University of BRISTOL

Self Supervised Learning
The next challenge for industrial AI

Robert J. Piechocki

26th May 2021

Outline

- Supervised Learning
- Self Supervised: representation learning
- Self Supervised: compute into data via Self Play
- Examples: NLP, Vision, AlphaGo Zero, Protein Folding, Open-RAN, Wireless Networks

Stage 1: Learn your World (auxiliary tasks)

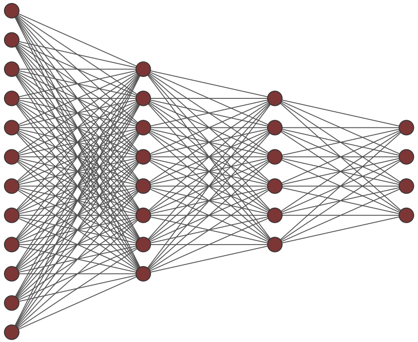
Stage 2: Solve tasks of interests

- *Human brain has 125 trillion synapses*
- *Life expectancy is 2.5 billion seconds*
- *Ergo: 50000 synapses per second (200 Kpbs)*
- *Ergo: unsupervised learning (G Hinton)*

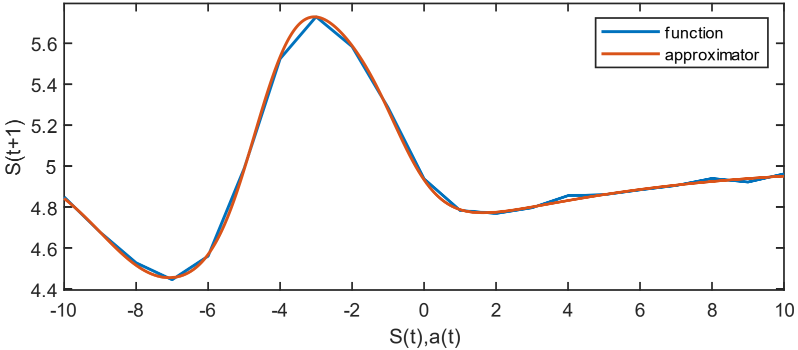
Approach evangelised by e.g: Y LeCun, Y Bengio, G Hinton "Unsupervised Learning"

Free clipart library
http://clipart-library.com

Supervised Learning – the only success so far



Dog



Universal function approximation (theorem).
Any piecewise continuous function can be approximated by NN: classification / regression etc

Examples of success:

- TESLA's Autopilot, speech recognition, image classification (medical etc), many more!
- Digital networks / spectrum: spectrum occupancy, channel estimation, traffic prediction, link evaluation, *anomaly detection*.

The problem: we need labels / examples / ground truth – very expensive!!

Self-Supervised Learning – the messy data problem

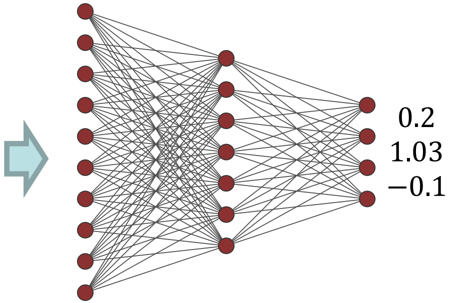
- Data deluge. We have abundance of data.
- Real word data is messy, not curated, unlabelled.
- Data labelling is very costly / impossible.
- Data distribution shift (drift).
- Self-supervised Learning to the rescue.

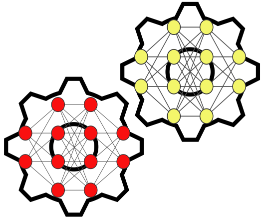
Image credit: unsplash.com

Representation Learning

Image/video, sound wave, text, network state

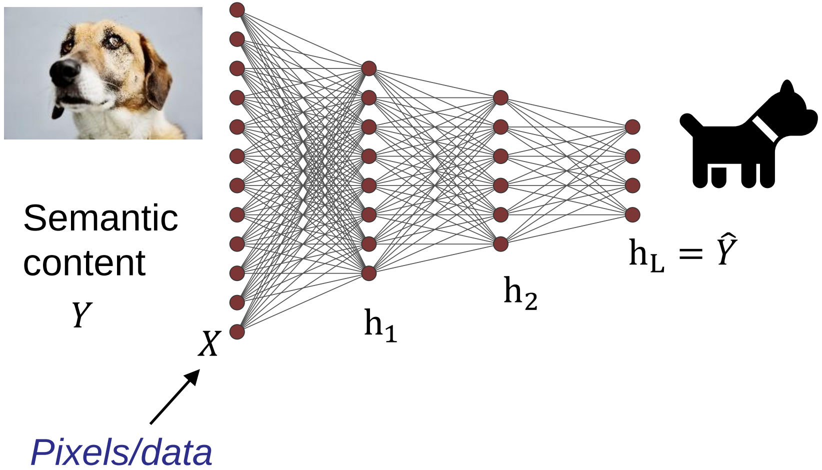
“Through this Prosperity Partnership..”





- Representation: a mapping from raw input data (specific for modality) to feature vector or tensor.
- Abstraction and invariance: capture salient information, remove all redundancy, be invariant to inconsequential changes (scale, rotation, translation etc).
- Disentanglement: each element should have an independent factor

(Latent) Representation Learning



$$I(Y;X) \geq I(Y;h_1) \geq I(Y;h_2) \geq I(Y;\hat{Y})$$

Mutual information

$$I(X;Y) = \mathrm{KL}(p(x,y) \| p(x),p(y)) = \int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy$$

$$I(X;Y) = H(Y) - H(Y|X)$$

Data processing inequality:

For Markov chains $X \rightarrow Y \rightarrow Z$

$$I(X;Y) \geq I(X;Z)$$

However, we cannot train the network in a supervised way as we have no access to the semantic content Y

Aiming at:

$$\inf[I(X;\hat{Y})]$$

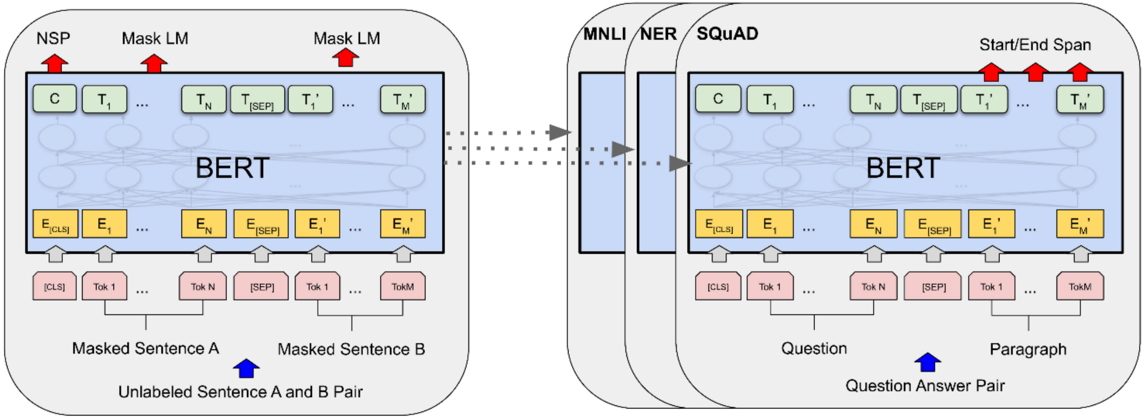
$$st.\ I(\hat{Y};Y) > \alpha$$

Connections with Rate Distortion Theory

R. Shwartz-Ziv, N. Tishby "Opening the Black Box of Deep Neural Networks via Information" 2017, https://arxiv.org/abs/1703.00810

Natural Language Models - BERT

Self-Supervised training



Pre-training

Fine-Tuning

Stage 1:
Pre-train the NLP model

Stage 2:
Fine tuning for NLP tasks

- **machine translation**
- **Q&A**
- **text summarisation**
- **etc**

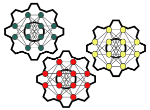
The masked language model randomly masks some of the tokens from the input, and the objective is to predict the original vocabulary id of the masked word based only on its context.

+ guess the sequence order given two sentences

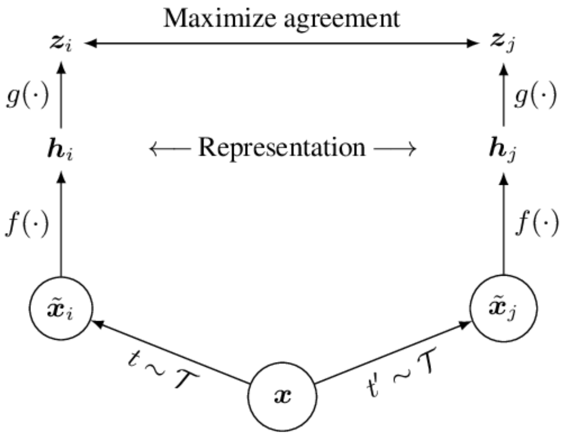
Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019.

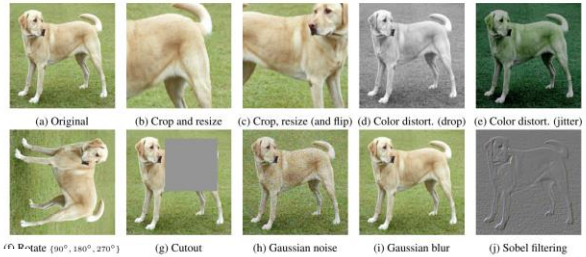
Contrastive Representation Learning - Vision

Stage 1:
Learn visual representation (self-supervised)



Stage 2:
Classification tasks





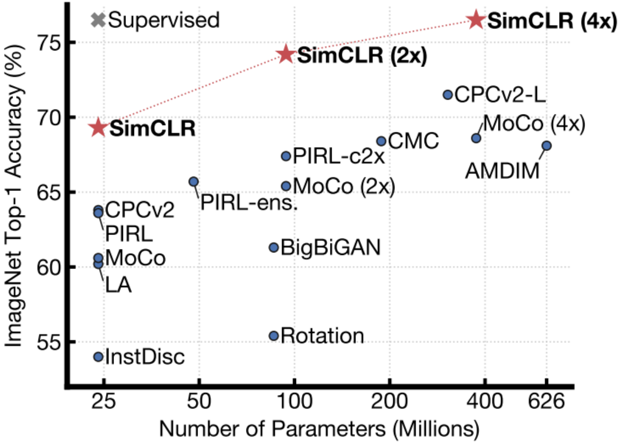


Figure 1. ImageNet Top-1 accuracy of linear classifiers trained on representations learned with different self-supervised methods (pretrained on ImageNet). Gray cross indicates supervised ResNet-50. Our method, SimCLR, is shown in bold.

Contrastive loss:

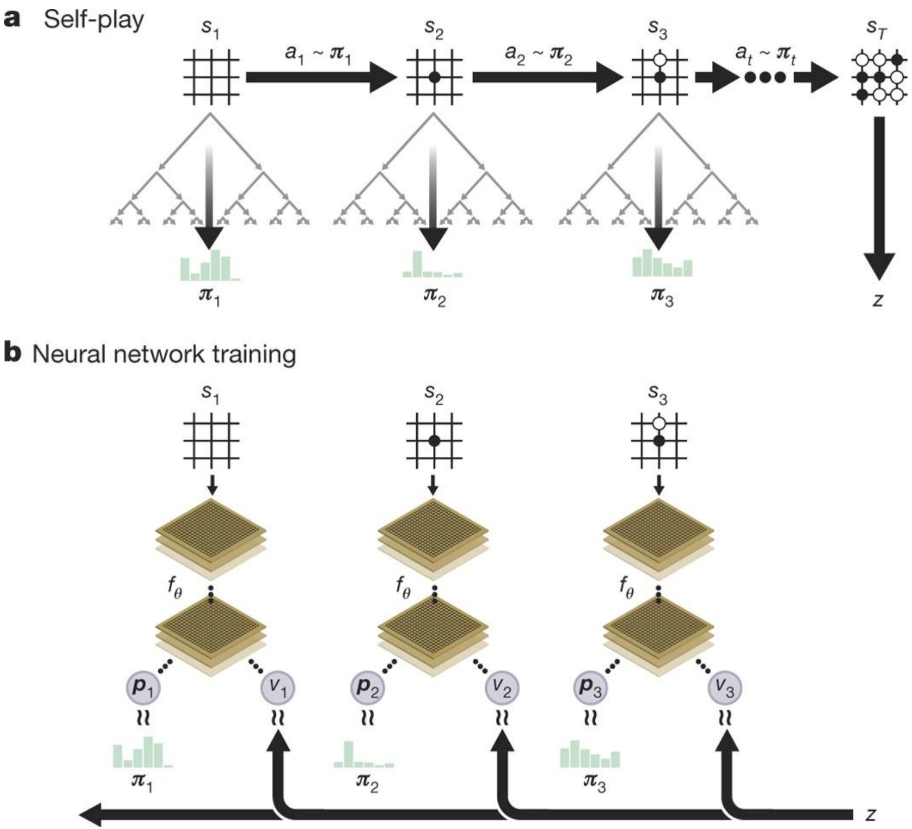
$$\ell_{i,j} = -\log \frac{\exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)}$$

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G.E. (2020). A Simple Framework for Contrastive Learning of Visual Representations. ArXiv, abs/2002.05709

AlphaGo Zero, Self-play

Image credit: DeepMind & Nature

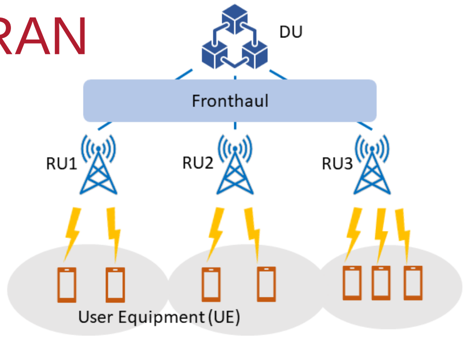
- The original AlphaGo (version that beat Lee Sedol) used expert data to pretrain neural networks.
- AlphaGo Zero trains without data!! Self-Play. And comfortably beats AlphaGo
- MuZero (latest) learns all representations



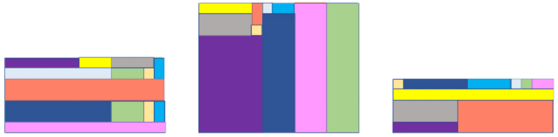
Images credit: DeepMind & Nature

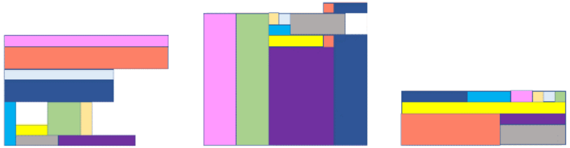
Silver, D., Schrittwieser, J., Simonyan, K. et al. Mastering the game of Go without human knowledge. Nature 550, 354–359 (2017)

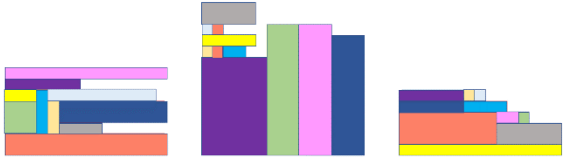
Self-play Learning for Resource Assignment in Open-RAN

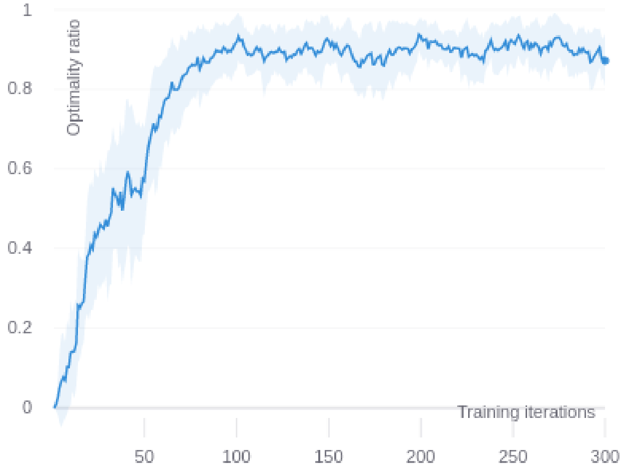


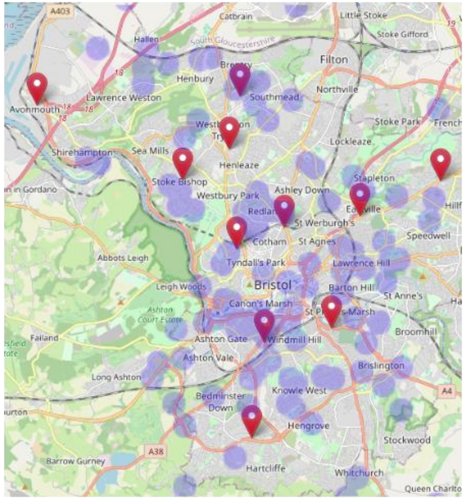
- Neural-MCTS (AlphaGo Zero) for "bin-packing" problem
- Curriculum Learning via Ranked Reward Self-play.


Self-play Neural-MCTS


HVRAA

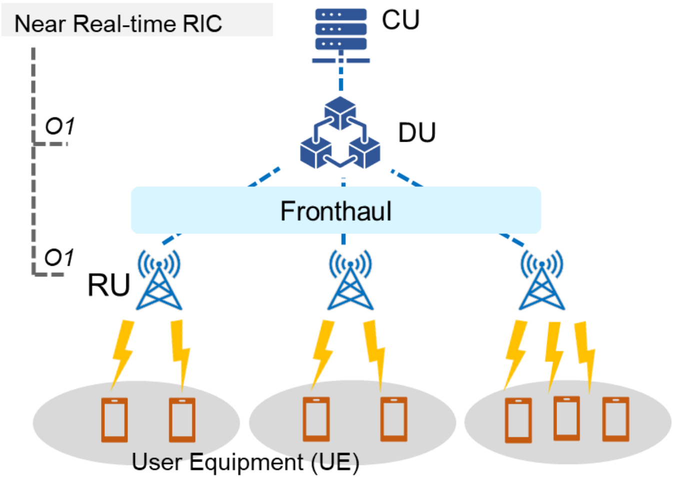

LEGO heuristics





Self-play Learning Strategies for Resource Assignment in Open-RAN Networks
X Wang, J Thomas, RJ Piechocki, S Kapoor, R Santos-Rodriguez, A Parekh,2021 https://arxiv.org/abs/2103.02649

Where can we use SSL?



- Smart Interference Management for QoS/QoE Optimisation
- Broadcast Beam Optimisation for Coverage Enhancements
- Massive MIMO Channel Estimation and Detection
- Massive MIMO Precoding and Scheduling
- Distributed and Cell-less Massive MIMO
- Disaggregated and Open Massive MIMO
- RAN Energy Efficiency Optimisation
- Intelligent Reflecting Surfaces & Holographic Beamforming

AI for Massive MIMO (AIMM) project

AIMM Project Key Information

Title: AIMM (AI-enabled Massive MIMO)

Project Coordinator: Arman Shojaeifard

Project Status: Running

Clusters: UK, Germany, France, Canada

Duration: 2 years

Start Date: Oct 2020

End Date: Sep 2022

Budget (total): 4,732 K€

Effort: 44.83 PY

 # Partners: 10

 # Work-Packages: 6

Project-ID: C2019/2-5

Website: https://www.celticnext.eu/project-aimm/

CELTIC-NEXT Next Generation Telecommunications

BT

NOKIA

University of BRISTOL

Vilicom signalling the future

Universität Stuttgart

interdigital

Supélec

thinkRF

Loughborough University

IMST

montimage

NG-CDI

Next Generation Converged Digital Infrastructure

NG-CDI's ambition is to develop a transformational approach to managing the next generation of digital infrastructure for the UK. This requires foundational research in a diverse range of areas from networking and communications, AI, industrial automation and organisational behaviour.

NG-CDI
NEXT GENERATION • CONVERGED DIGITAL INFRASTRUCTURE

https://www.ng-cdi.org/

AGILE:

A completely new architecture for digital infrastructures, composed of highly dynamic network functions based on a micro-NFV approach that are collectively able to adapt to the real-time requirements of future digital services.

AUTONOMIC:

Creating a new autonomic framework for digital infrastructure to equip the nodes of the infrastructure network with the ability to understand their state, detect and diagnose disruptions to service, and take autonomous actions.

AUTONOMOUS:

Implementing approaches for the successful integration of these technologies within the business functions with an aim to improve service assurance and organisational value.

University of BRISTOL | UNIVERSITY OF CAMBRIDGE | Lancaster University | UNIVERSITY OF SURREY | UKRI Engineering and Physical Sciences Research Council | BT

Wrap up: What's next for Self Supervised Learning?

- The case for Self-Supervised Learning is very strong for natural signals.
- In Digital Comms we use "human/machine" made signals. This makes the problem somewhat different than NLP, Image/Video etc.
- Decades of research into classical Signal Processing / detection / estimation etc. It makes no sense to replace it wholesale with AI/SSL.
- Many classical techniques require good hardware and will not work with non-linear distortions.
- Algorithms' approximation (Universal Function Approximation). Take advantage of HW accelerators (TPUs).
- Reinforcement Learning – beyond reward only guided learning.

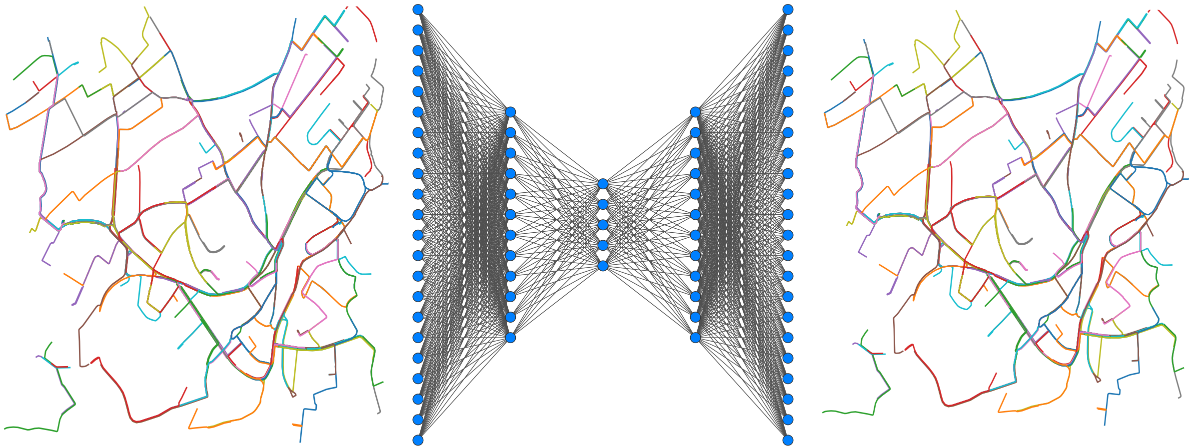
@BristolCSN
@bristol_smart

Thank you!

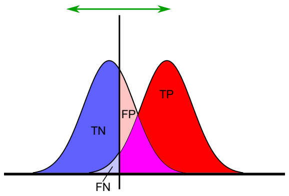
Special thanks to:
X Wang, J Thomas, A Khalil, R Santos-Rodriguez, R McConville, T Yamagata, J Bocus, A Parekh, S Kapoor, N Race, A Shojaeifard, S Cassidy, M Beach, N Lane, K Chetty

Anomaly Detection unsupervised learning

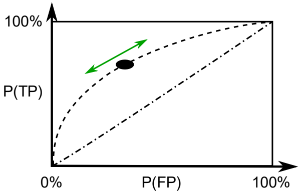
- Intrusion Detection Systems
- Anomaly Detection
- Outlier Detection



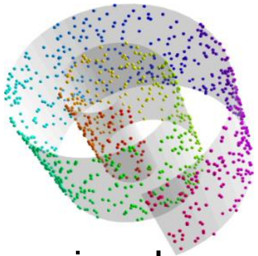
Deep Autoencoder Network



TP	FP
FN	TN



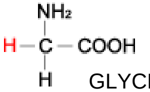
TP - True Positive
FP - False Positive
TN - True Negative
FN - False Negative



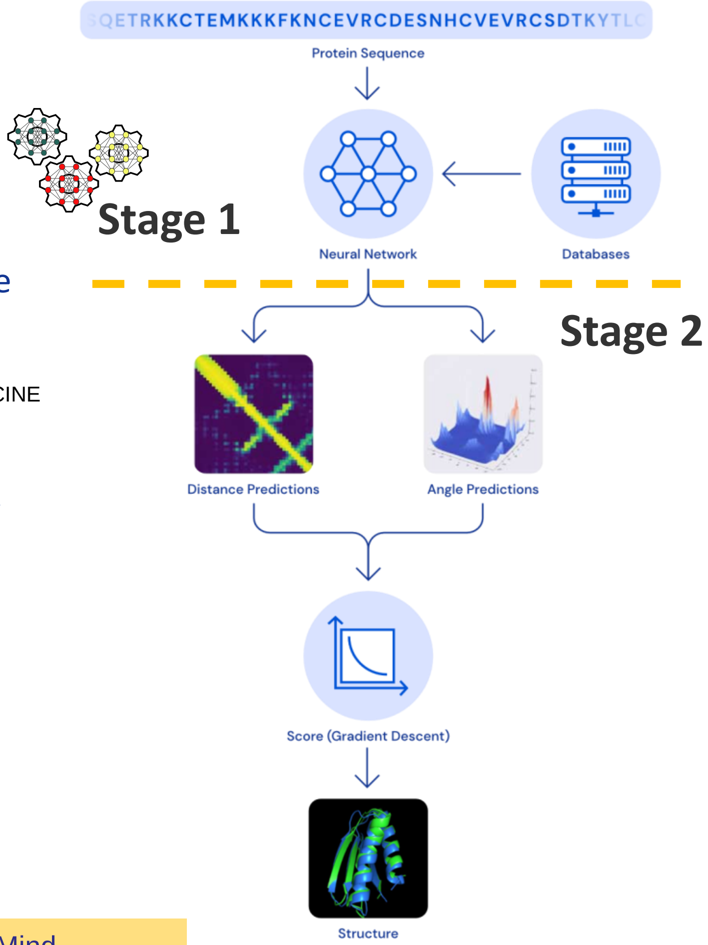
Low dimensional representation (manifold)

AlphaFold

- Proteins are large, complex molecules essential to all of life.
- What any given protein can do depends on its unique 3D structure.
- There are 20 amino acids that make up proteins and all have the same basic structure, differing only in the side chain they have.



GLYCINE

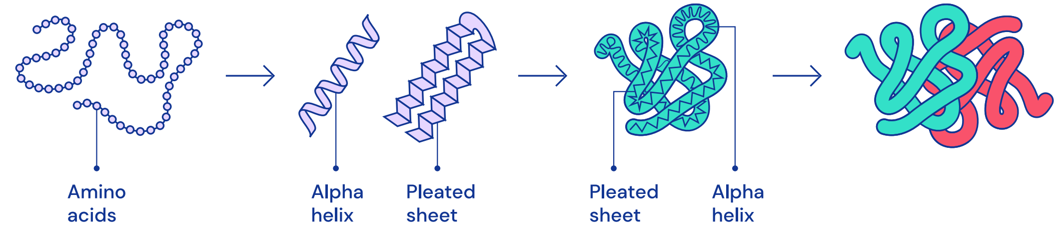


Every protein is made up of a sequence of amino acids bonded together

These amino acids interact locally to form shapes like helices and sheets

These shapes fold up on larger scales to form the full three-dimensional protein structure

Proteins can interact with other proteins, performing functions such as signalling and transcribing DNA



https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery, images: DeepMind