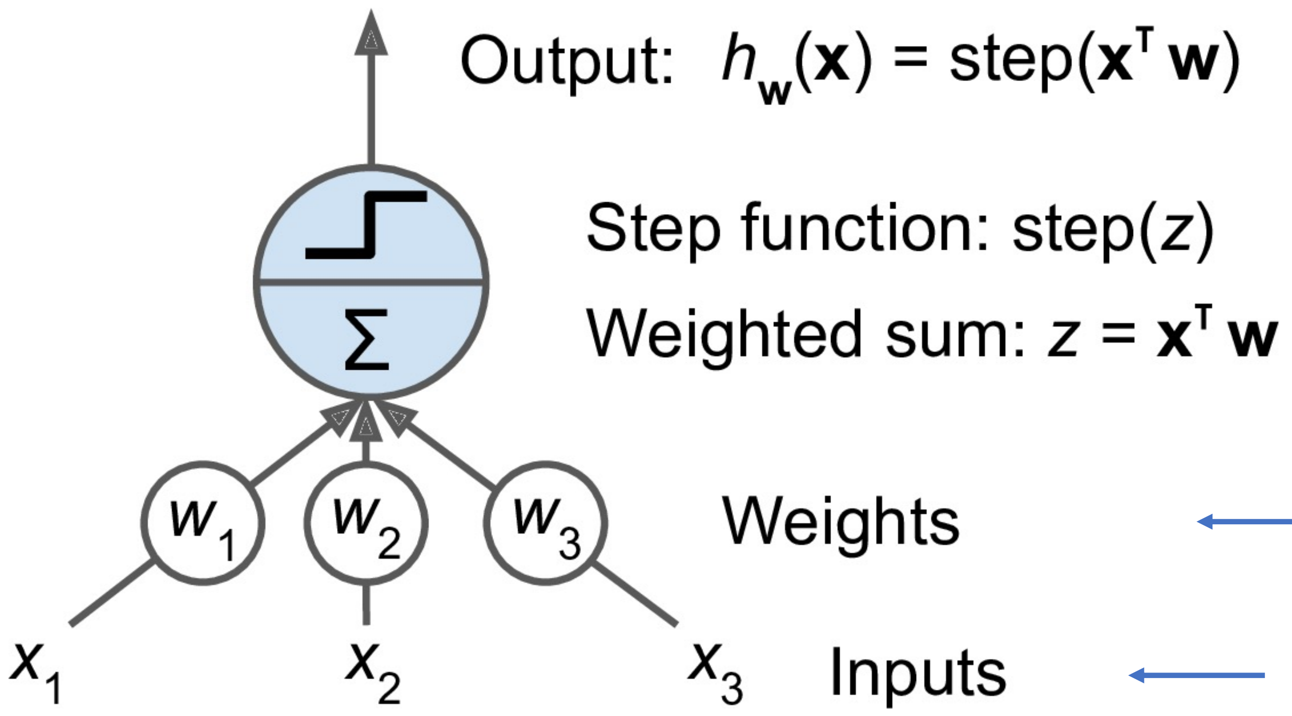
neuron called an
threshold logic unit
(TLU)




Hands-On Machine Learning with Scikit-Learn & TensorFlow, A. Geron

# From biological to artificial neurons

# The Perceptron (1957)




output now *continuous* number

each input associated with a *weight*

inputs now *continuous* numbers

Hands-On Machine Learning with Scikit-Learn & TensorFlow, A. Geron

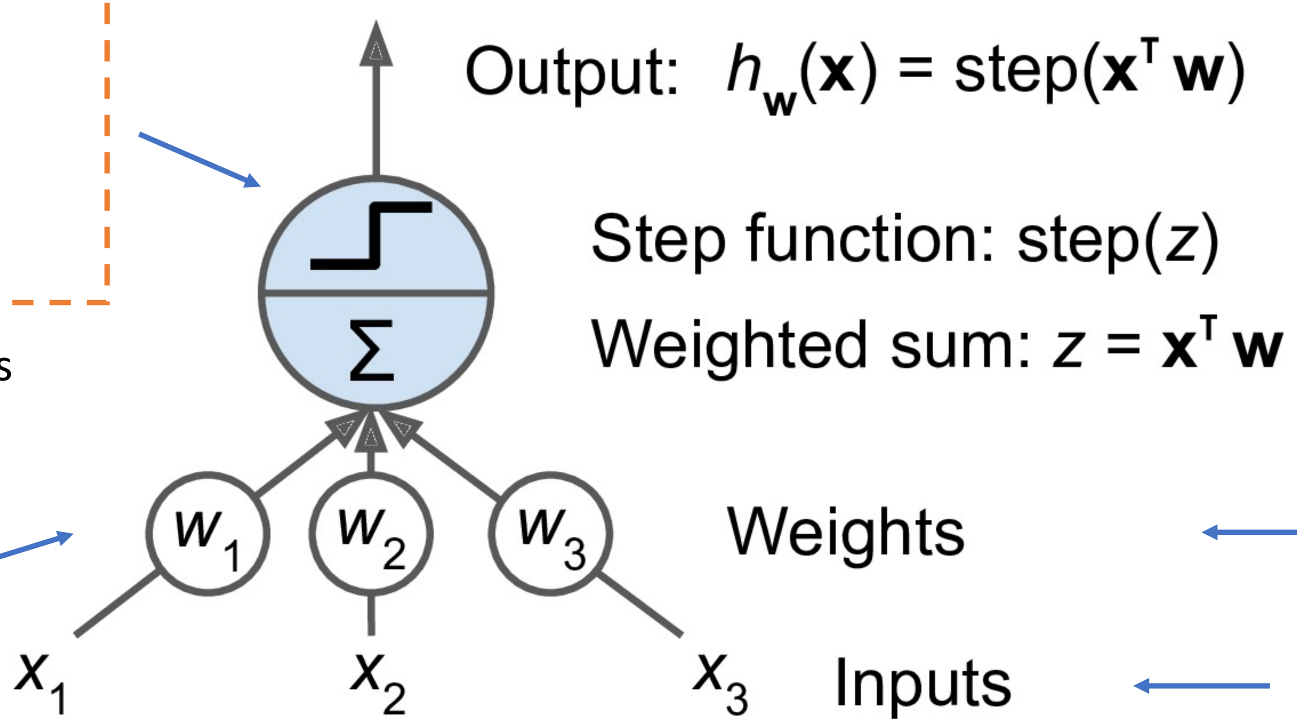
# From biological to artificial neurons

# The Perceptron (1957)

$$\text{heaviside}(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases}$$

$$\operatorname{sgn}(z) = \begin{cases} -1 & \text{if } z < 0 \\ 0 & \text{if } z = 0 \\ +1 & \text{if } z > 0 \end{cases}$$

common step functions




output now *continuous* number

each input associated with a *weight*

inputs now *continuous* numbers

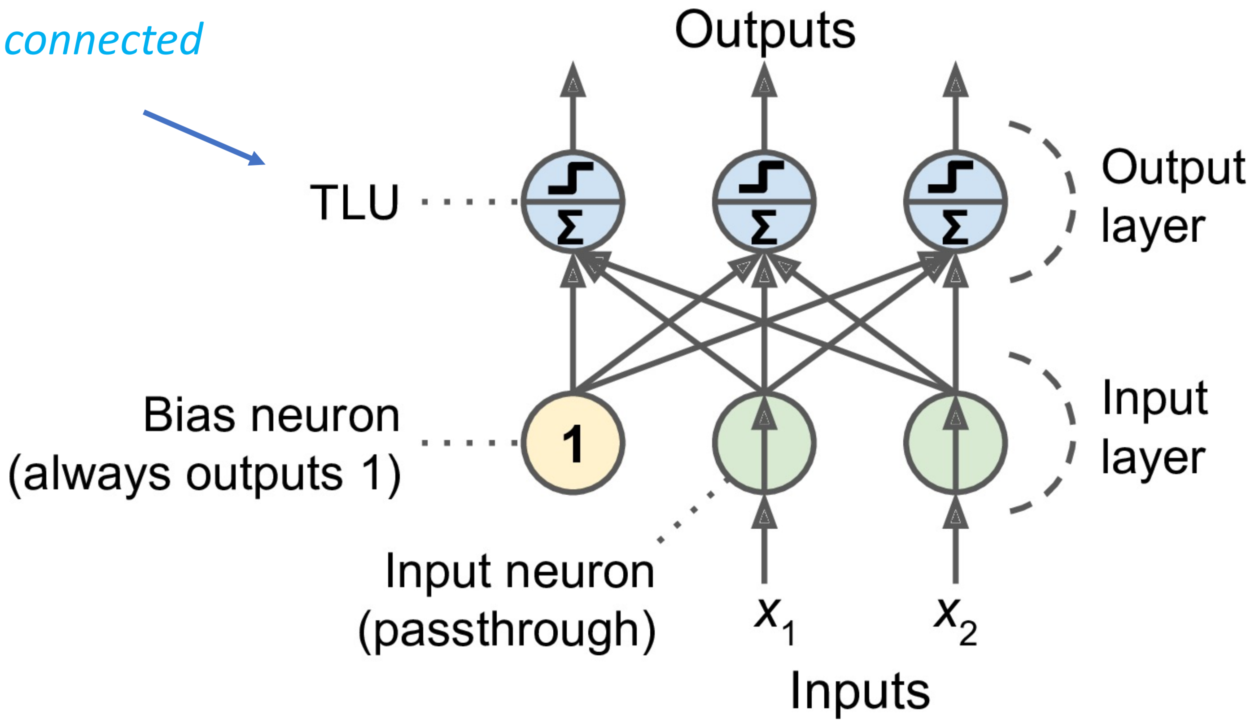
*training* means finding "right" values of $w_i$

Hands-On Machine Learning with Scikit-Learn & TensorFlow, A. Geron

# From biological to artificial neurons

# The Perceptron (1957)

each TLU connected to all inputs, i.e. *fully connected* or *dense layer*




Hands-On Machine Learning with Scikit-Learn & TensorFlow, A. Geron

*weights* or *parameters*

$$\begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix}$$

$$h_{\mathbf{W},\mathbf{b}}(\mathbf{X}) = \phi(\mathbf{X}\mathbf{W} + \mathbf{b})$$

$$\begin{pmatrix} x_1^{(1)} & x_2^{(1)} \\ x_1^{(2)} & x_2^{(2)} \\ \dots & \dots \end{pmatrix}$$

input *features*

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

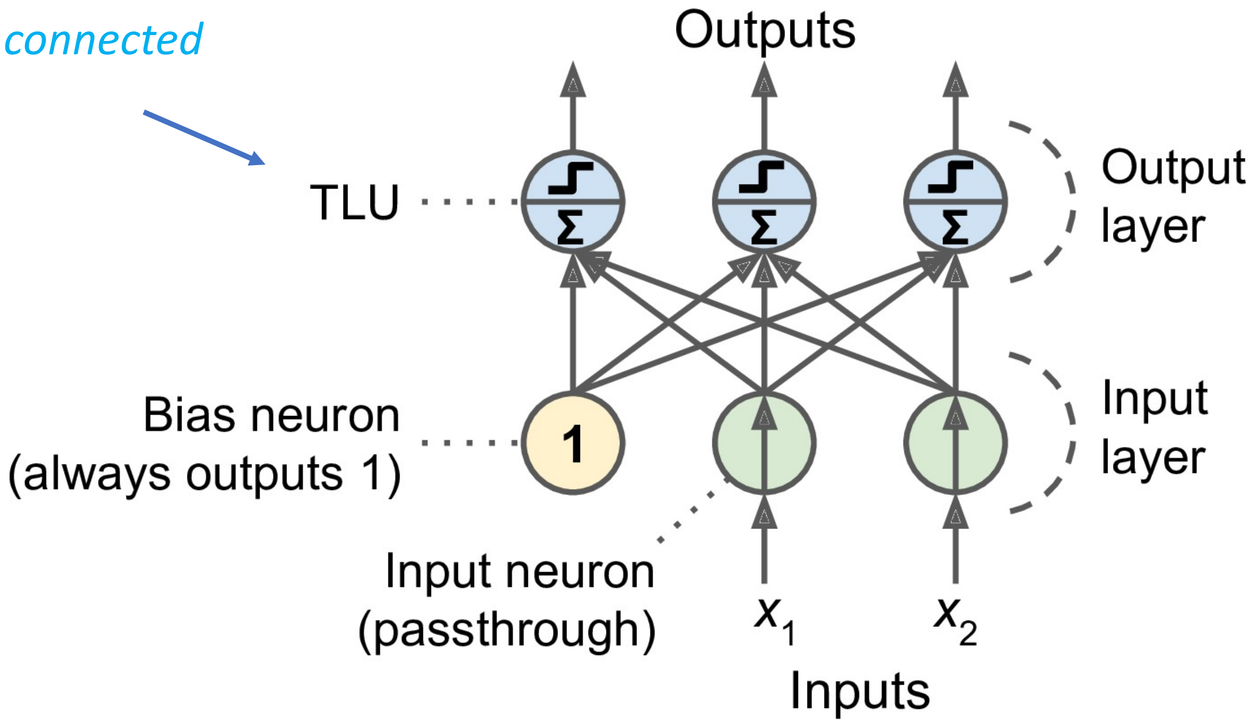
*bias*

*activation function*

# From biological to artificial neurons

## The Perceptron (1957)

each TLU connected to all inputs, i.e. *fully connected* or *dense layer*




*weights* or *parameters*

$$\begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix}$$

$$h_{\mathbf{W},\mathbf{b}}(\mathbf{X}) = \phi(\mathbf{X}\mathbf{W} + \mathbf{b})$$

$$\begin{pmatrix} x_1^{(1)} & x_2^{(1)} \\ x_1^{(2)} & x_2^{(2)} \\ \dots & \dots \end{pmatrix}$$

input *features*

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

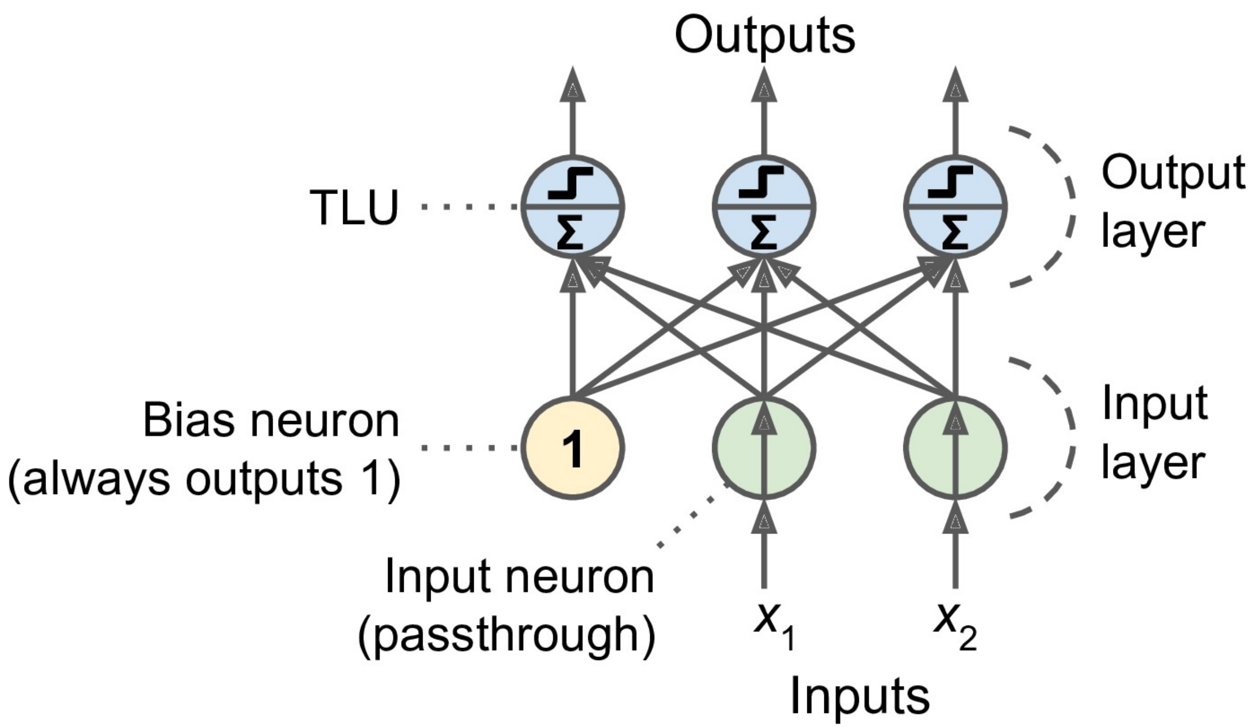
*bias*

*activation function*

Hands-On Machine Learning with Scikit-Learn & TensorFlow, A. Geron

# From biological to artificial neurons

# The Perceptron (1957)




*learning rate*

$$\Delta w_{ij} = \eta(y_j - \hat{y}_j)x_i$$

trained via Hebb's rule
"cells that fire together, wire together"

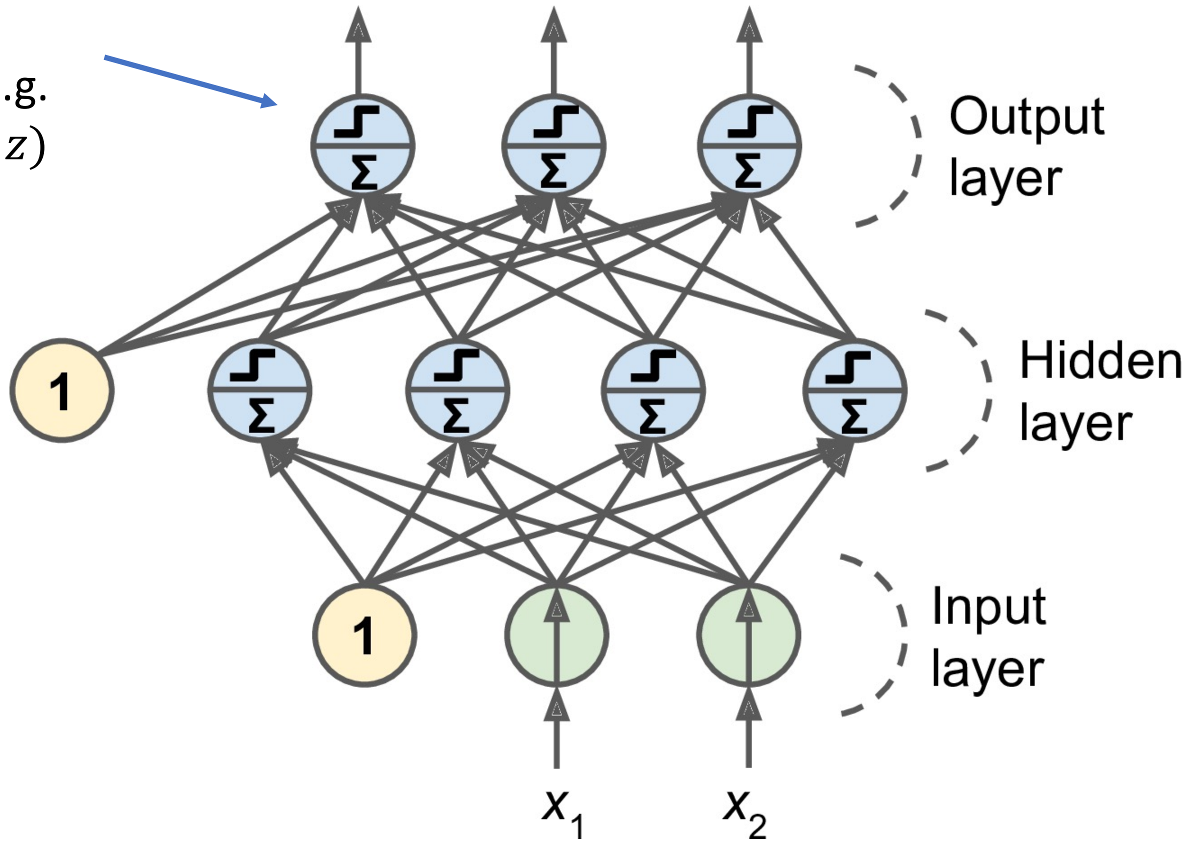
Hands-On Machine Learning with Scikit-Learn & TensorFlow, A. Geron

# From biological to artificial neurons

# The Multi-Layer Perceptron (MLP)

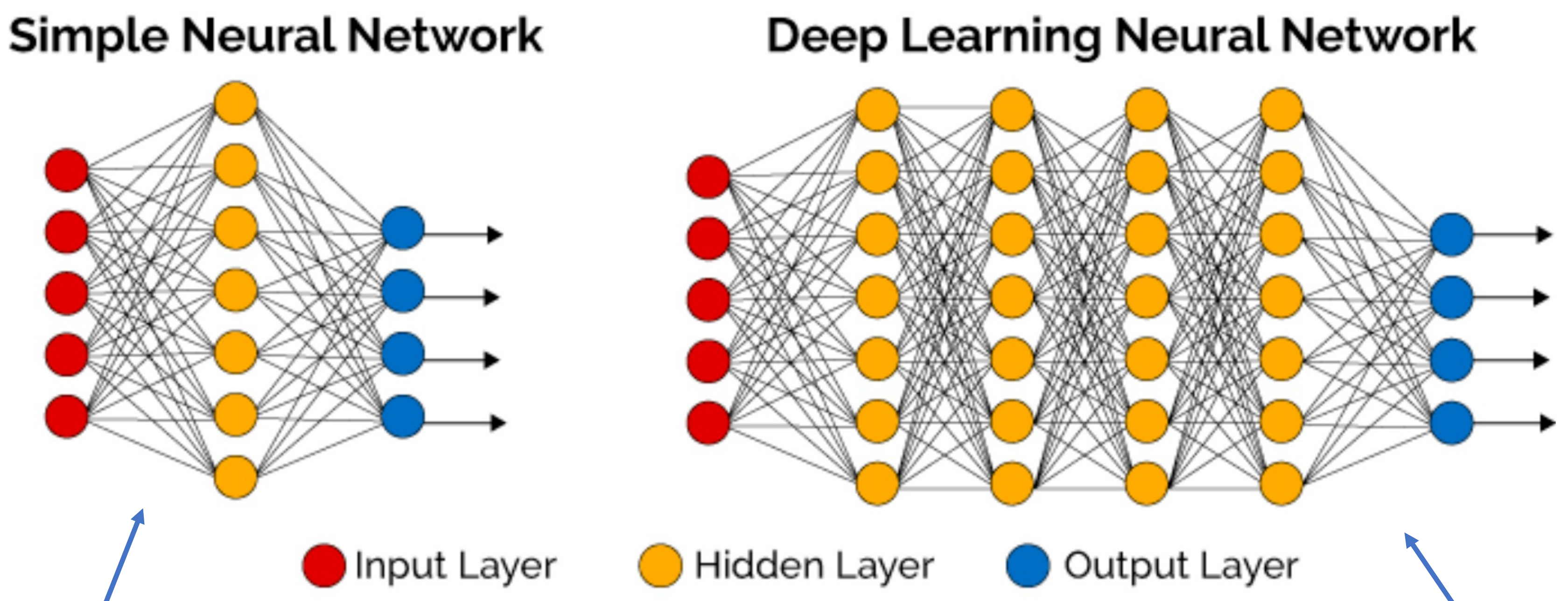
use *non-linear* activation function e.g. $ReLU(z) = \max(0, z)$

adding one or more *hidden layers* allows the network to model *non-linear functions*




trained using *stochastic gradient descent (SGD)*

Hands-On Machine Learning with Scikit-Learn & TensorFlow, A. Geron

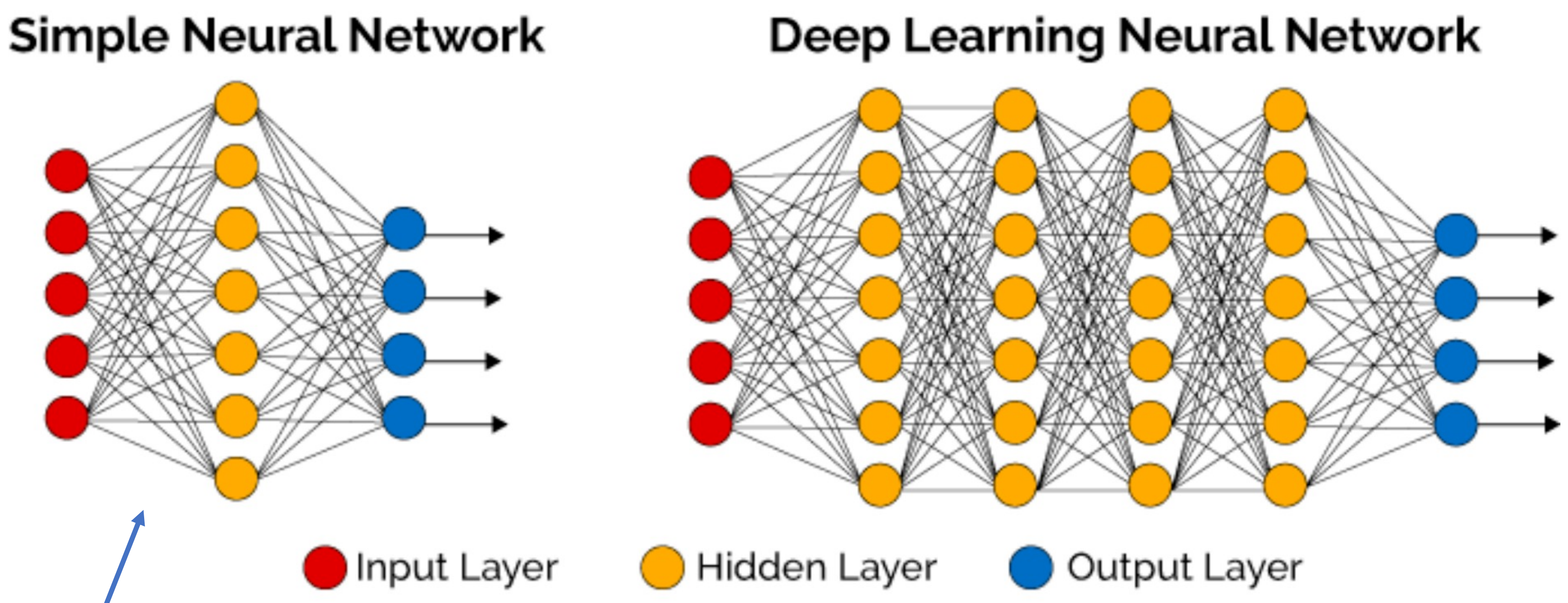



thedatascientist.com/what-deep-learning-is-and-isnt/

can learn any function with
*infinite data and width*
(universal approximation theorem)

going deep is more *parameter efficient*
⇒ better performance with same amount of training data

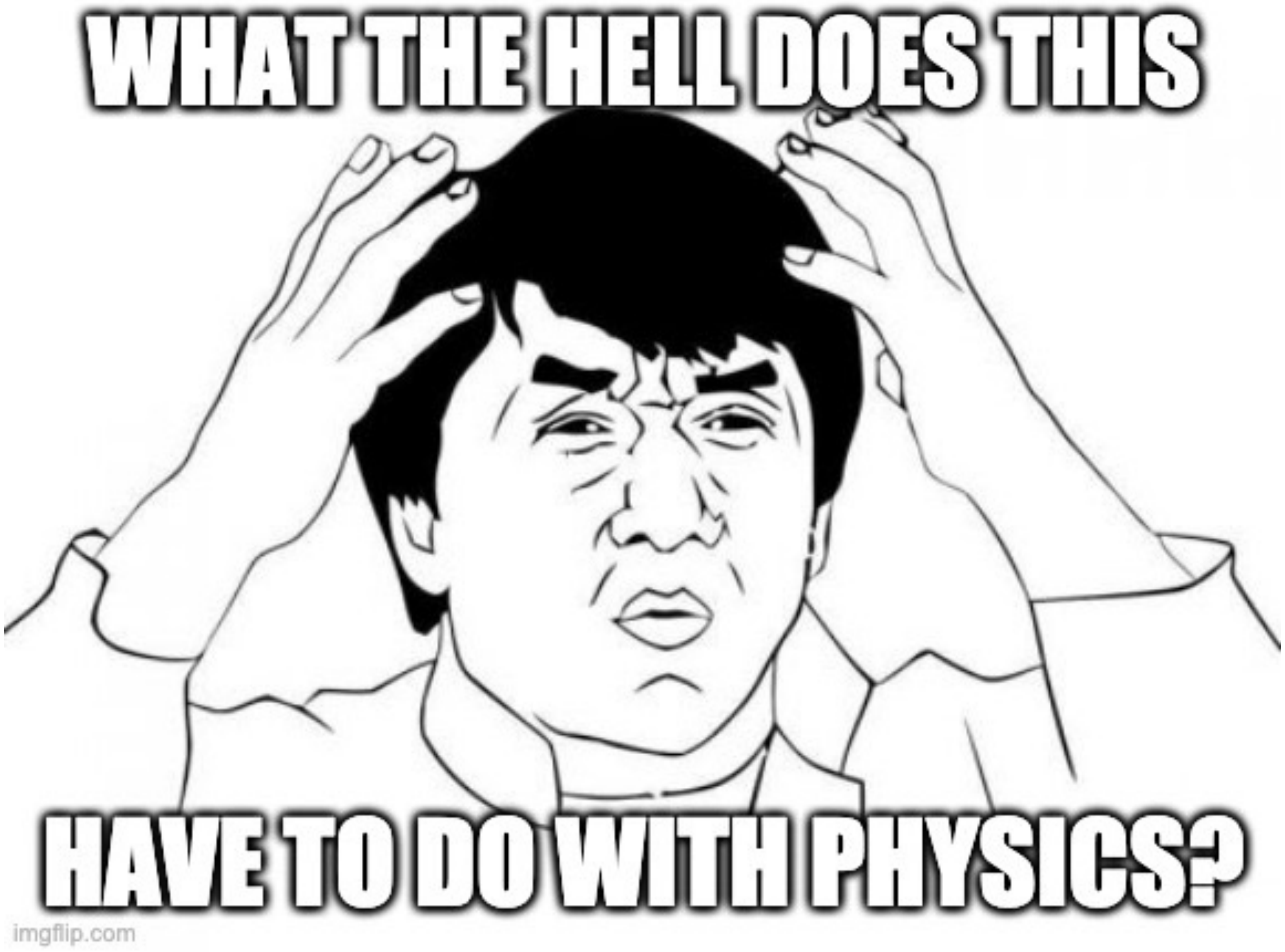
# *Shallow* vs *deep*




thedatascientist.com/what-deep-learning-is-and-isnt/

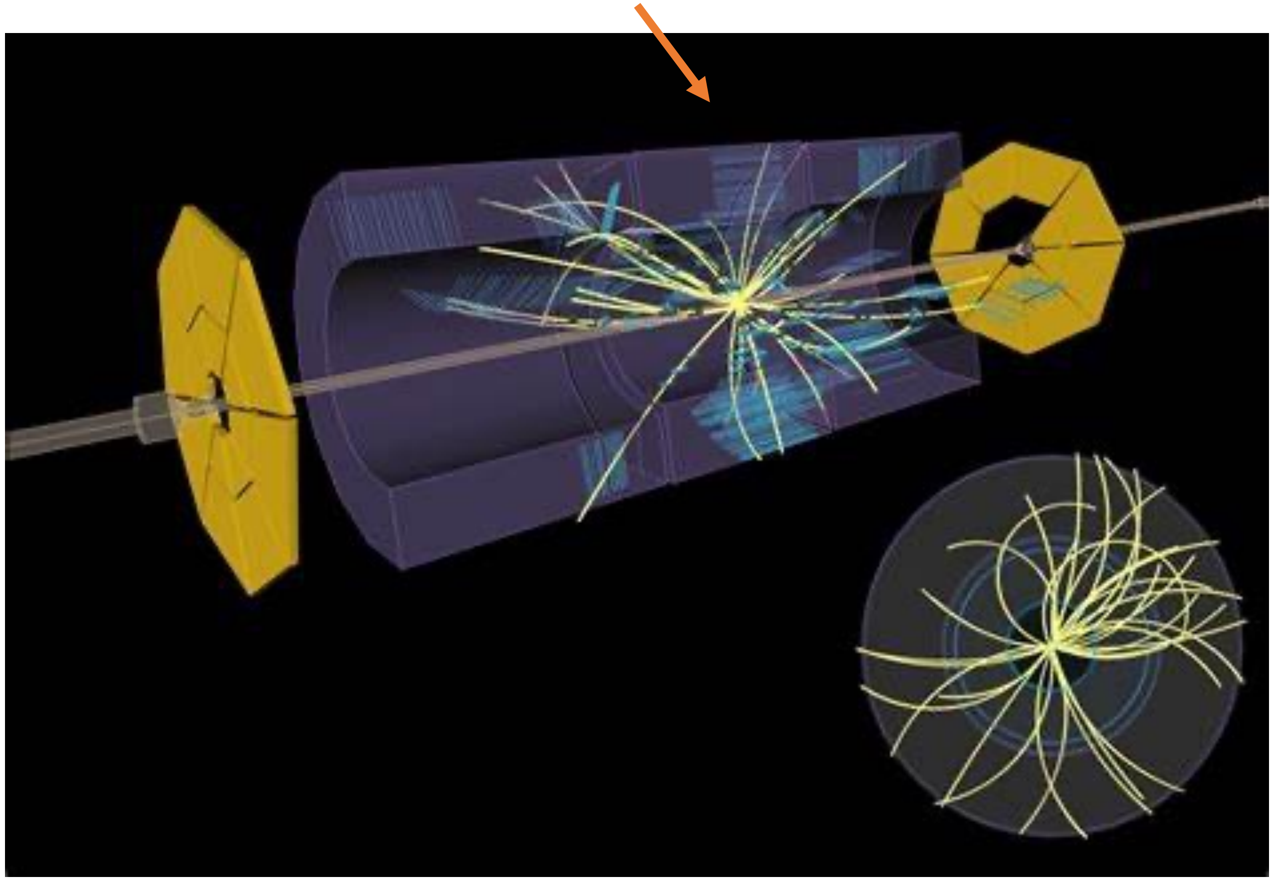
can learn any function with
*infinite data and width*
(universal approximation theorem)

# *Shallow* vs *deep*


WHAT THE HELL DOES THIS
HAVE TO DO WITH PHYSICS?
imgflip.com

*observable* decay products

- electrically charged leptons ($e$ or $\mu$)
- particle jets originating from quarks / gluons

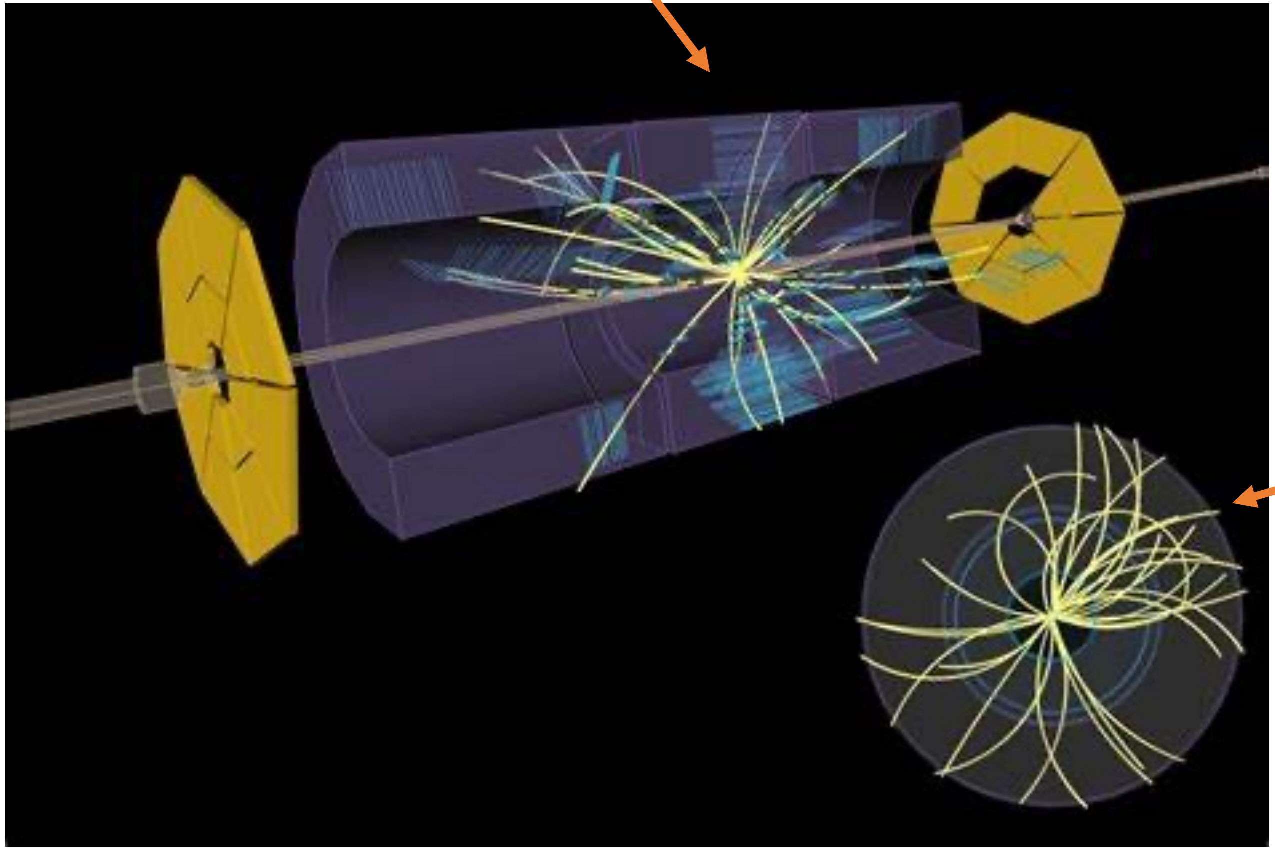


*huge datasets* & *accurate simulations*
⇒ ideal playground for ML methods


# Deep learning at the LHC

*observable* decay products

- electrically charged leptons ($e$ or $\mu$)
- particle jets originating from quarks / gluons



*huge datasets* & *accurate simulations*
⇒ ideal playground for ML methods


detector measures energies, *transverse momenta*, charges, particle ID of decay products

$$p_T^{initial} = 0$$

$$\sum_{i \in event} p_T^{(i)} = 0$$

imbalance ⇒ invisible particles ($\upsilon$) or exotic stuff like SUSY

# Deep learning at the LHC

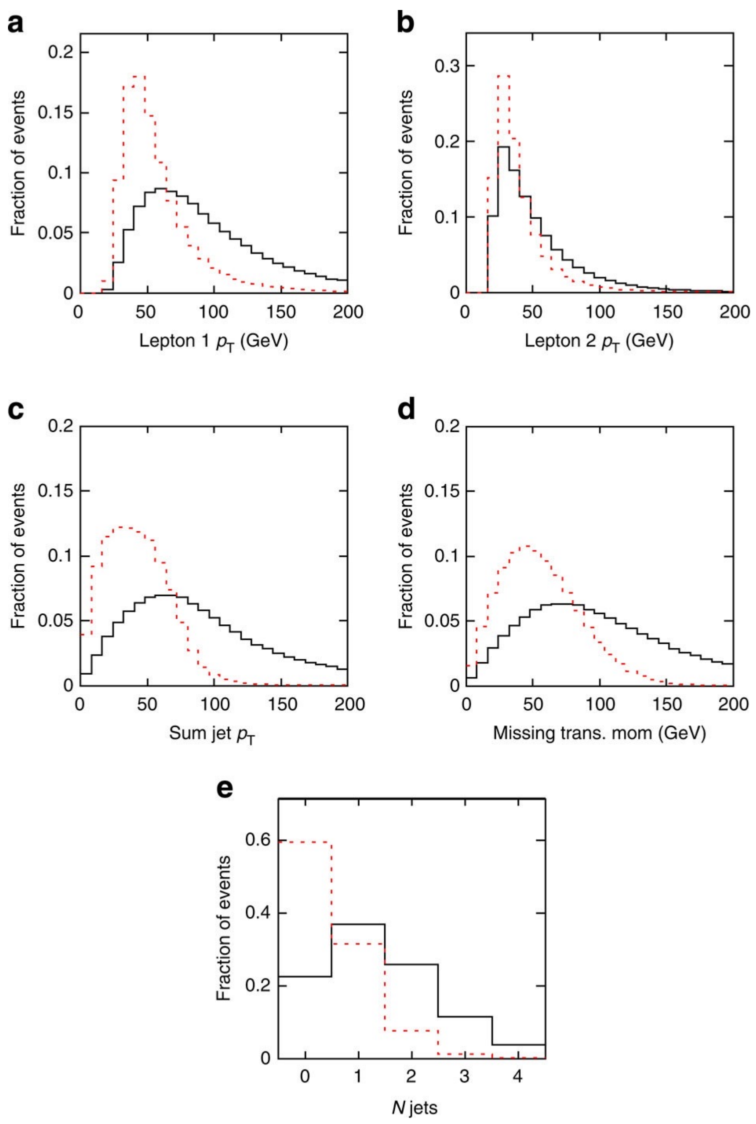
# Searching for exotic particles in high-energy physics with deep learning

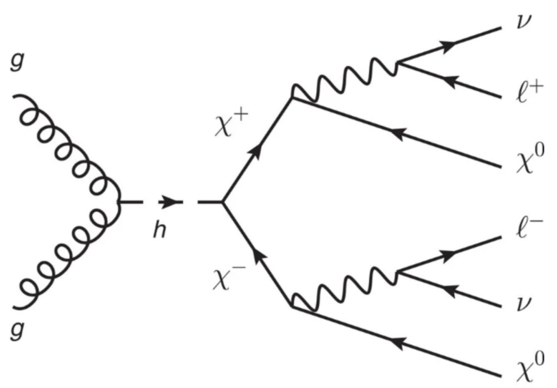
P. Baldi, P. Sadowski & D. Whiteson

Nature Communications 5, Article number: 4308 (2014) | Cite this article

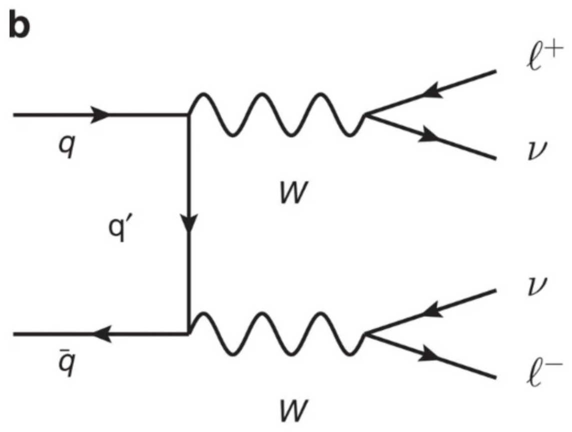


“low level” input features




*signal*




*background*

# Deep learning at the LHC

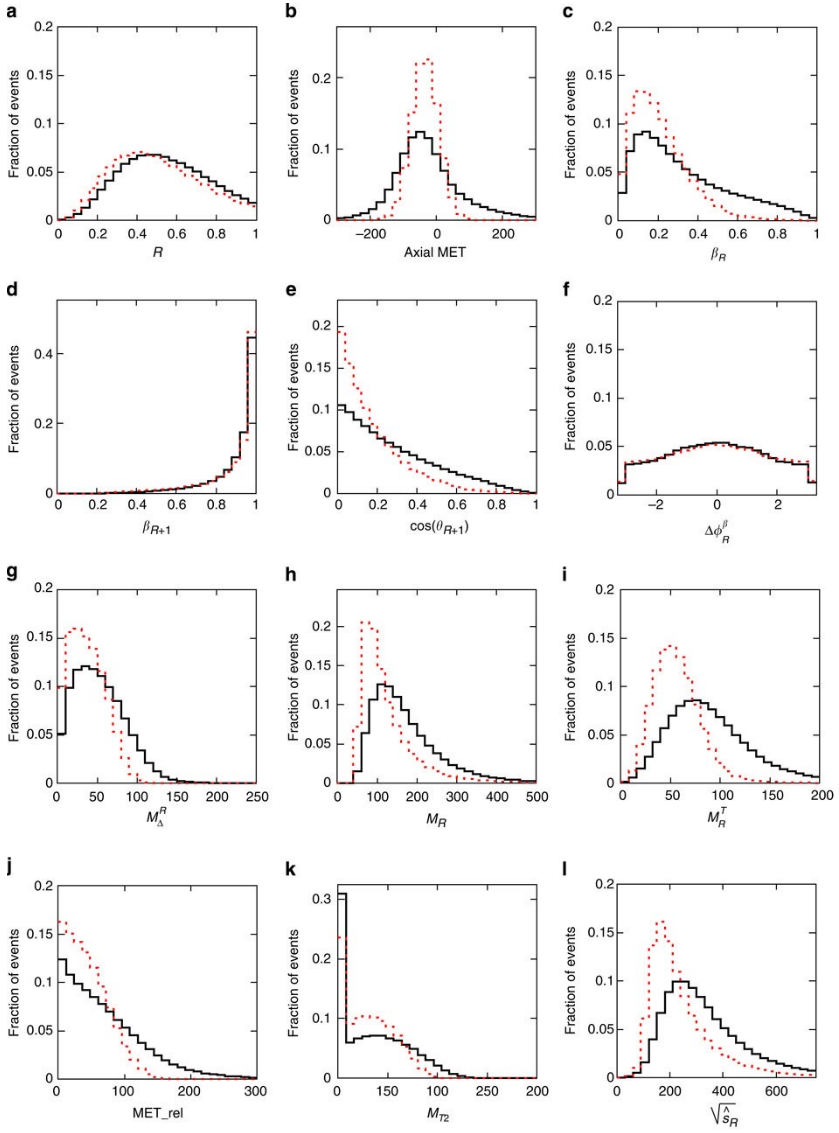
# Searching for exotic particles in high-energy physics with deep learning

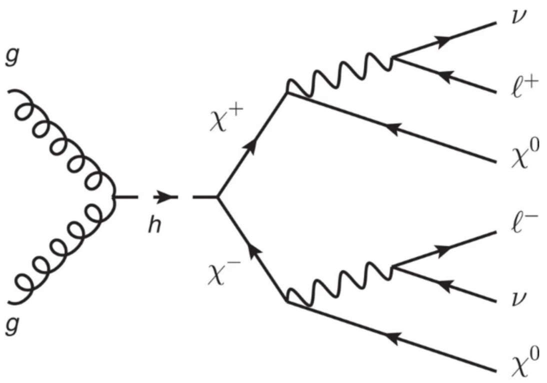
P. Baldi, P. Sadowski & D. Whiteson

Nature Communications 5, Article number: 4308 (2014) | Cite this article



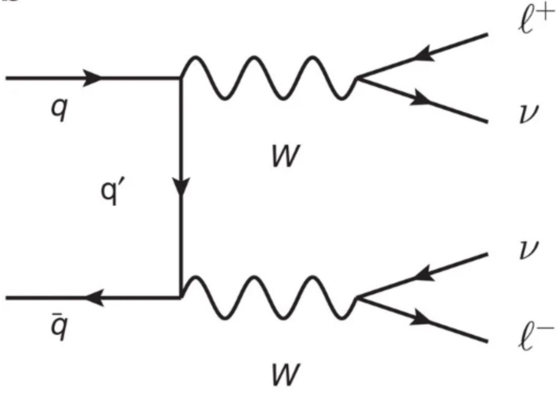
“high level”
input features




signal

b




background

# Deep learning at the LHC

# Searching for exotic particles in high-energy physics with deep learning

P. Baldi, P. Sadowski & D. Whiteson

*Nature Communications* **5**, Article number: 4308 (2014) | Cite this article



shallow NN (1-layer)

| **Technique** | **Low-level** | **High-level** | **Complete** |
|---|---|---|---|
| | | *AUC* | |
| BDT | 0.850 (0.003) | 0.835 (0.003) | 0.863 (0.003) |
| NN | 0.867 (0.002) | 0.863 (0.001) | 0.875 (<0.001) |
| $NN_{dropout}$ | 0.856 (<0.001) | 0.859 (<0.001) | 0.873 (<0.001) |
| DN | 0.872 (0.001) | 0.865 (0.001) | 0.876 (<0.001) |
| $DN_{dropout}$ | 0.876 (<0.001) | 0.869 (<0.001) | 0.879 (<0.001) |
| | | *Discovery significance* | |
| NN | $6.5\sigma$ | $6.2\sigma$ | $6.9\sigma$ |
| DN | $7.5\sigma$ | $7.3\sigma$ | $7.6\sigma$ |

# Deep learning at the LHC

# Searching for exotic particles in high-energy physics with deep learning

P. Baldi, P. Sadowski & D. Whiteson

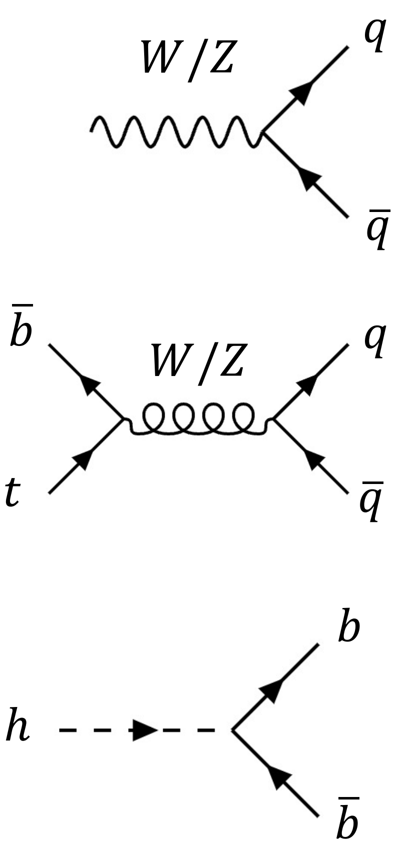
*Nature Communications* **5**, Article number: 4308 (2014) | Cite this article

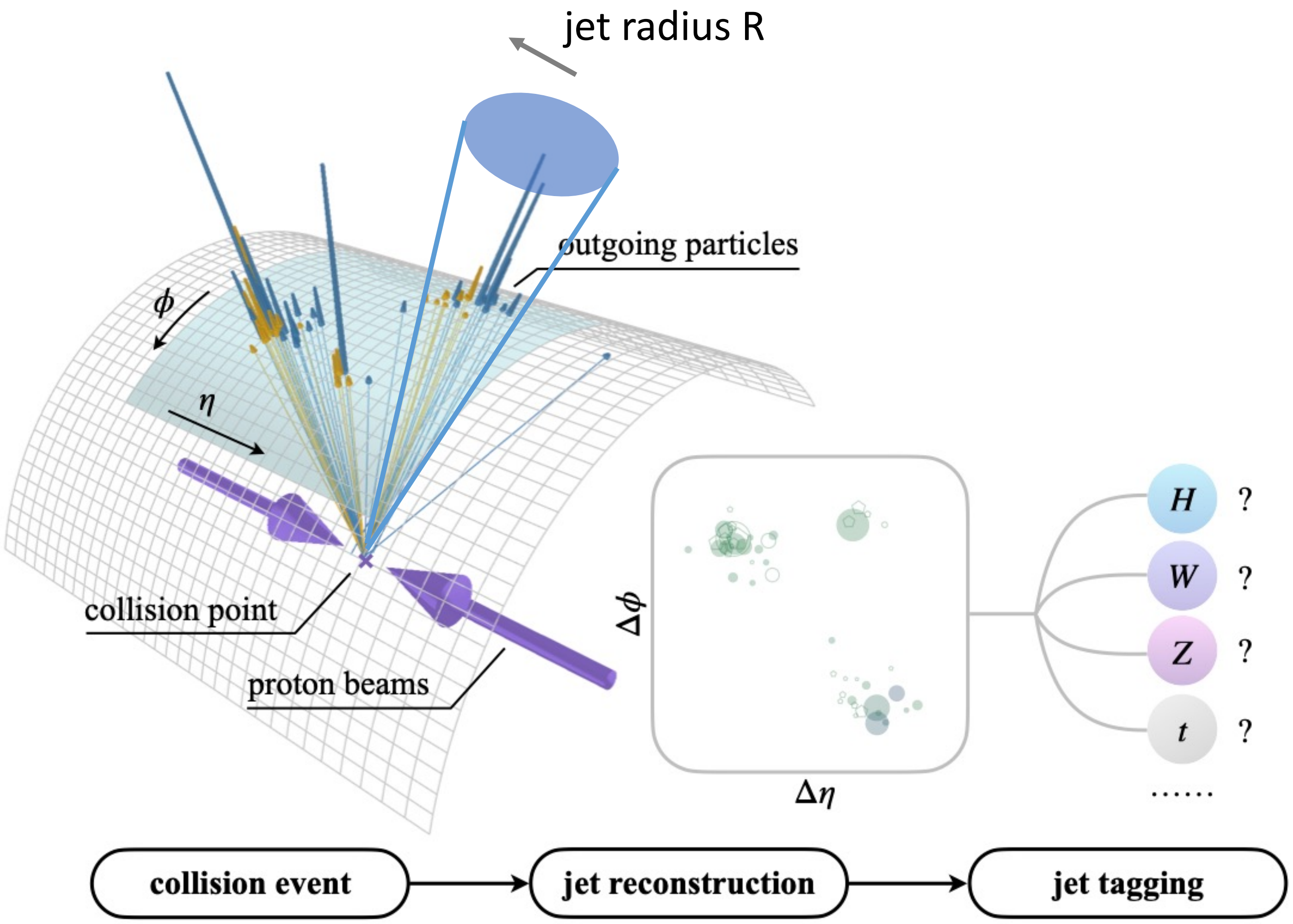


| **Technique** | **Low-level** | **High-level** | **Complete** |
|---|---|---|---|
| | | *AUC* | |
| BDT | 0.850 (0.003) | 0.835 (0.003) | 0.863 (0.003) |
| NN | 0.867 (0.002) | 0.863 (0.001) | 0.875 (<0.001) |
| $NN_{dropout}$ | 0.856 (<0.001) | 0.859 (<0.001) | 0.873 (<0.001) |
| DN | 0.872 (0.001) | 0.865 (0.001) | 0.876 (<0.001) |
| $DN_{dropout}$ | 0.876 (<0.001) | 0.869 (<0.001) | 0.879 (<0.001) |
| | | *Discovery significance* | |
| NN | $6.5\sigma$ | $6.2\sigma$ | $6.9\sigma$ |
| DN | $7.5\sigma$ | $7.3\sigma$ | $7.6\sigma$ |

deep NN (5-layer)

# Deep learning at the LHC




how to tell apart?

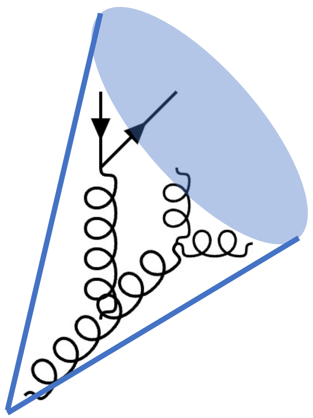
[Particle Transformer for Jet Tagging](#)

# Jet tagging at the LHC

SciPost Physics Submission

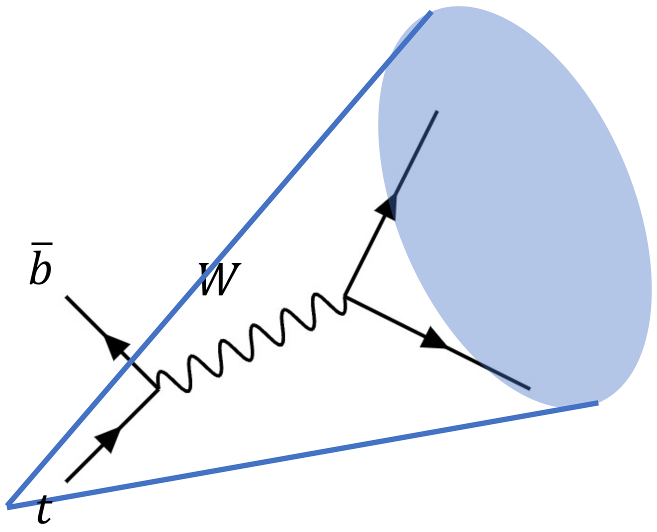
## The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

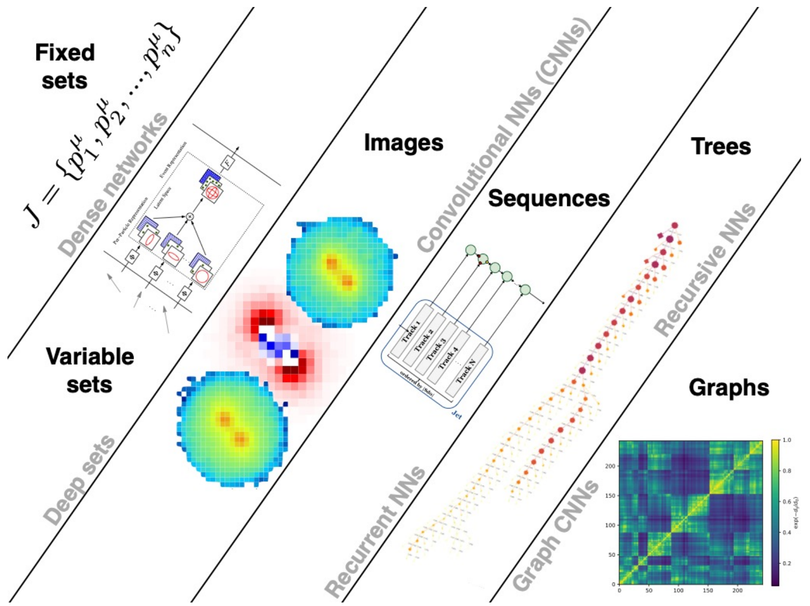


quark / gluon jet

vs




top jet




many ways to *represent jets* for NNs
which one is best?

# Jet tagging at the LHC