PreciDIAB
National Center for Precision Diabetic Medicine

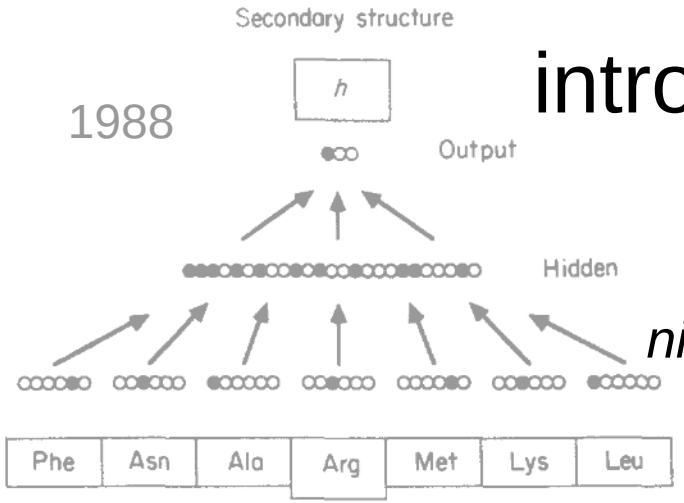
Egid
European Genomic Institute for Diabetes

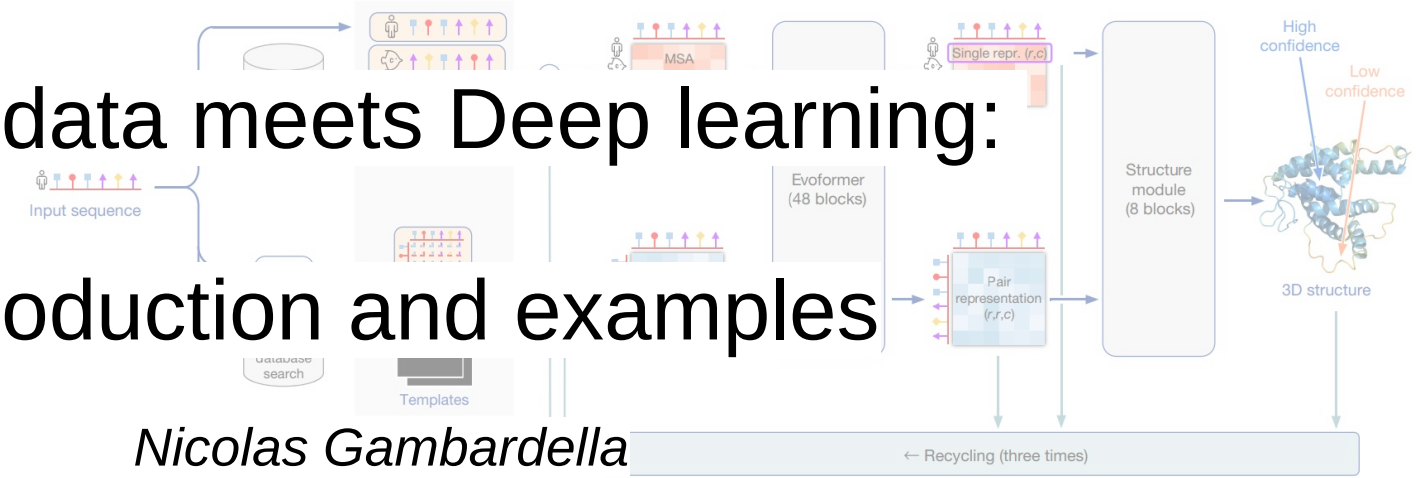
# Omics data meets Deep learning: introduction and examples

*Nicolas Gambardella*

*nicolas.gambardella@univ-lille.fr*

1988



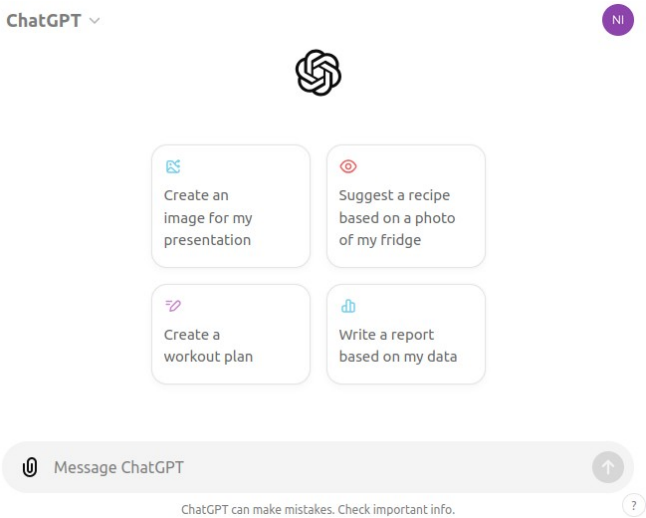

2021

Région Hauts-de-France

UMR 1283 8199
Functional (Epi)genomics and Mechanisms of Type 2 Diabetes and Related Diseases


Google
Self-driving car


Boston
HD-03


ChatGPT
NI
Create an image for my presentation
Suggest a recipe based on a photo of my fridge
Create a workout plan
Write a report based on my data
Message ChatGPT
ChatGPT can make mistakes. Check important info.


AlphaGo
Lee Sedol

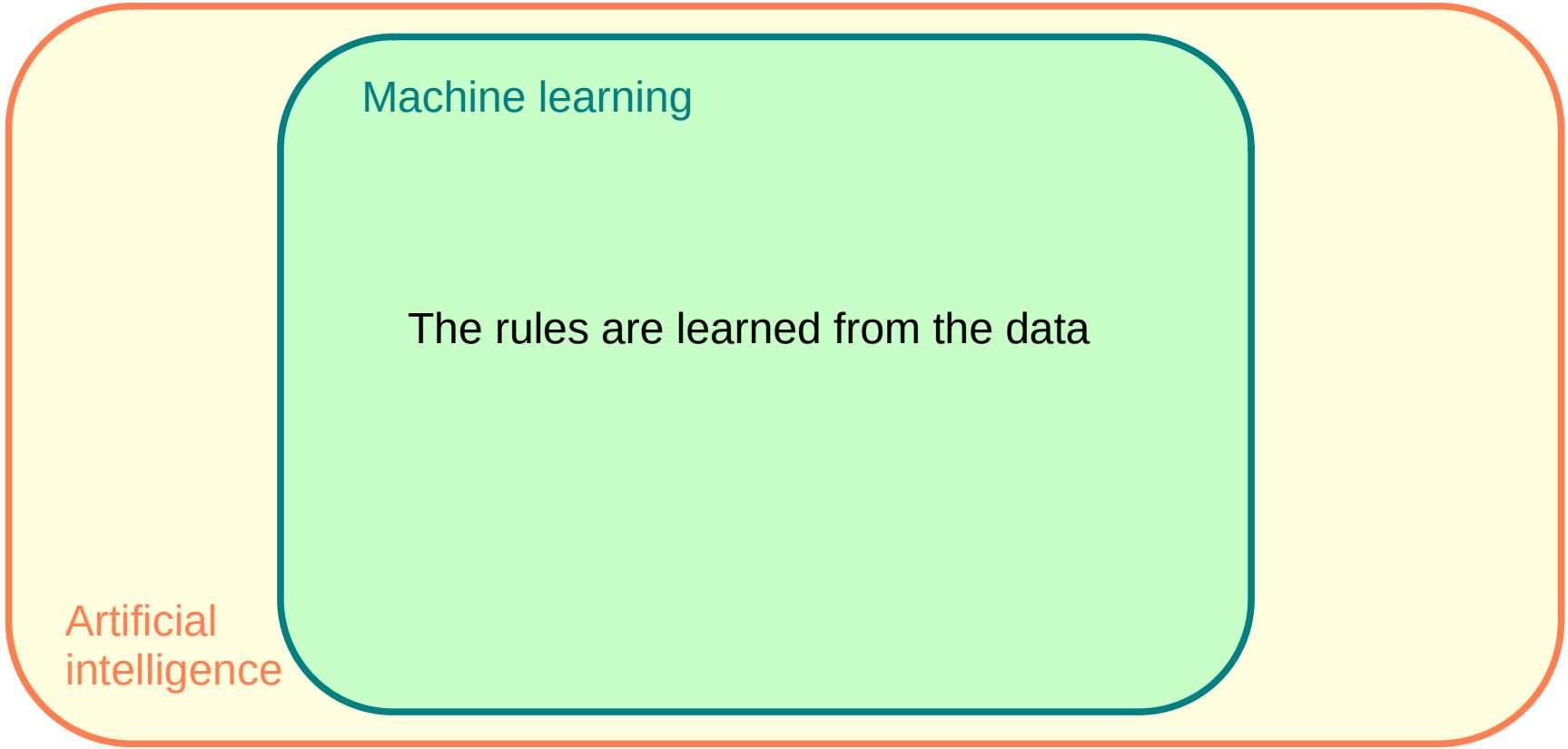
AlphaFold

# Some terminology

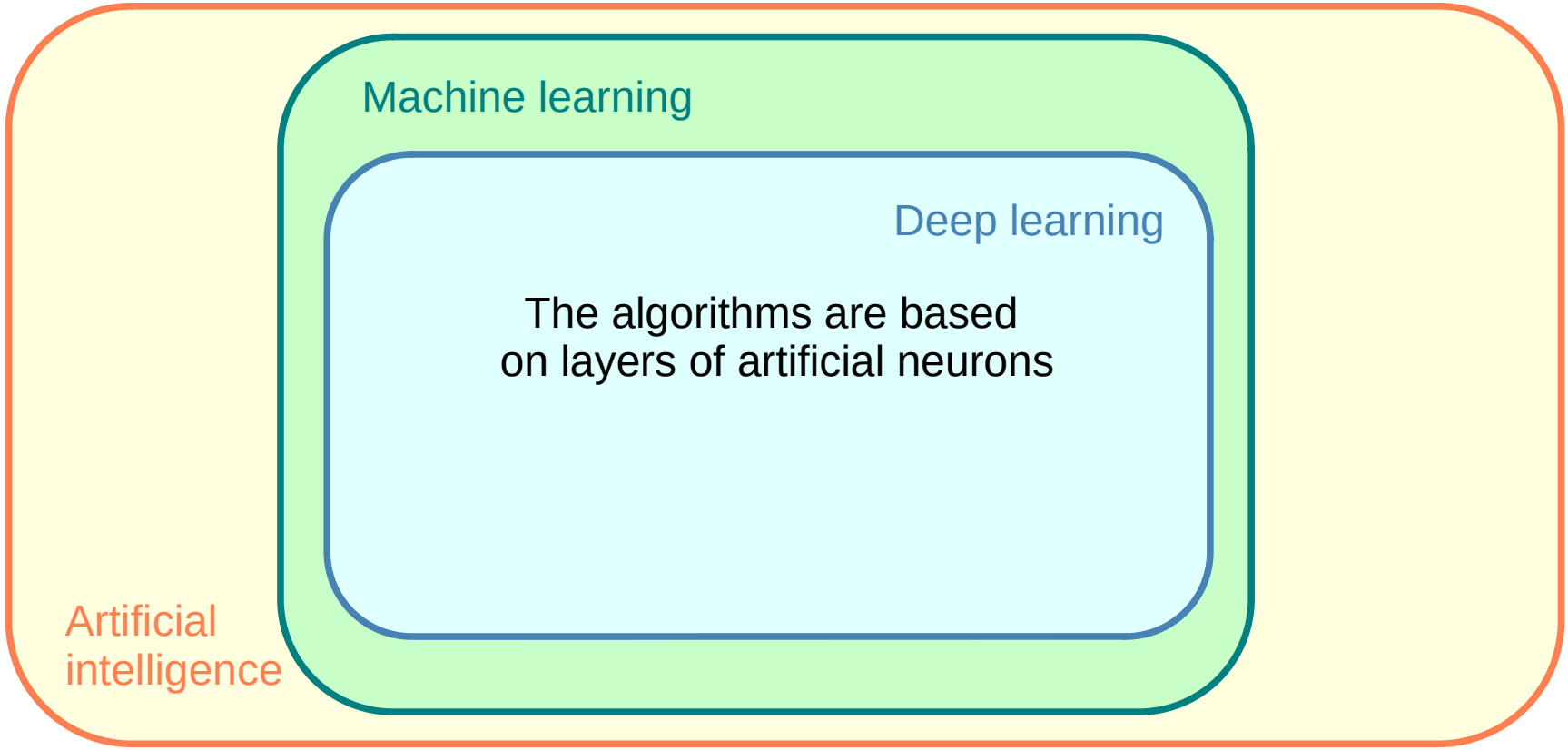
Assessments, evaluations, decisions, predictions made by software tools

Artificial intelligence

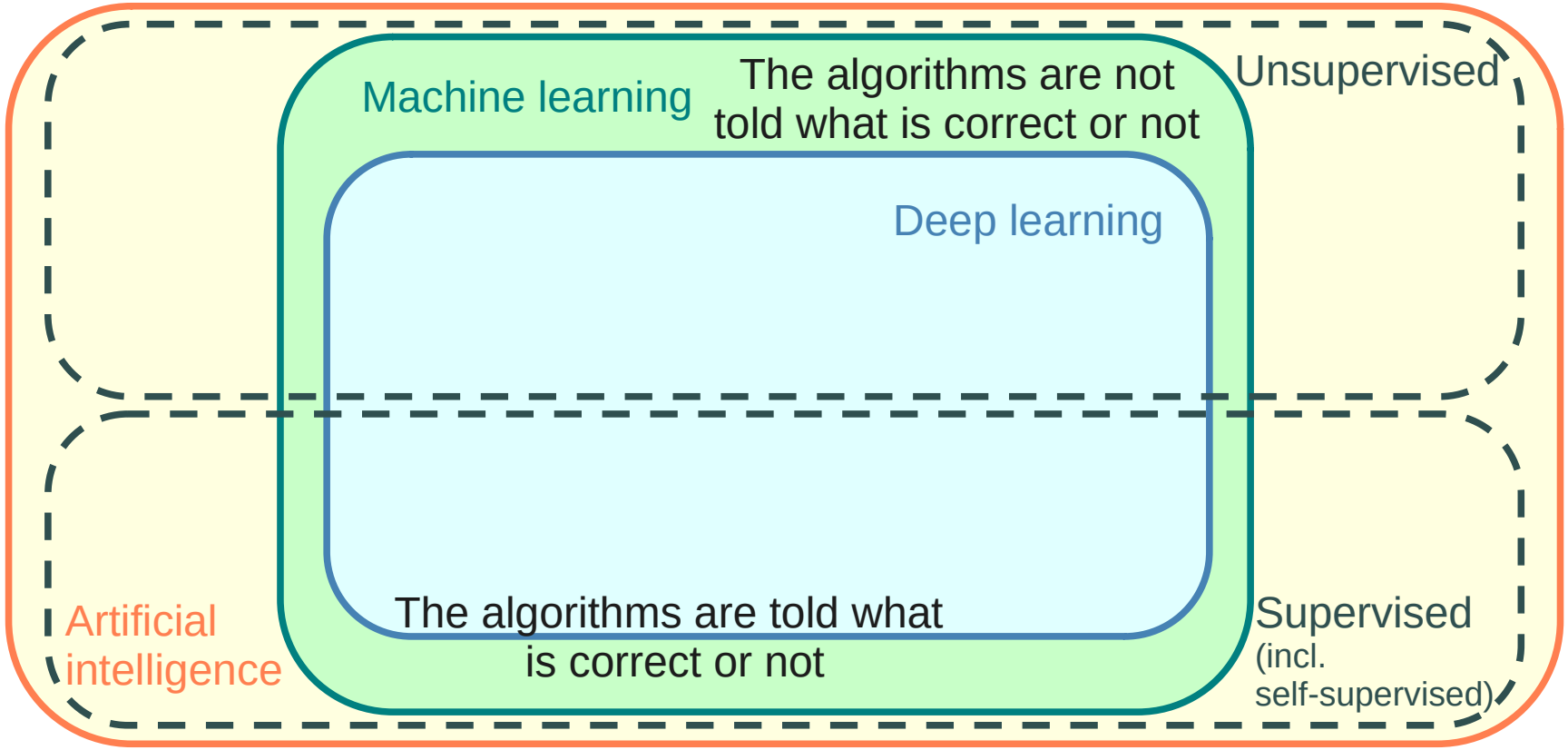
# Some terminology



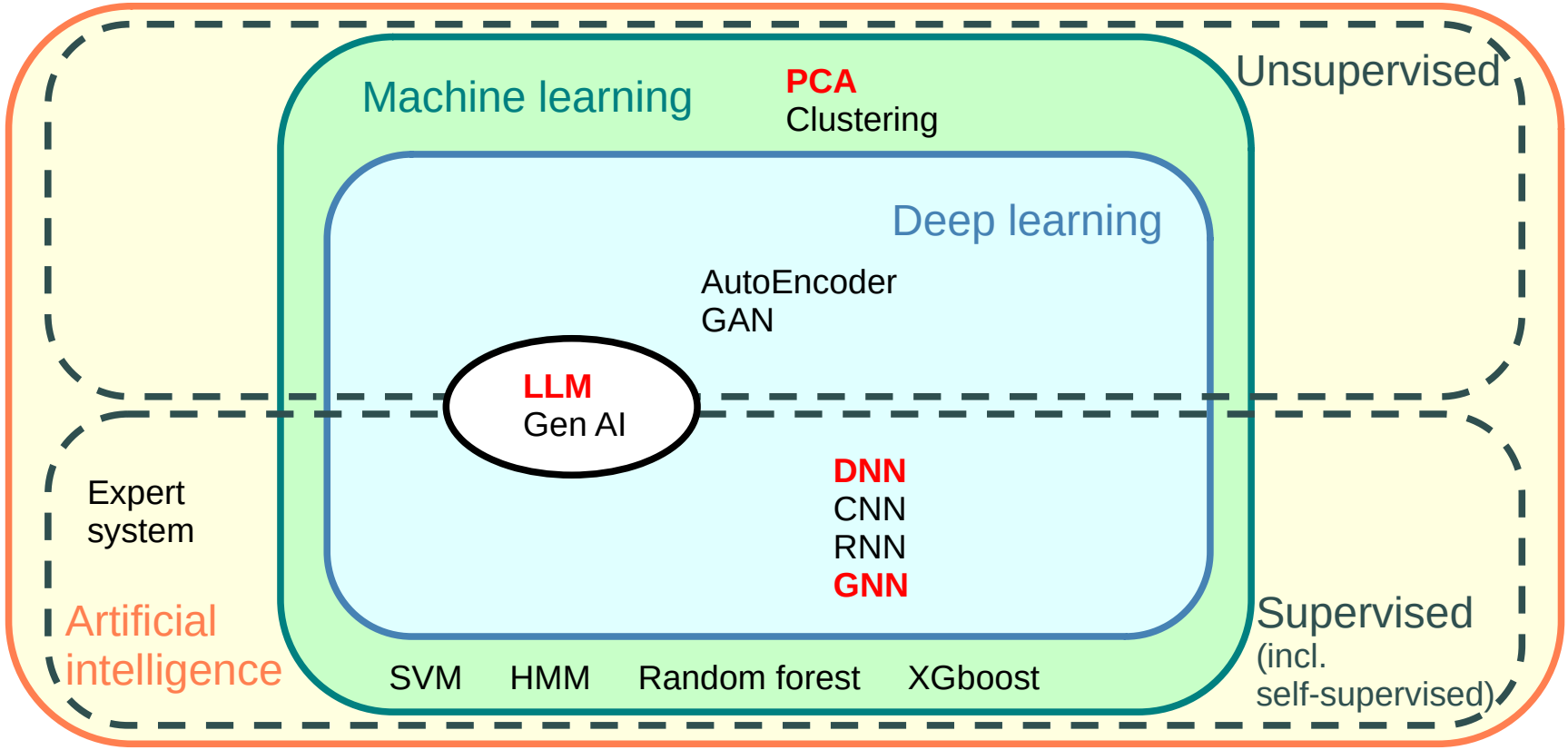
# Some terminology



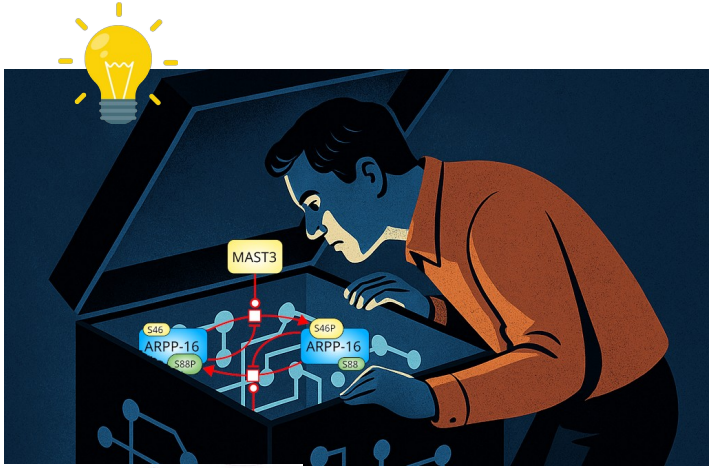
# Some terminology

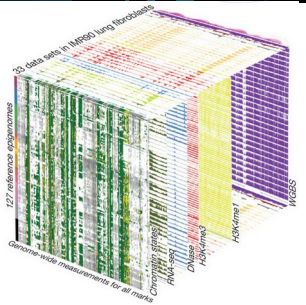


# Some terminology

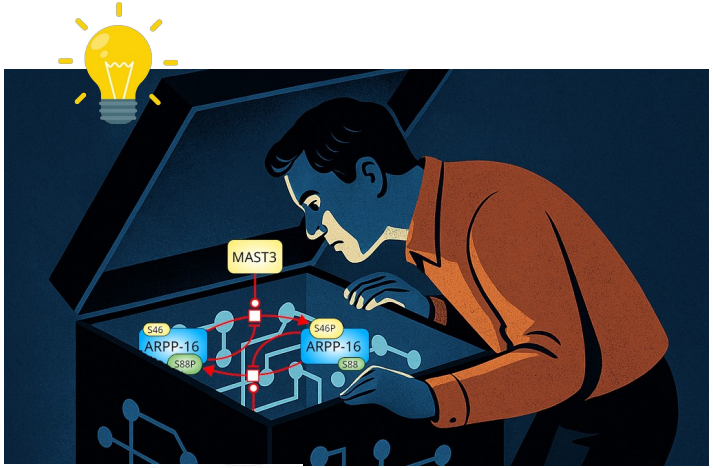


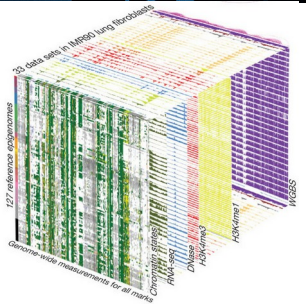
# Why should we use machine learning?

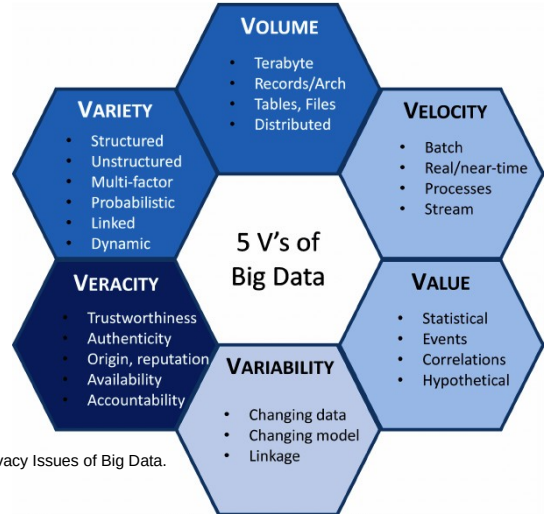


# Why should we use machine learning?

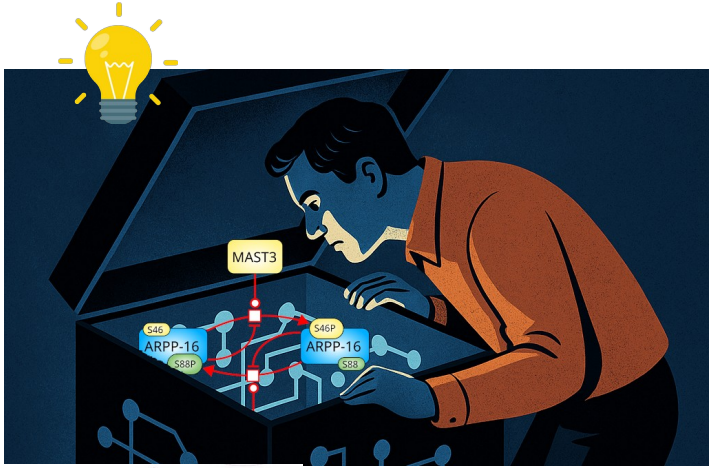


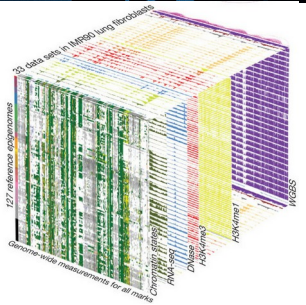



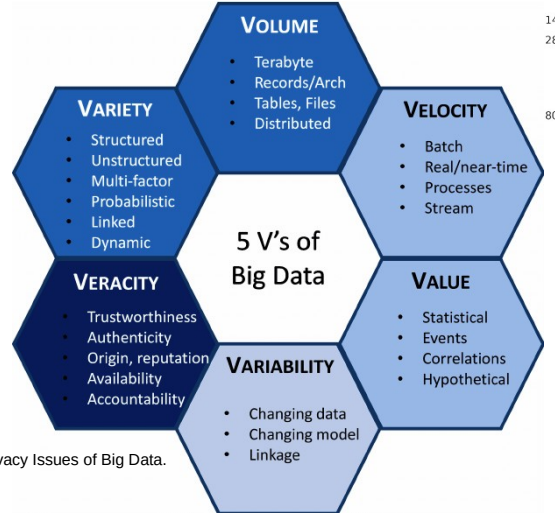



source: Moura, Serrão. Security and Privacy Issues of Big Data.
doi:10.4018/978-1-4666-8505-5.ch002

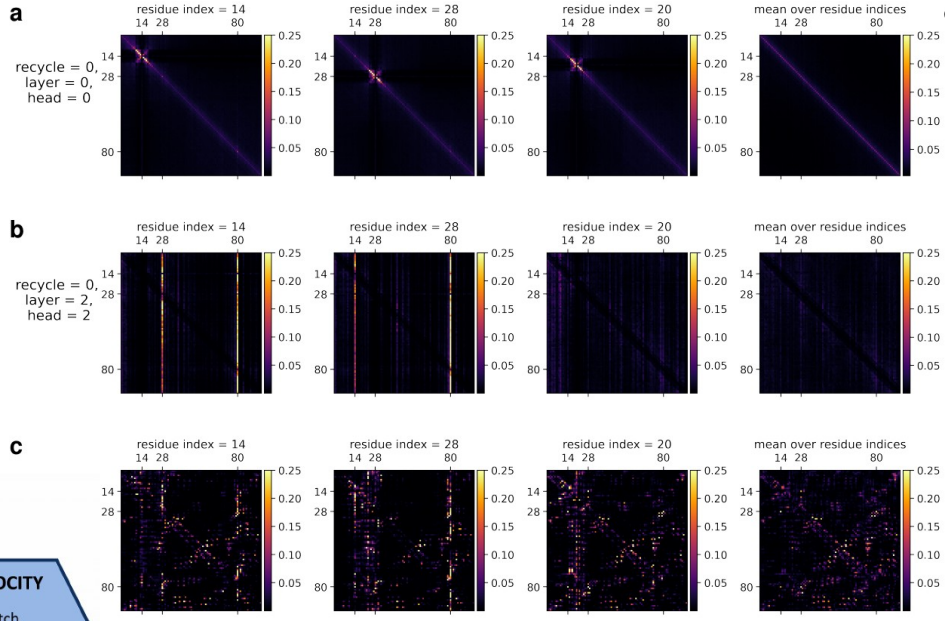
# Why should we use machine learning?



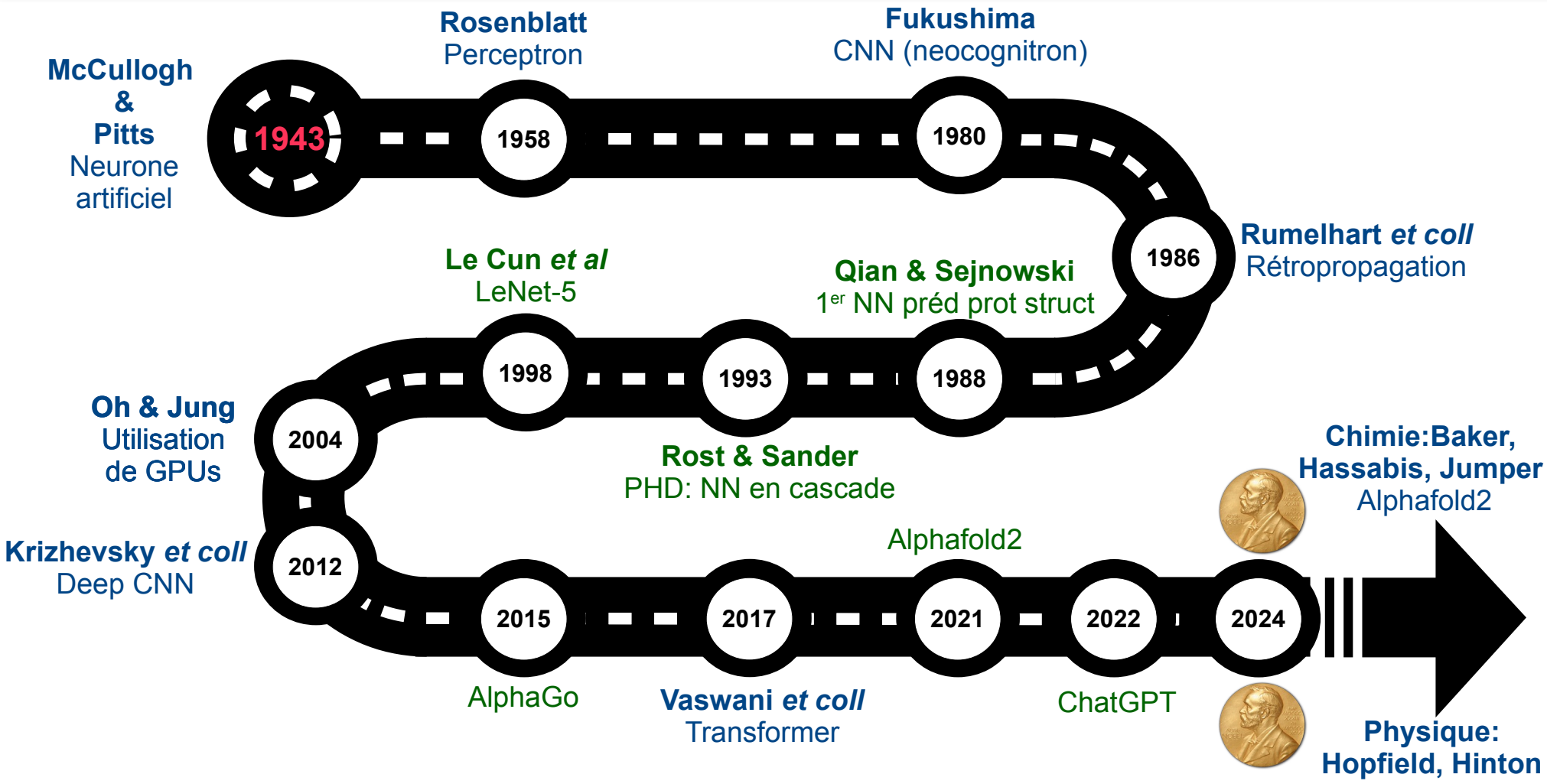



source: Moura, Serrão. Security and Privacy Issues of Big Data.
doi:10.4018/978-1-4666-8505-5.ch002

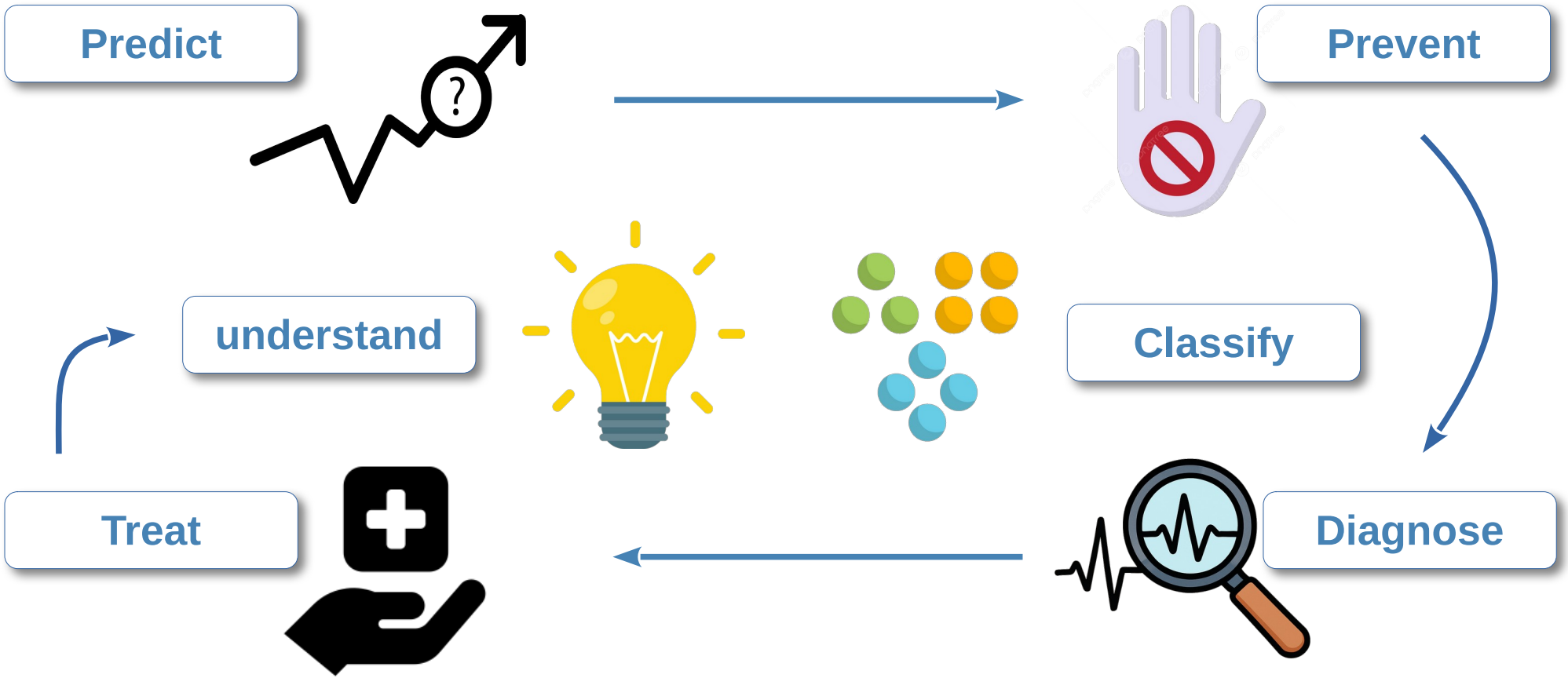



Source: Jumper *et al* (2021). Highly accurate protein structure prediction with Alphafold
doi:10.1038/s41586-021-03819-2 supplementary information

# Some history

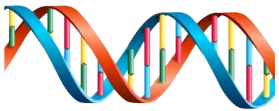


# AI in health, what for?



# Which data for AI?

**Genetic**



**From birth**

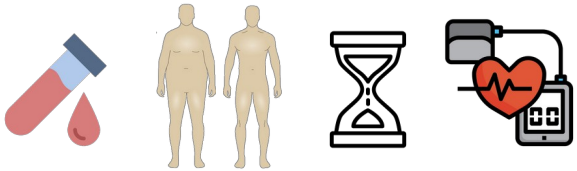
**Medical history**

**Environment**

**Accumulation**

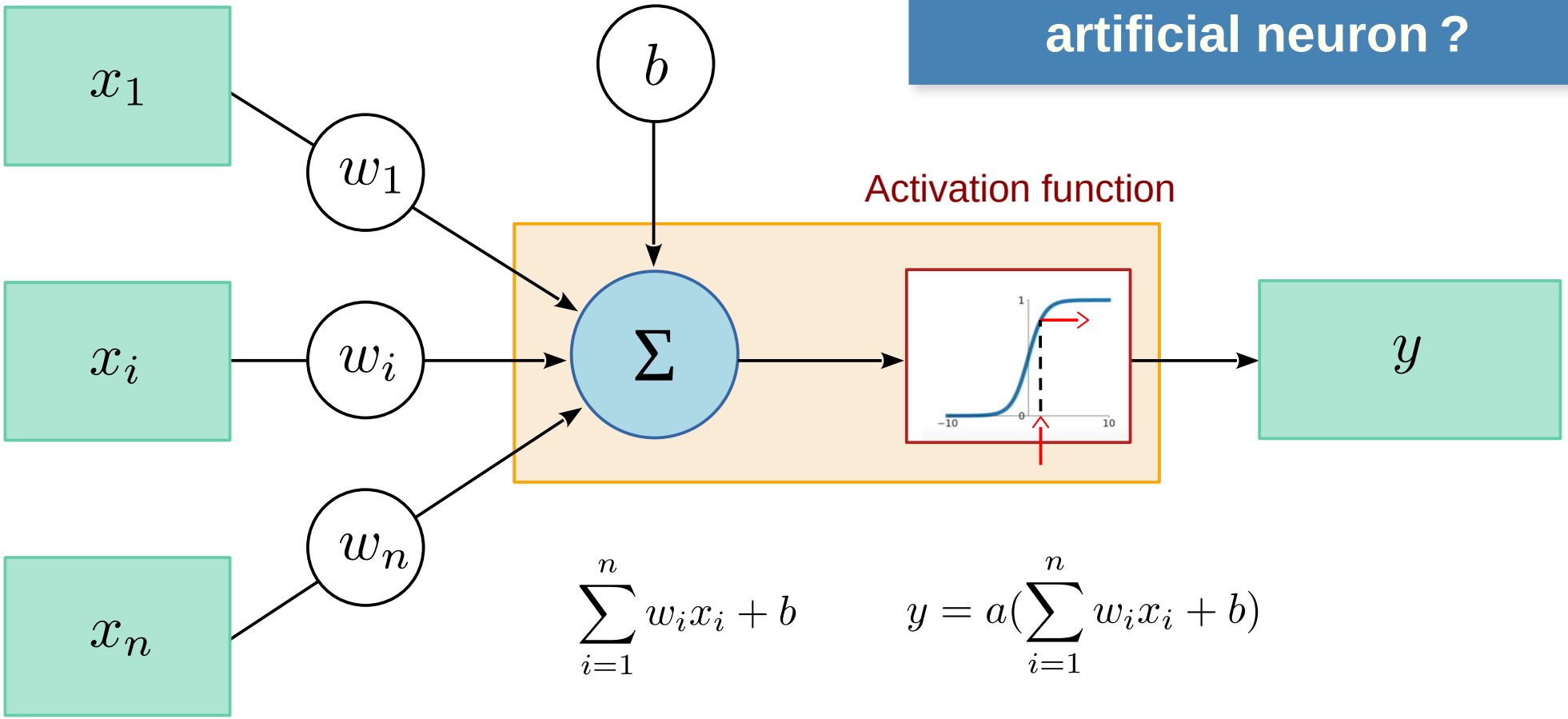
**Lifestyle**

**Clinical data**



**Instantaneous**
**Updated**

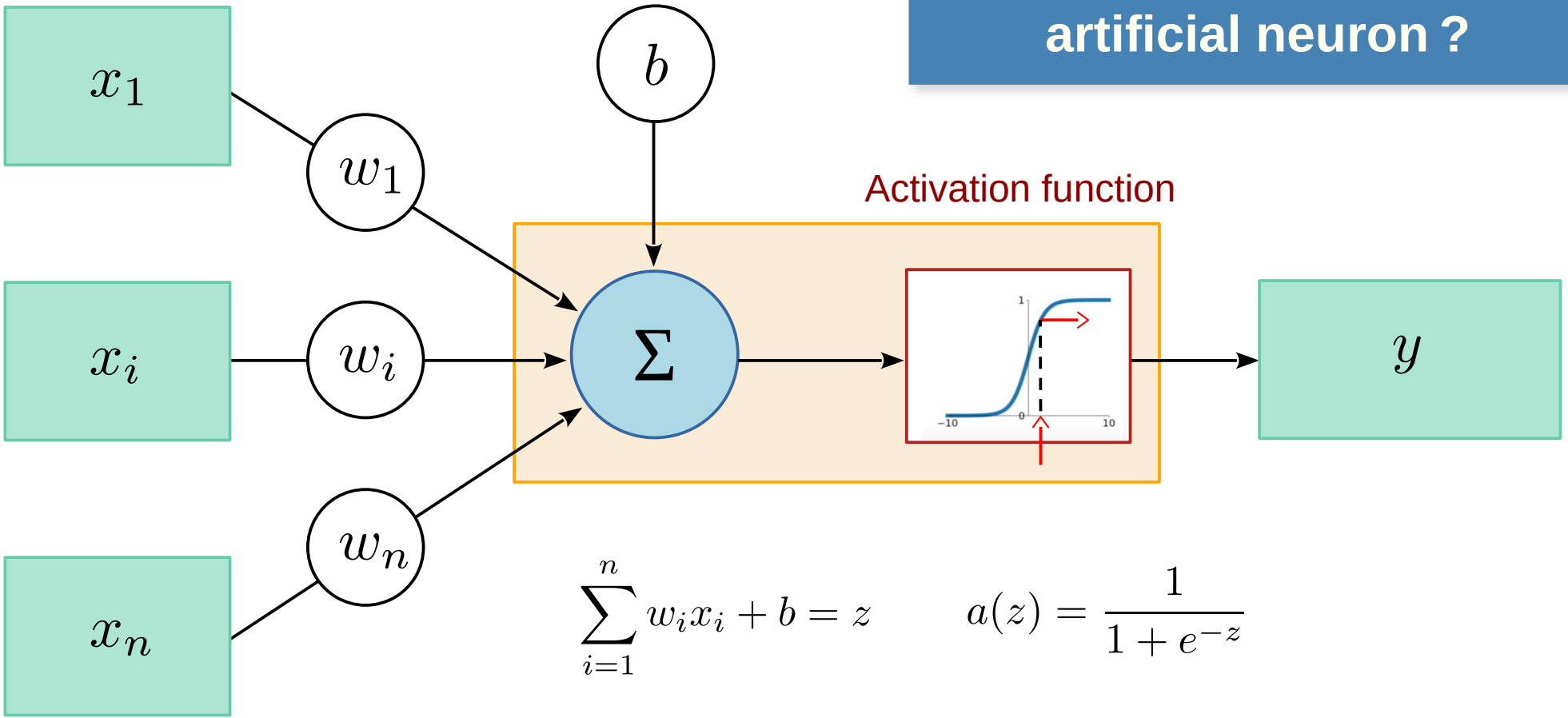
# What is an artificial neuron ?




$$\sum_{i=1}^{n} w_i x_i + b \qquad y = a\left(\sum_{i=1}^{n} w_i x_i + b\right)$$


McCulloch and Pitts (1943) A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5:115-133
Rosenblatt (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 65(6):386-408
Widrow and Hoff (1960) Adaptive Switching circuits. *WESCON Convention record* part IV: 96-104

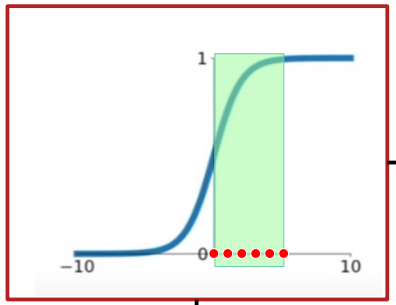
# What is an artificial neuron ?



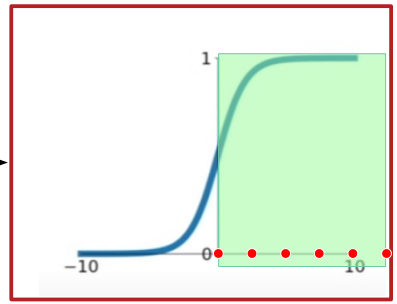

$$\sum_{i=1}^{n} w_i x_i + b = z \qquad a(z) = \frac{1}{1+e^{-z}}$$

NB: when the activation function is logistic (sigmoid), this is actually a logistic regression...

# Impact of the weights and the bias

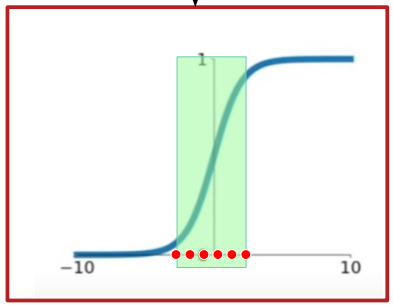



$w = 3$



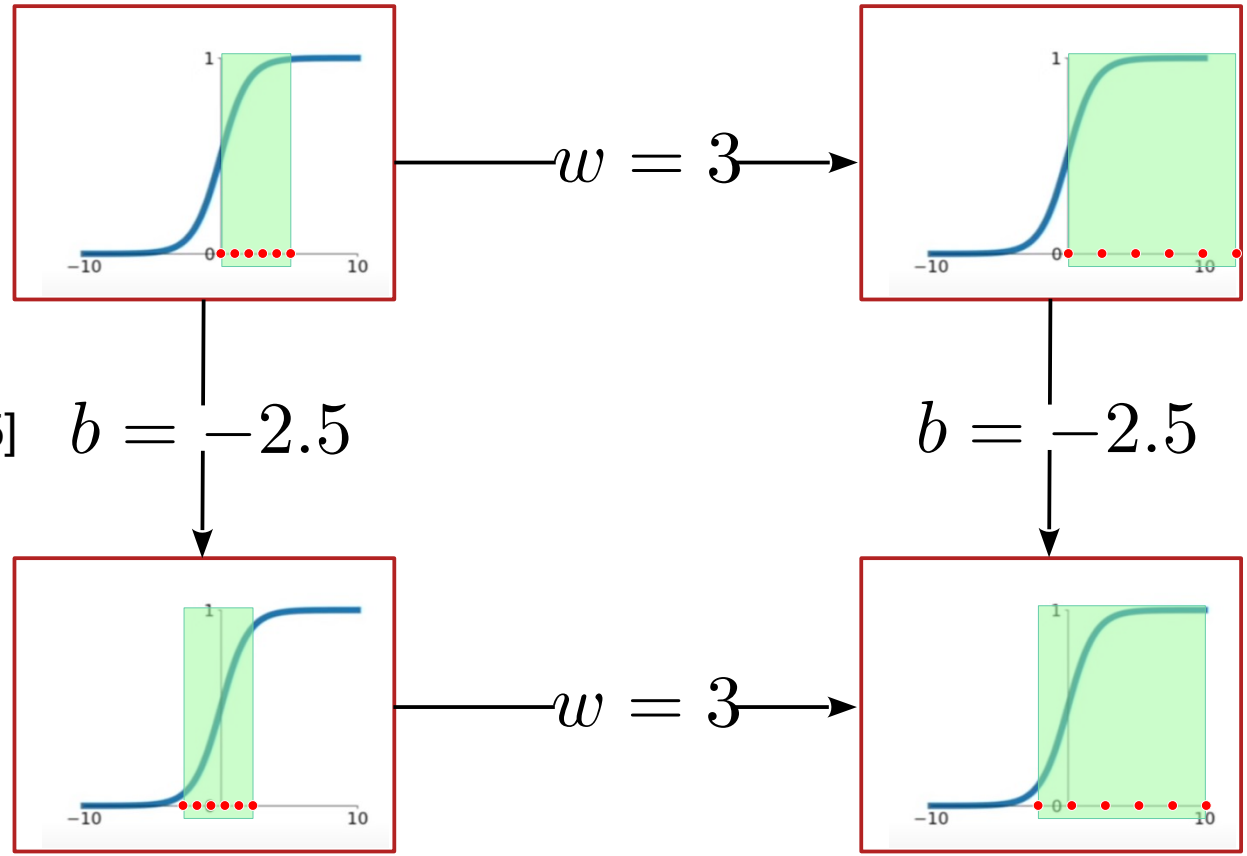

Successive inputs = [0,1,2,3,4,5]

$b = -2.5$

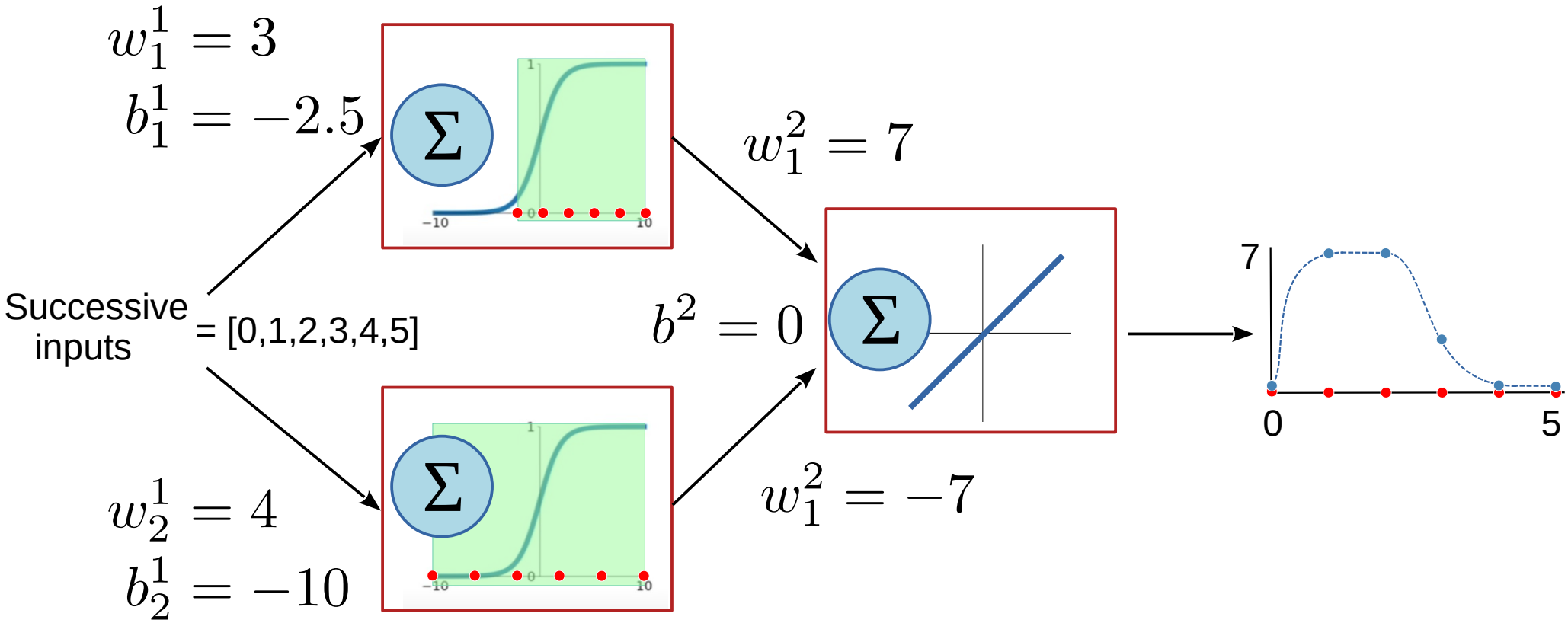


# Impact of the weights and the bias

Successive inputs = [0,1,2,3,4,5]

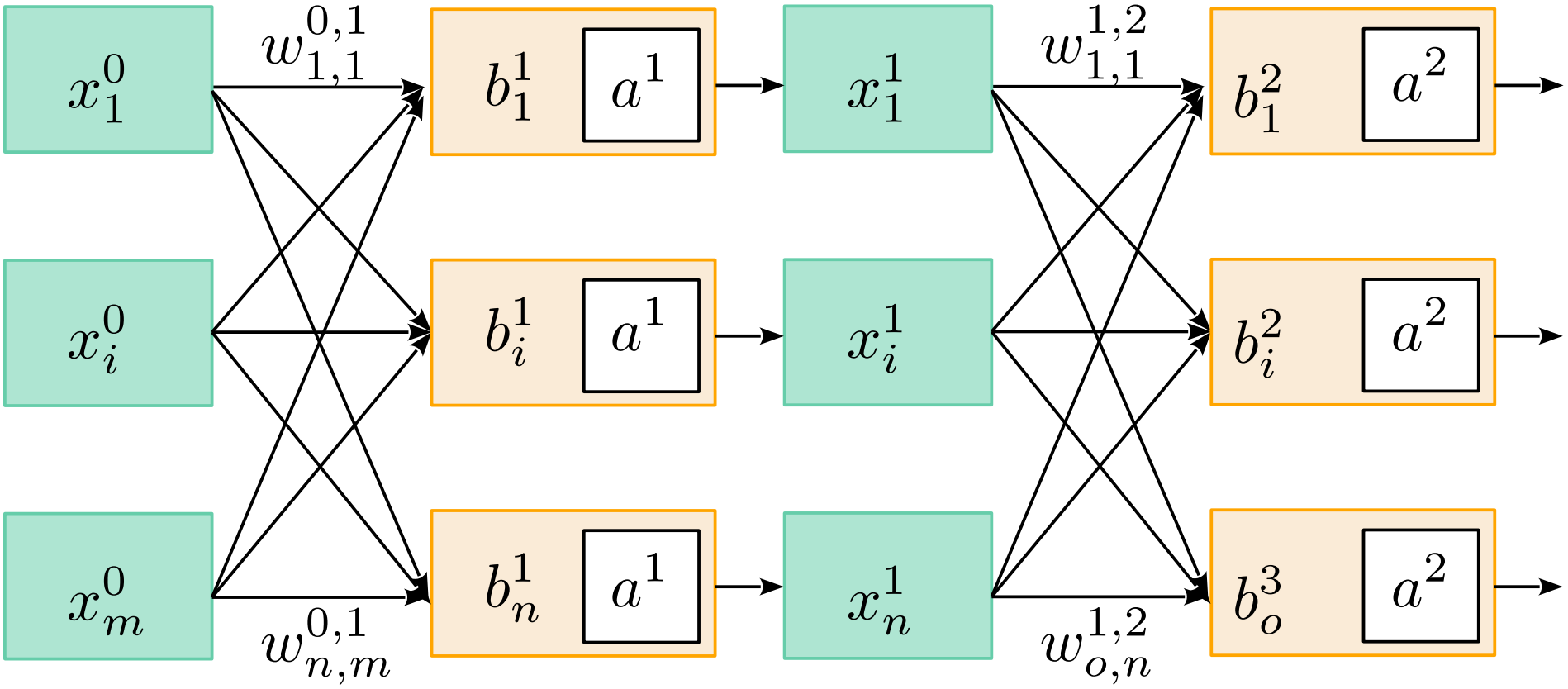


# The magic happens with several neurons



# The multi-layer perceptron

# And then we add layers (the “Deep”)




Rosenblatt (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 65(6):386-408

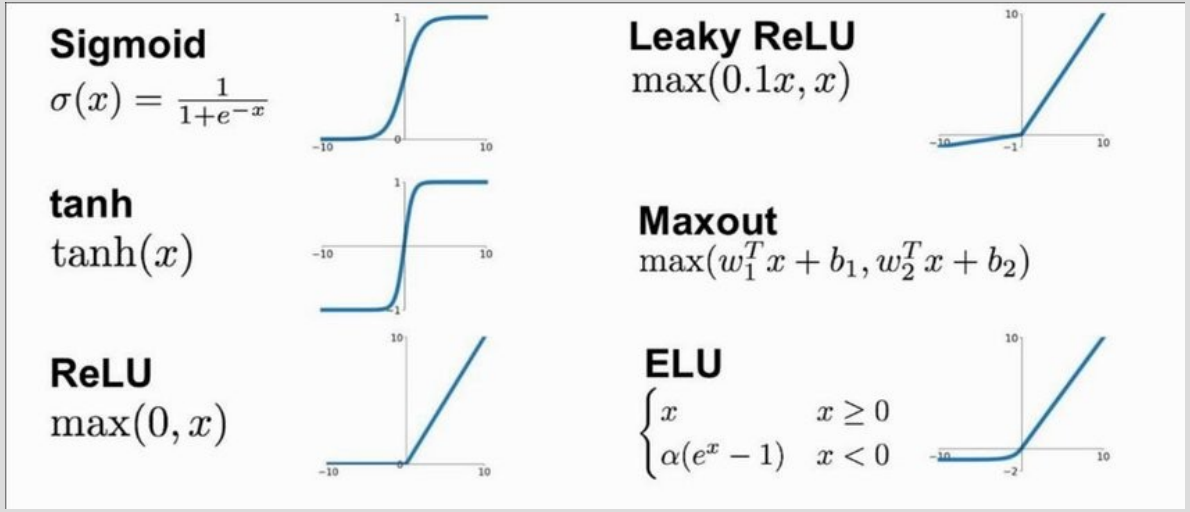
# Activation functions can be (almost) anything

$x_1^0$

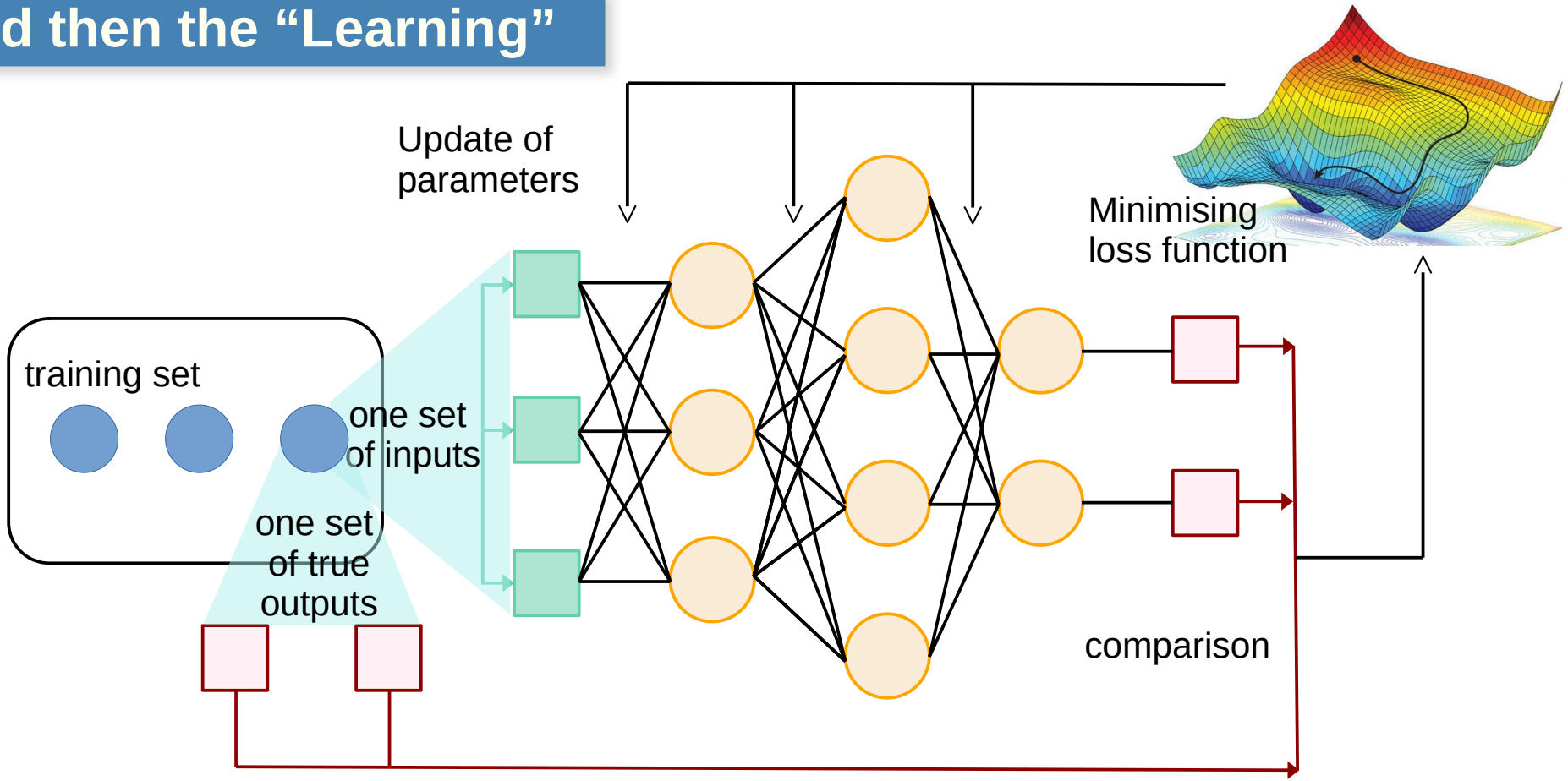
$x_i^0$

$x_m^0$

Activations on a given layer are the same,
but can be different on different layers

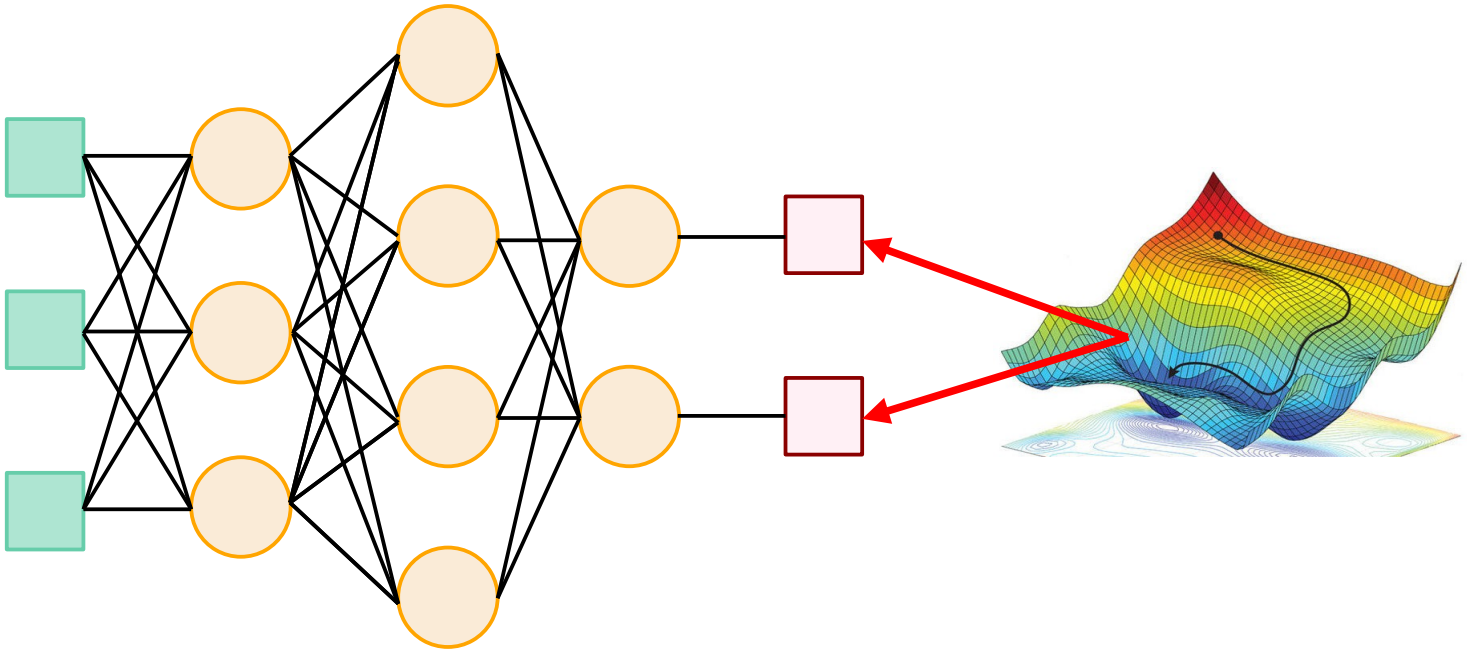


# And then the “Learning”

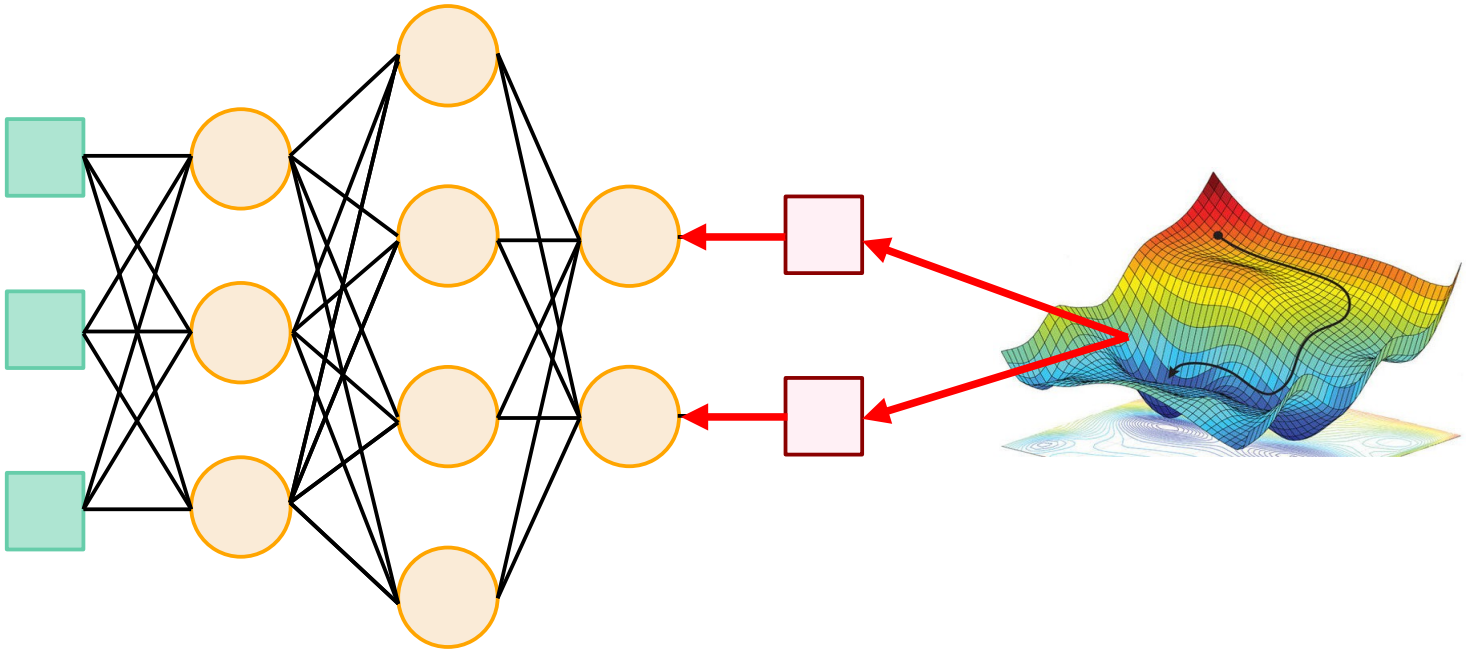



Widrow and Hoff (1960) Adaptive Switching circuits. *WESCON Convention record* part IV: 96-104; S Amari (1967). A theory of adaptive pattern classifier. *IEEE Transactions*. EC (16): 279–307; S Linnainmaa (1970-1976). The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors (Masters). University of Helsinki. p. 6–7; P Werbos (1971-1982) Applications of advances in non-linear sensitivity analysis. *LNCIS* 38: 762-770
LeCun Y (1985) Une procédure d'apprentissage pour réseau à seuil asymétrique. *Proc Cognitiva* 85, 599-604.
Rumelhart, Hinton, and Williams (1986) Learning representations by back-propagating errors." *Nature* 323(6088): 533-536.

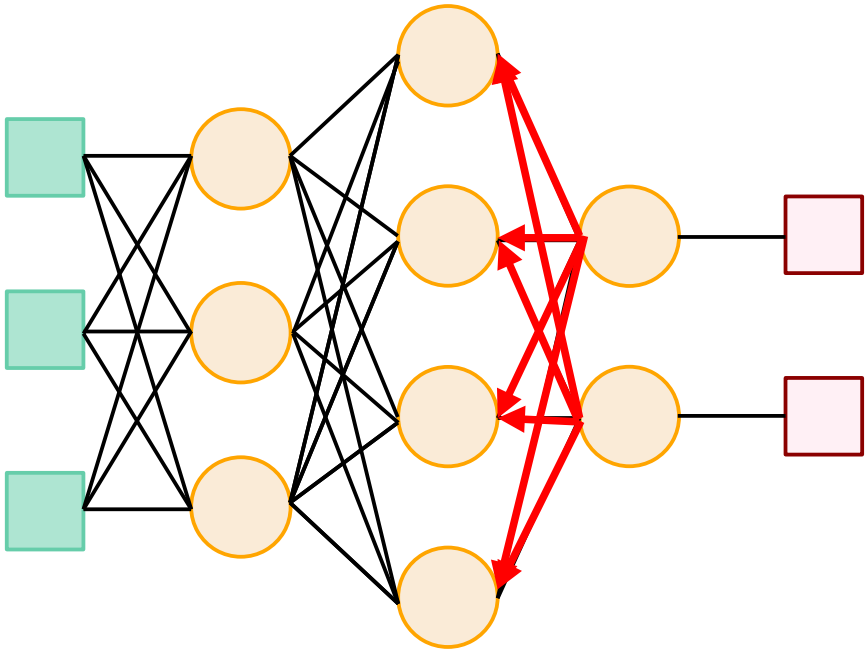
# Error backpropagation one layer at a time

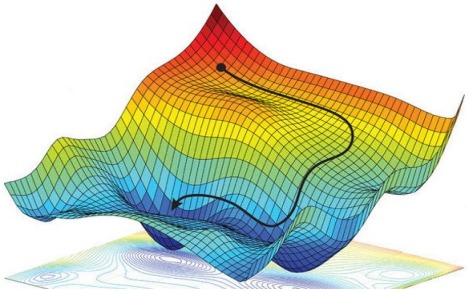


# Error backpropagation one layer at a time

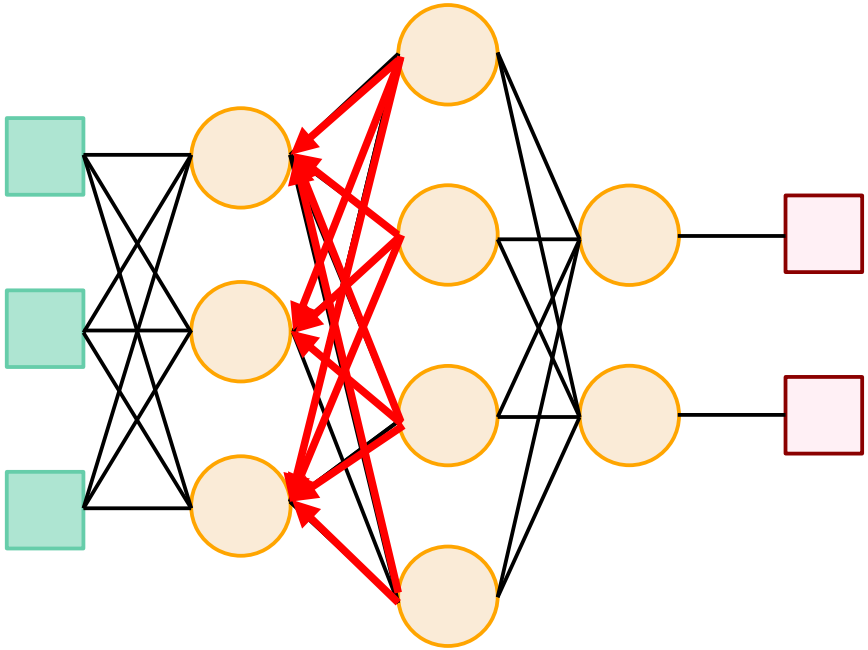


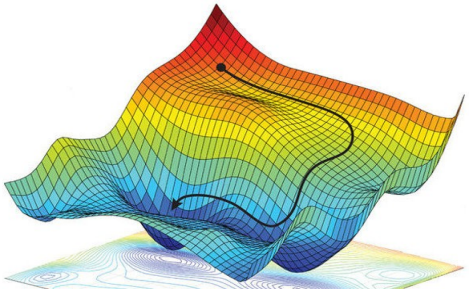
# Error backpropagation one layer at a time



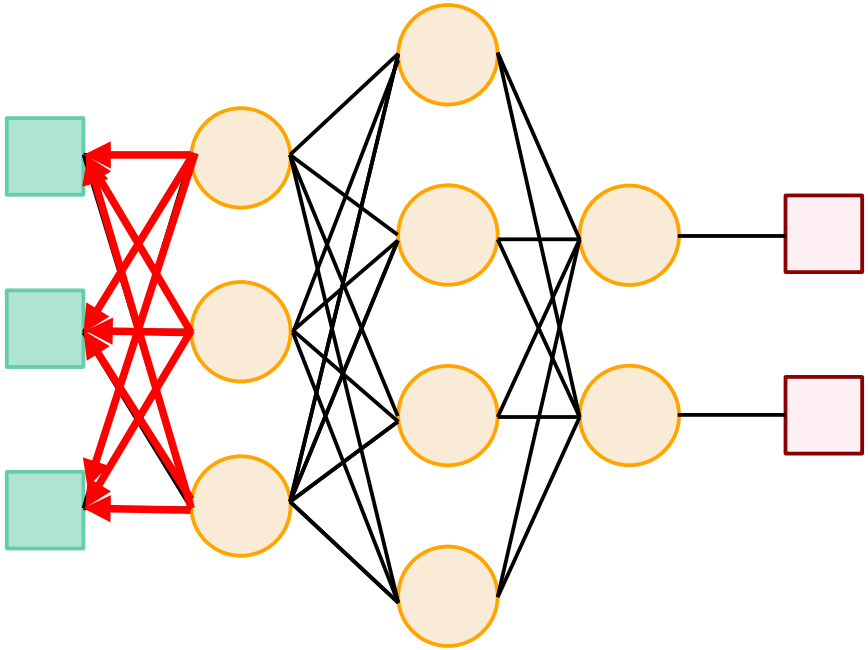


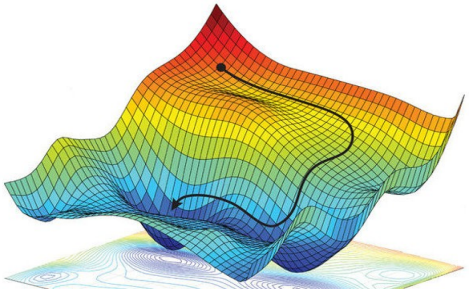
# Error backpropagation one layer at a time





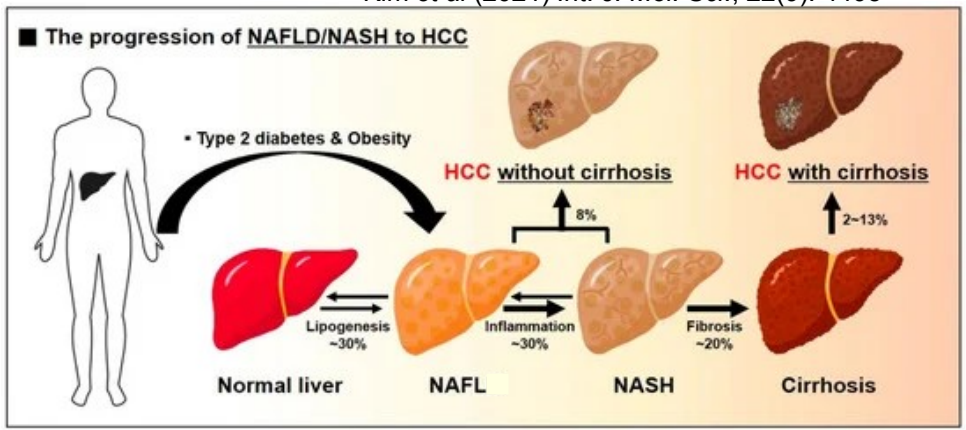
# Error backpropagation one layer at a time

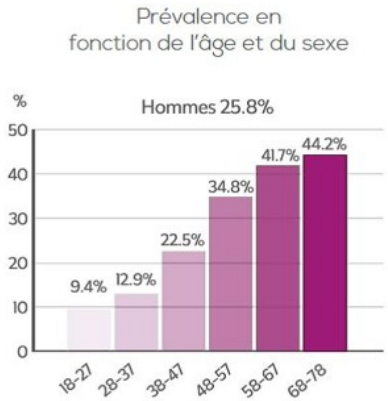


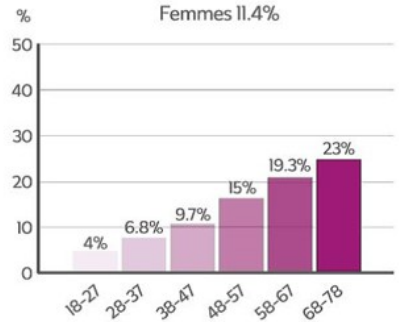


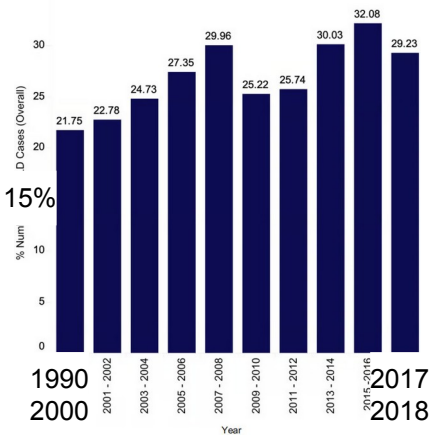
# Let’s try to recognise the severity of a disease: MASLD

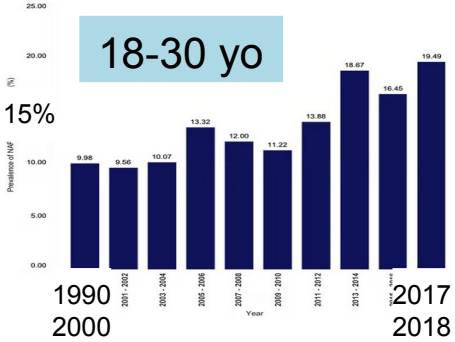
Kim *et al* (2021) *Int. J. Mol. Sci.*, 22(9): 4495



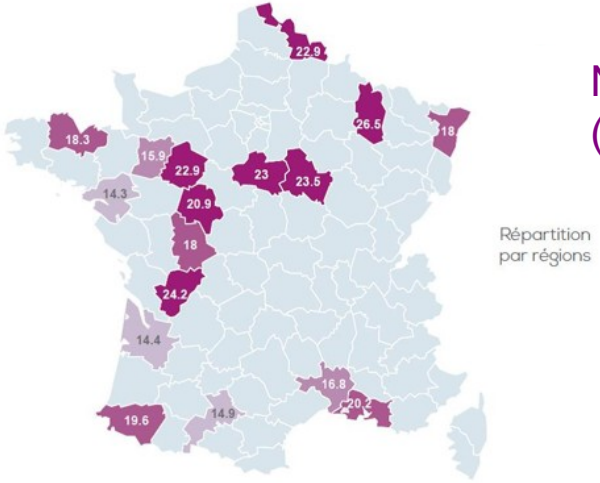




Kim *et al* (2022) *Met. Target Organ Damage*, 2: 19

NALFD (now MASLD)

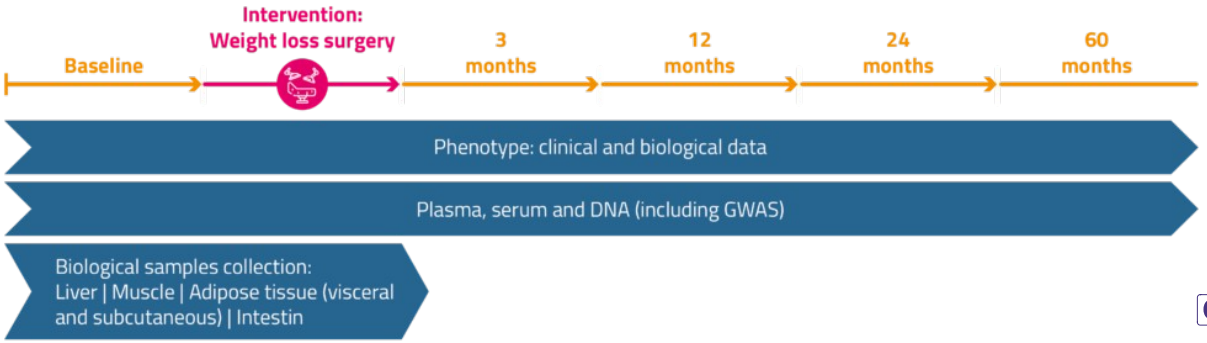



NASH
(now MASH)

Paris MASH Meeting
(11-12 juillet 2019)

PreciNASH
recherche hospitalo-universitaire

# ABOS and PreciNASH

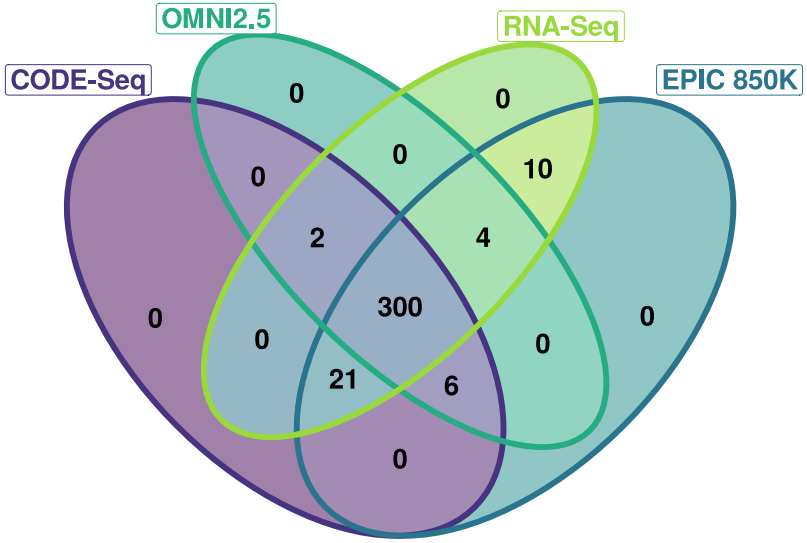



**ABOS (Biological Atlas of Severe Obesity)**

All subjects had bariatric surgery

**PreciNASH project**

ABOS subset: Only European ancestry and unrelated individuals

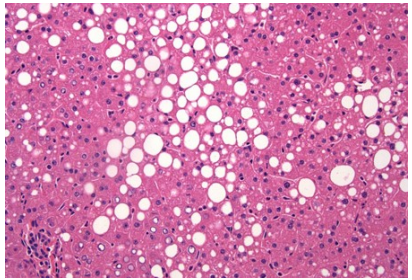



(+66 clinical and personal data
+1076 identified metabolites
in blood and liver)

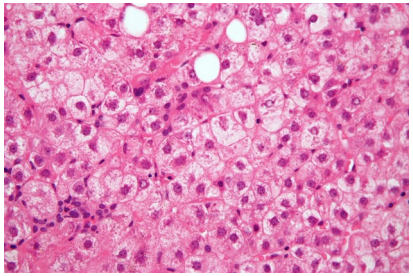
# Subject grouping

Scoring on liver biopsy with the method from Kleiner and Brunt 2005

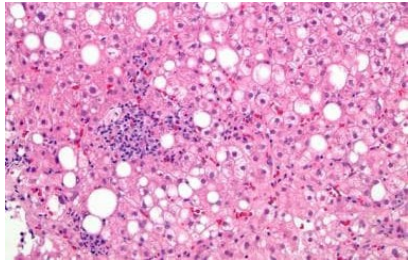
**Steatosis**
Categorical [0-3] from quantitative measurement



**Ballooning**
Categorical [0-2]
= {none, some, much}



**Inflammation**
Categorical [0-3] from number of foci



Final score:

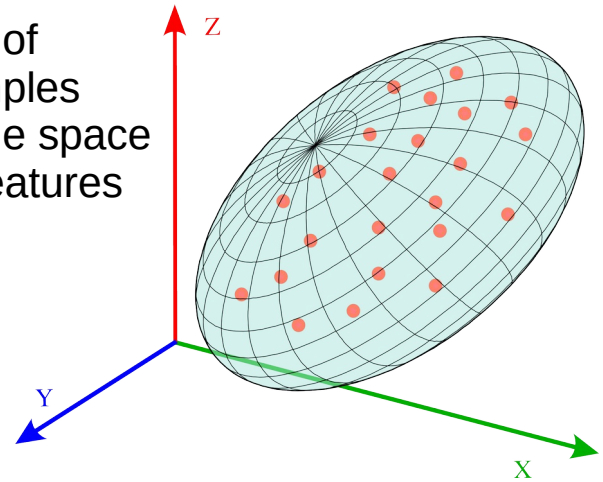
**Healthy**: $S = 0$, $B = 0$, $I = 0$ $n = 80$

**NAFL**: $S > 1$, $B = 0$, $I \geq 1$ $n = 137$
$S > 1$, $B > 1$, $I = 0$

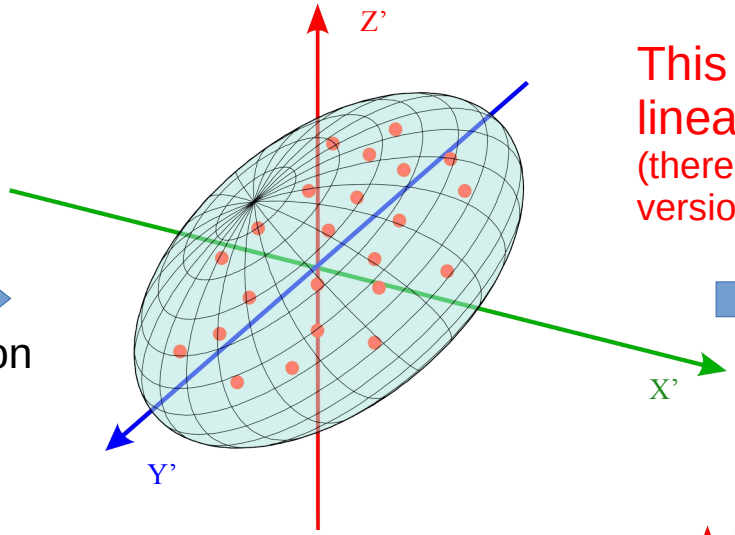
**NASH**: $S > 0$, $B > 0$, $I > 0$ $n = 83$

# Principal component analysis (PCA)

plot of samples in the space of features

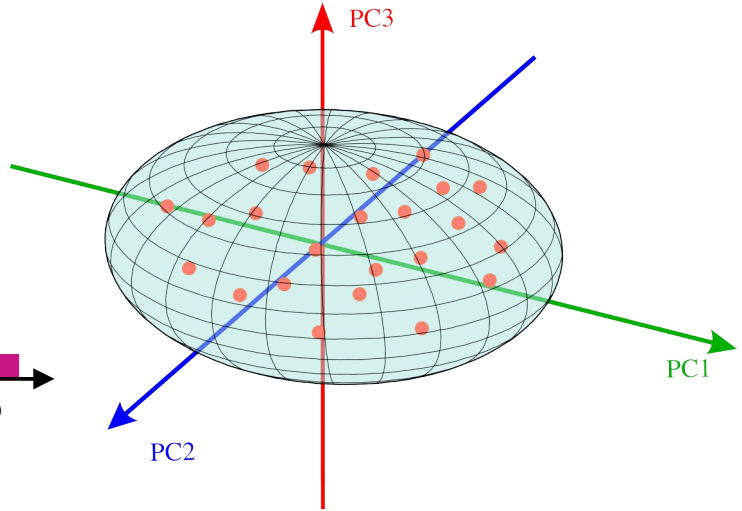



normalisation

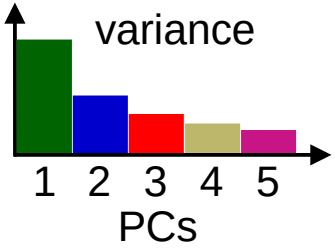



This is a linear transformation!
(there are non-linear versions, e.g. kernel PCA)

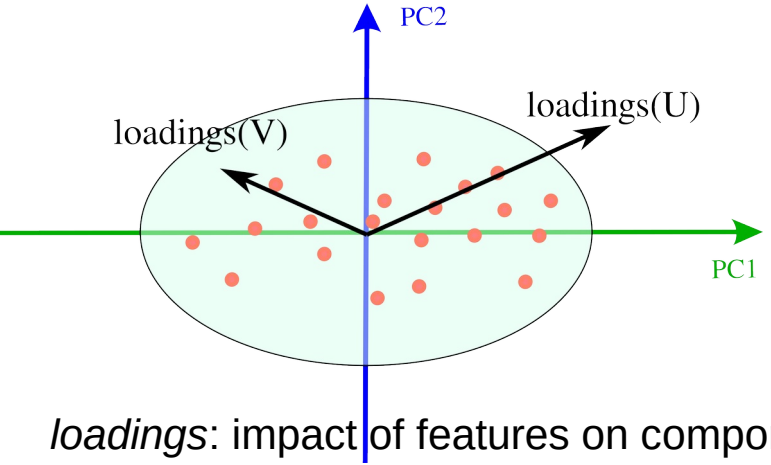
"rotation"




variance

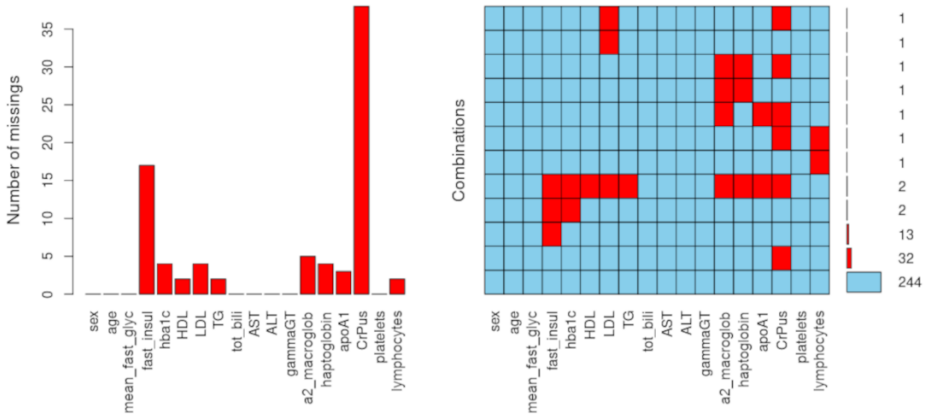


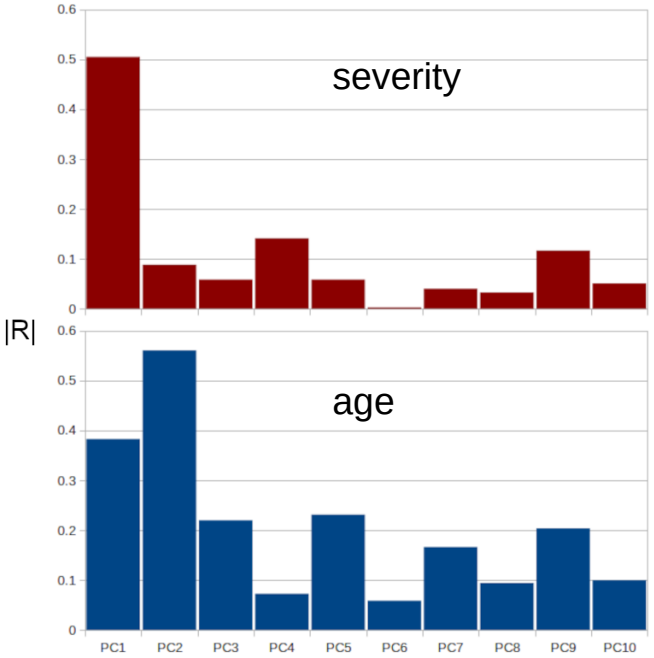

projection



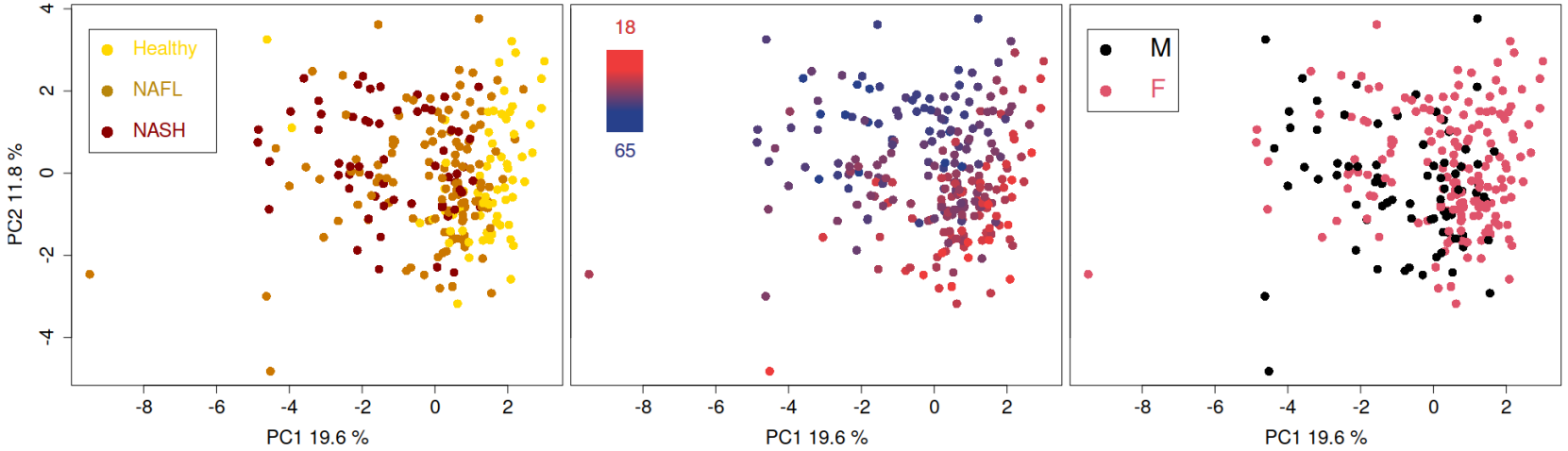

*loadings*: impact of features on components

# Clinical data

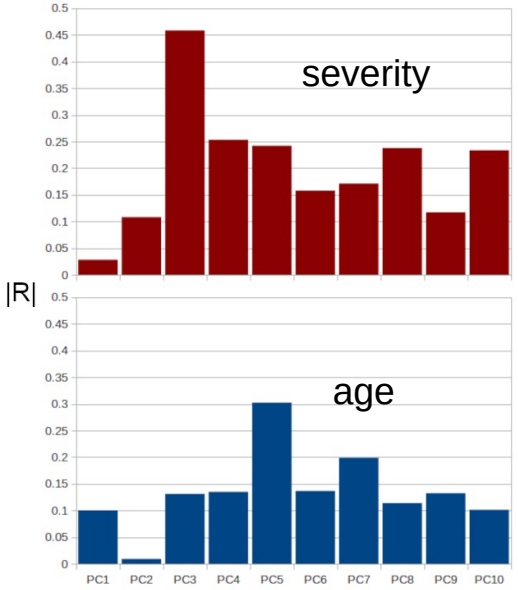


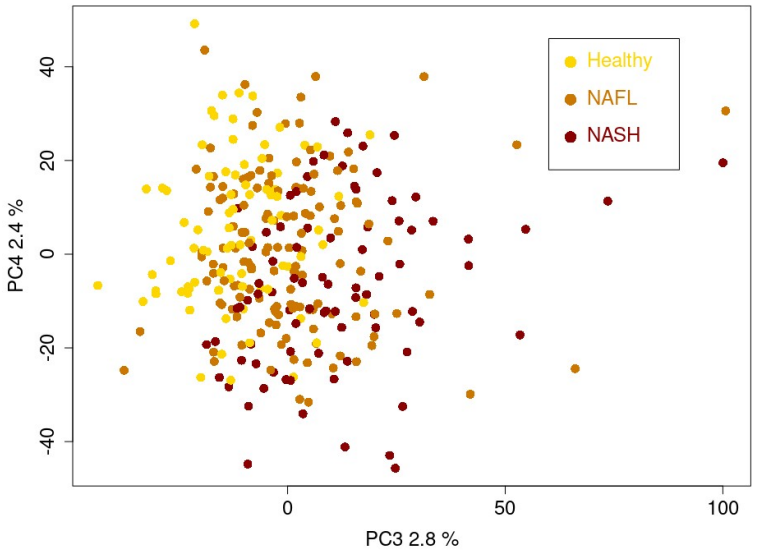




16 clinical variables
+ sex
+ age



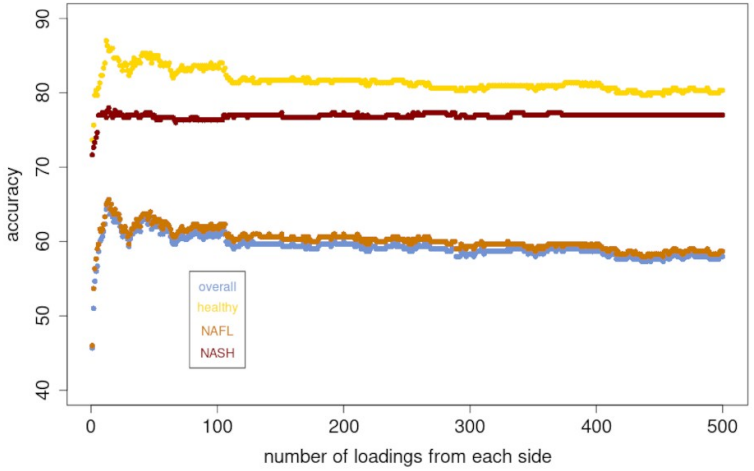
# RNA-seq

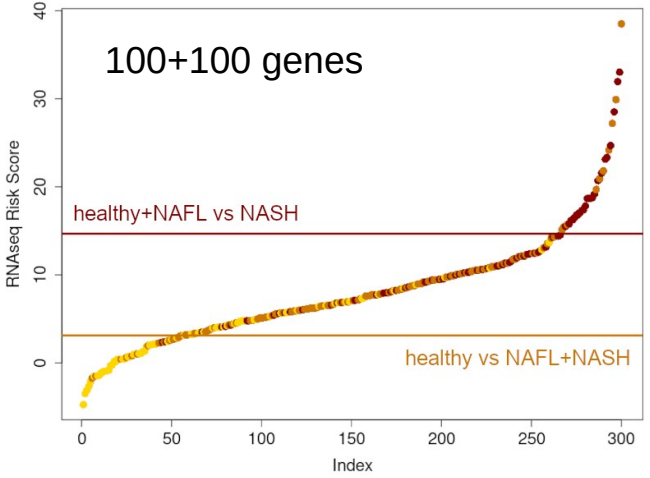


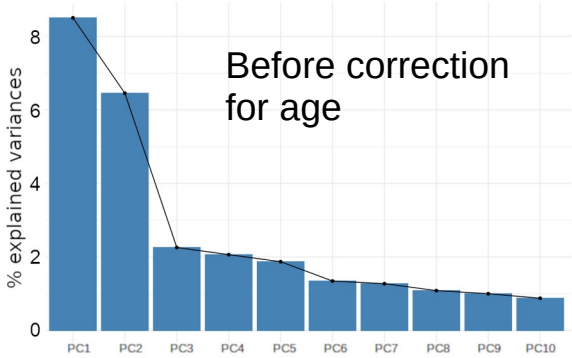

Score based on gene expression and gene “loadings” (impact of a gene on a given principal component)

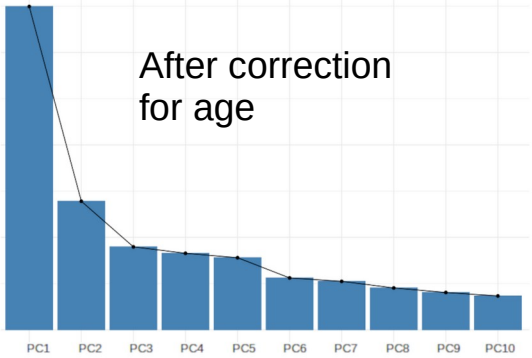
Logistic regression to find the thresholds best separating the severity groups

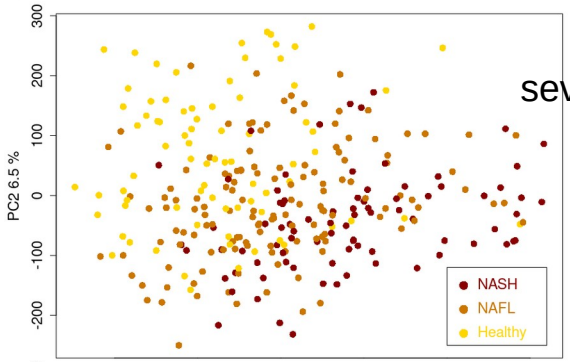


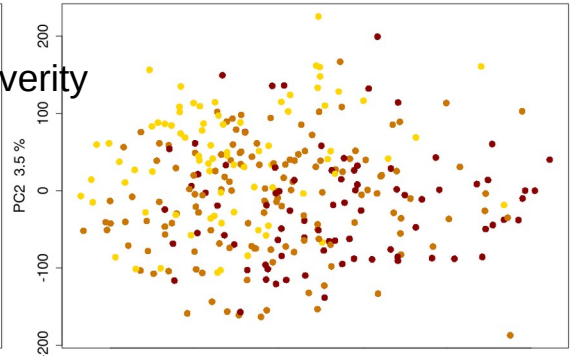



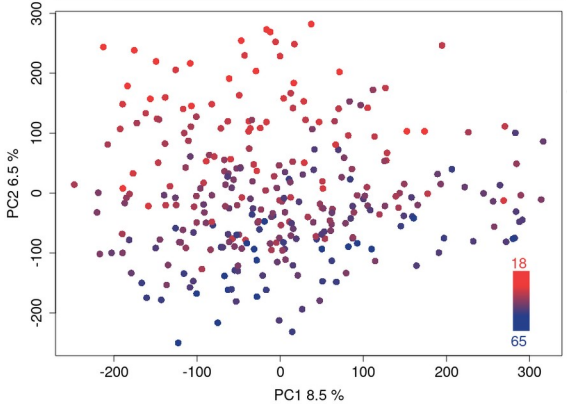
# DNA Methylation

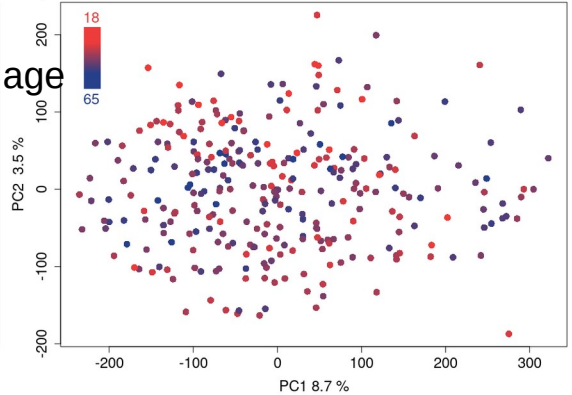


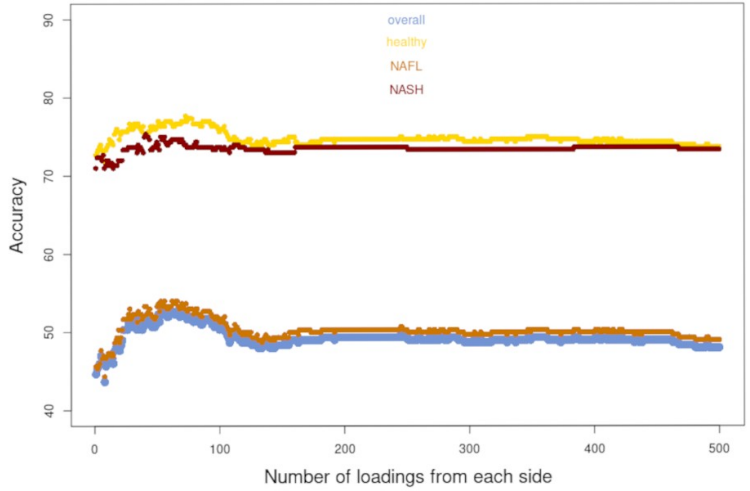



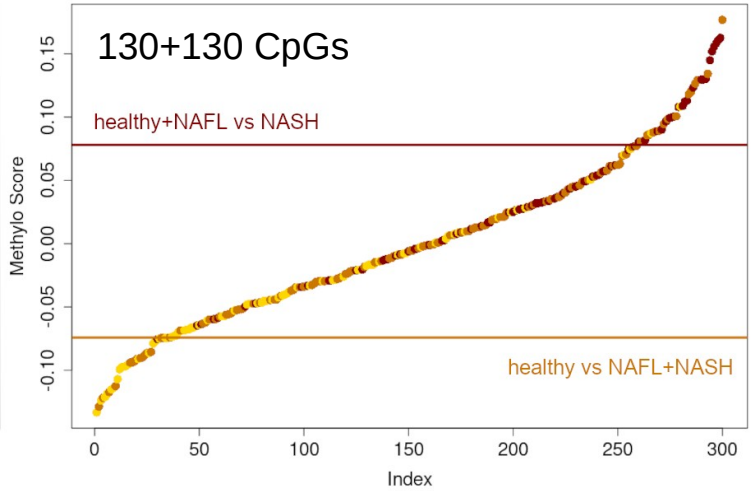



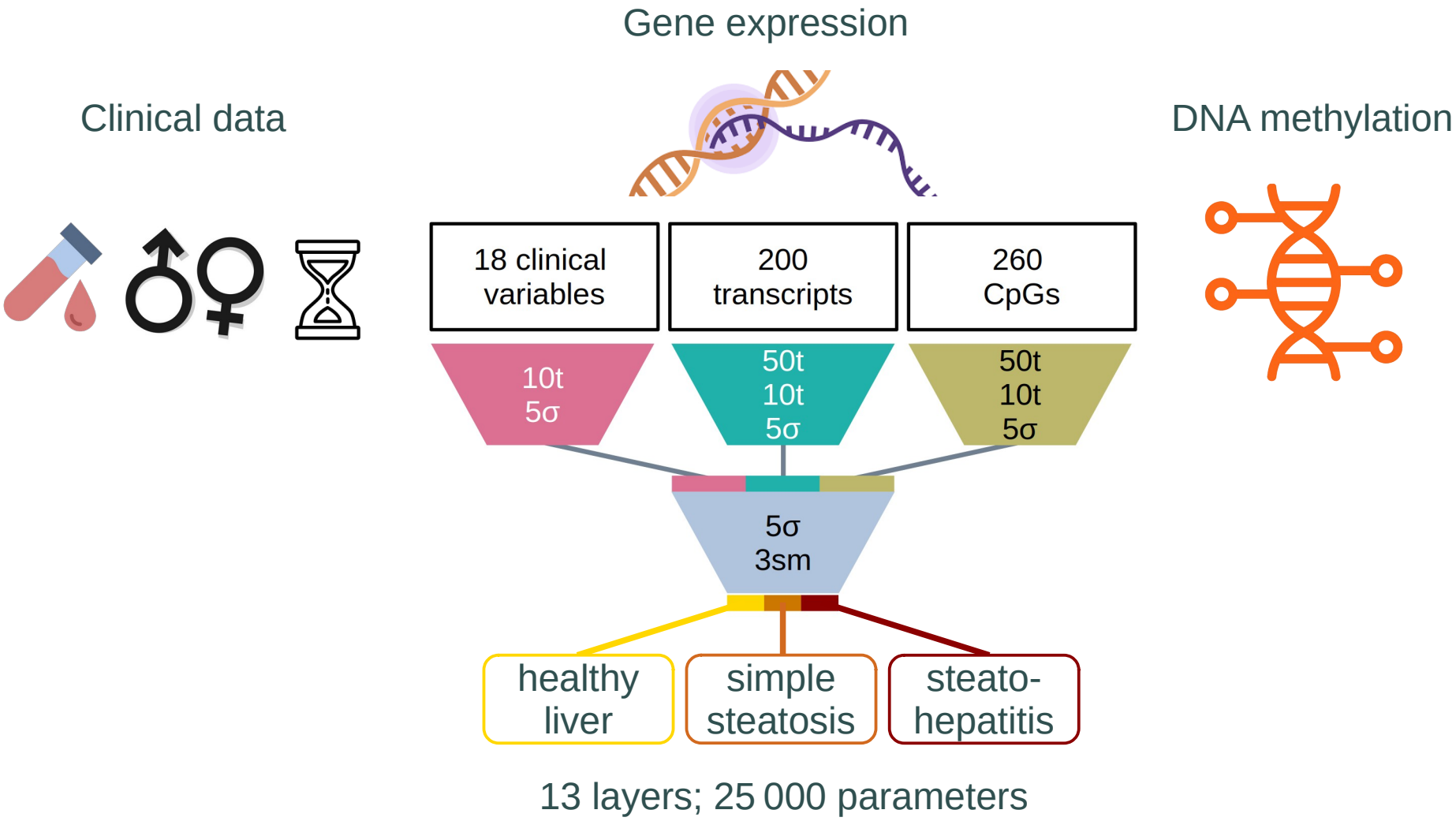



# Deep learning network reading clinical and omics data

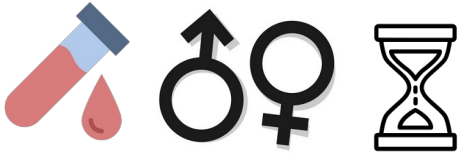


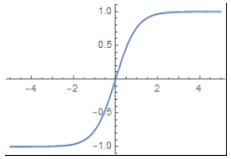
# Deep learning network reading clinical and omics data

Gene expression

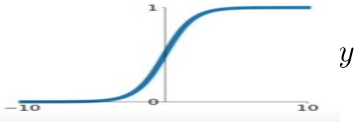
Clinical data

DNA methylation

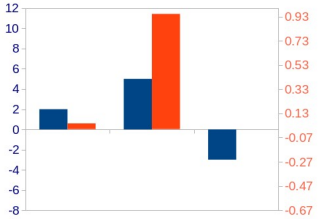




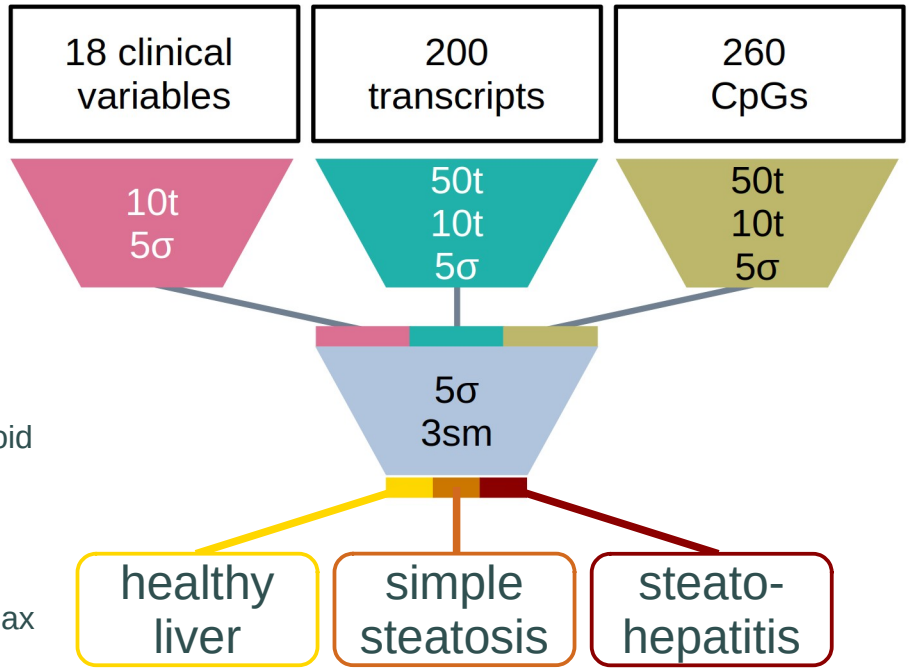
$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$ Tanh



$$y = \frac{1}{1 + e^{-x}}$$ Sigmoid



$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}$$ Softmax



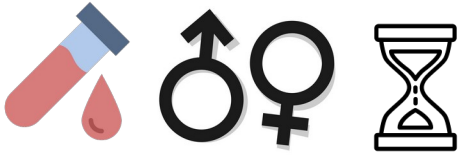

13 layers; 25 000 parameters

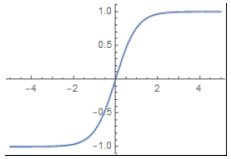
# Deep learning network reading clinical and omics data

Gene expression

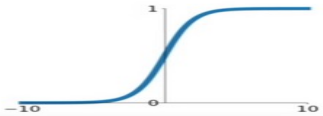
Clinical data

DNA methylation

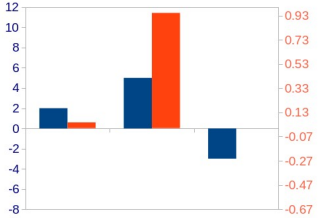




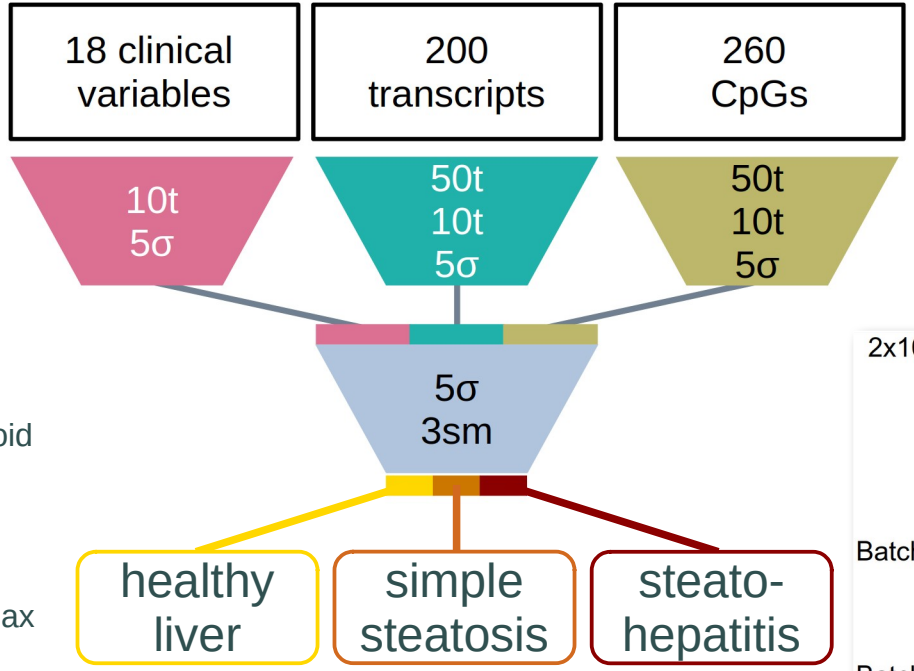
$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$ Tanh

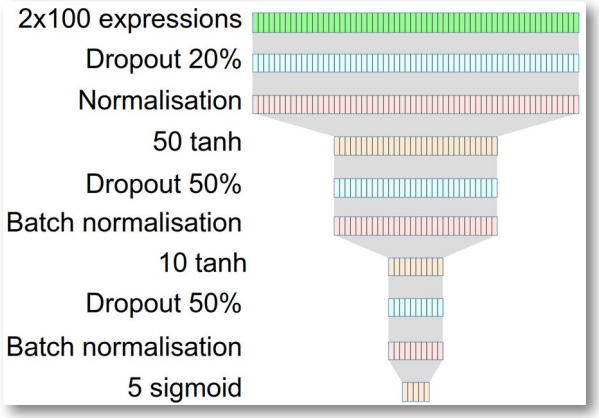


$$y = \frac{1}{1 + e^{-x}}$$ Sigmoid



$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}$$ Softmax



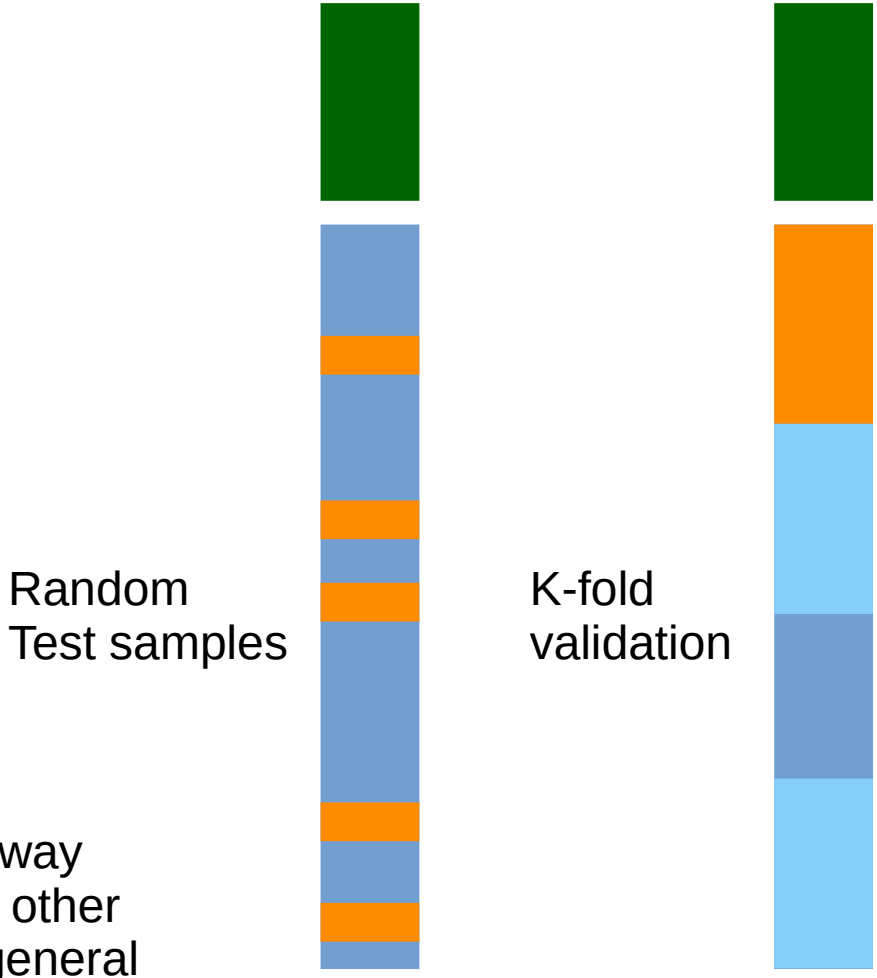

13 layers; 25 000 parameters

# Training, testing, and validation sets

"validation" (never seen)
Same for all model instances
Used to assess the model at the end

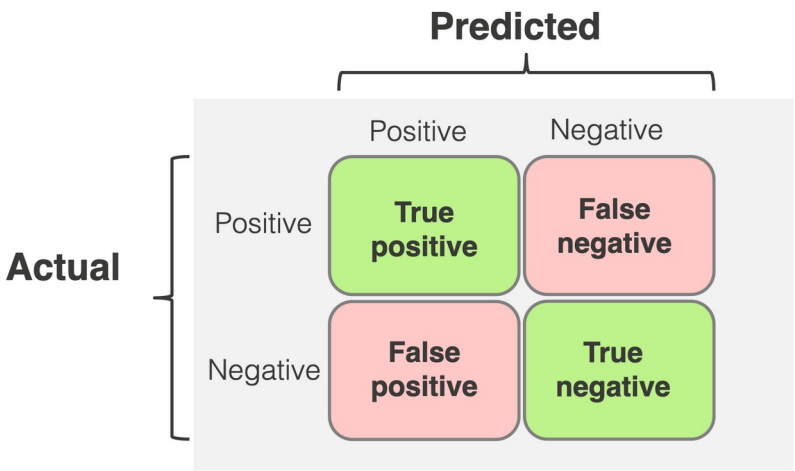
Training set: used to learn

"test" set: used to assess the model during the learning phase
Different for each model instance




**<u>Beware</u>**: "validation" and "test" are used the other way around a lot in deep learning, at the opposite of all other fields of machine learning, or even life science in general

# Evaluating a model's performance




Accuracy = (TP+TN)/(TP+FN+TN+FP)

Precision = TP/(TP+FP)

Sensitivity (true positive rate) = TP/(TP+FN)

Specificity (true negative rate) = TN/(TN+FP)

# Evaluating a model's performance

| | | Predicted: Positive | Predicted: Negative |
|---|---|---|---|
| Actual | Positive | True positive | False negative |
| | Negative | False positive | True negative |

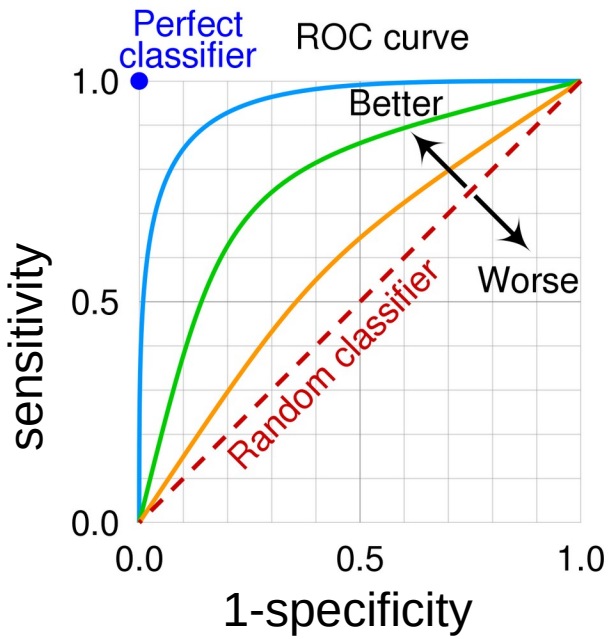
Accuracy = (TP+TN)/(TP+FN+TN+FP)

Precision = TP/(TP+FP)

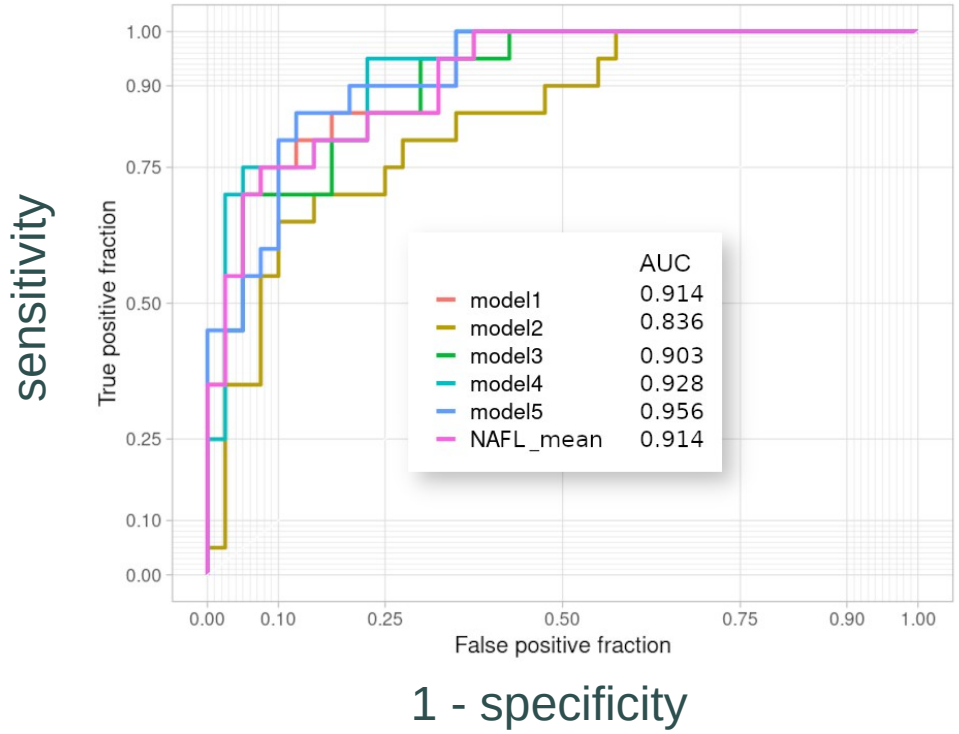
Sensitivity (true positive rate) = TP/(TP+FN)

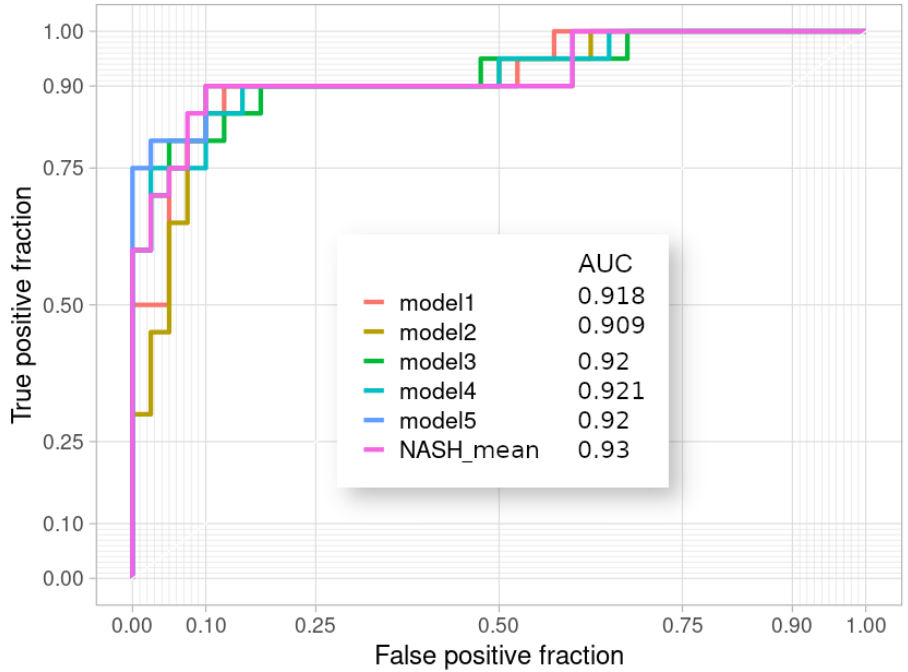
Specificity (true negative rate) = TN/(TN+FP)

Receiver operating characteristic (ROC) curve



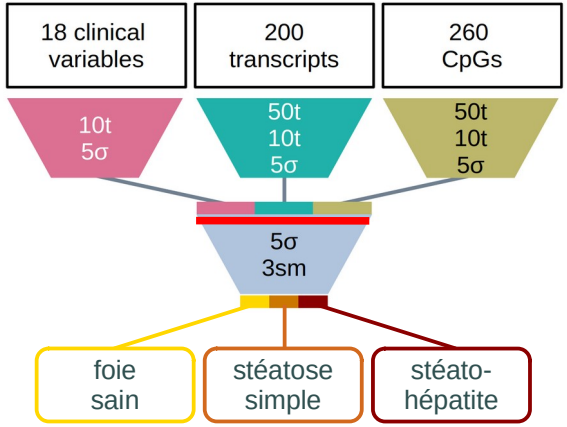
# How good is the model to distinguish NAFL and NASH?



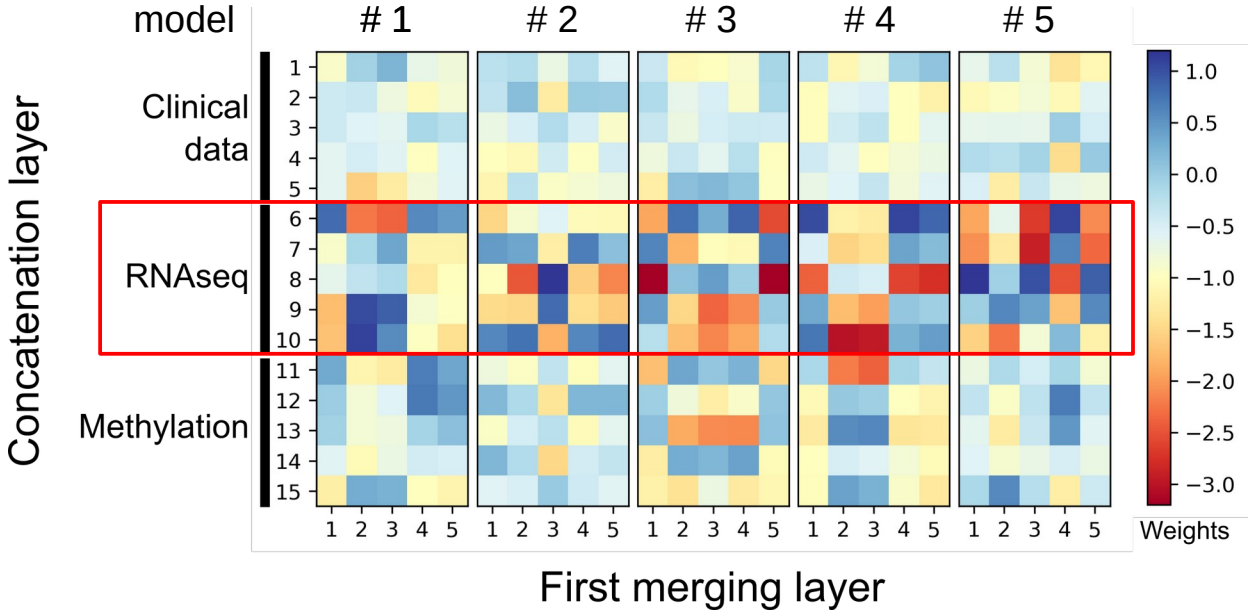




On the validation set, never seen by the model *and the modeller* during training

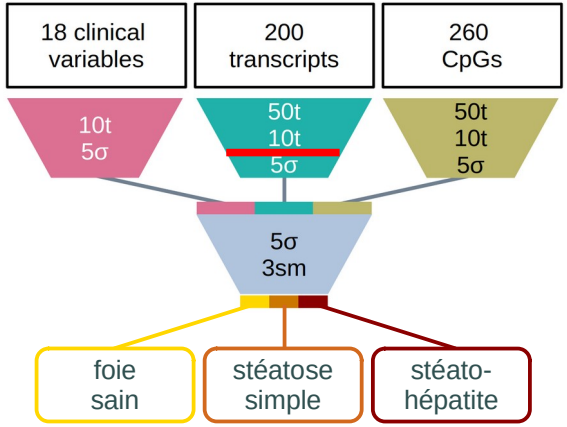
# AI models are not (always) black boxes

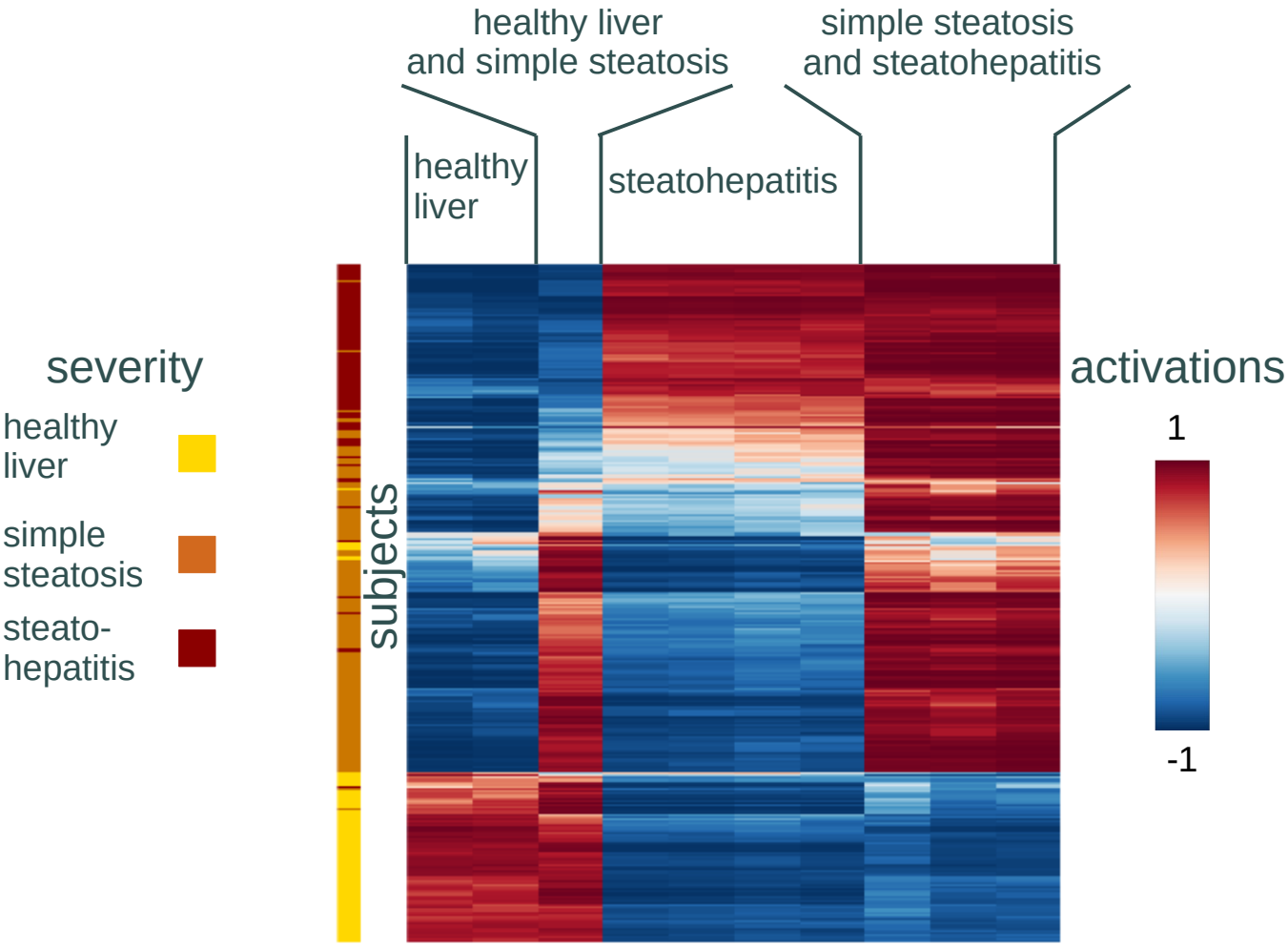



The weights reading the RNAseq module are larger → most impact on output

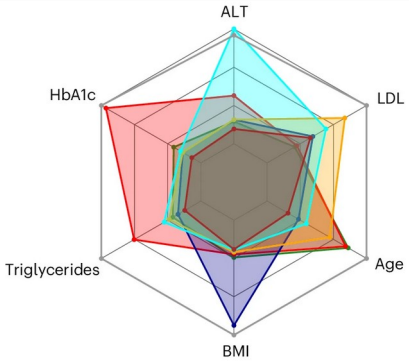


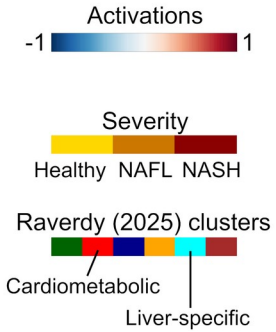
# AI models are not (always) black boxes

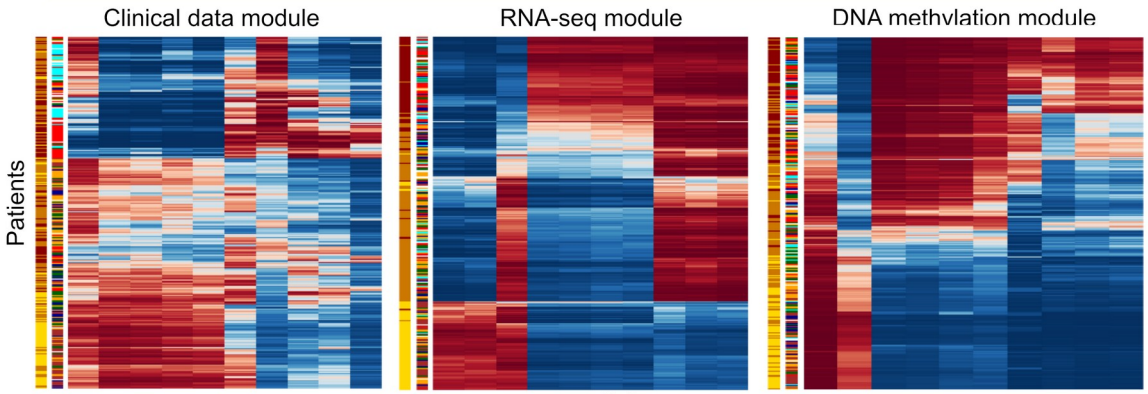


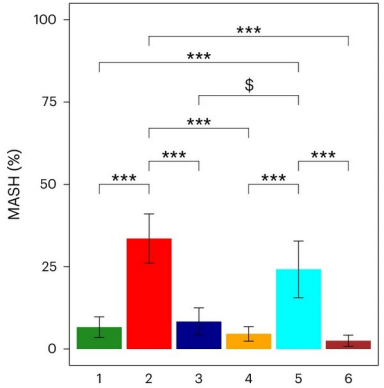



# Exploring latent spaces



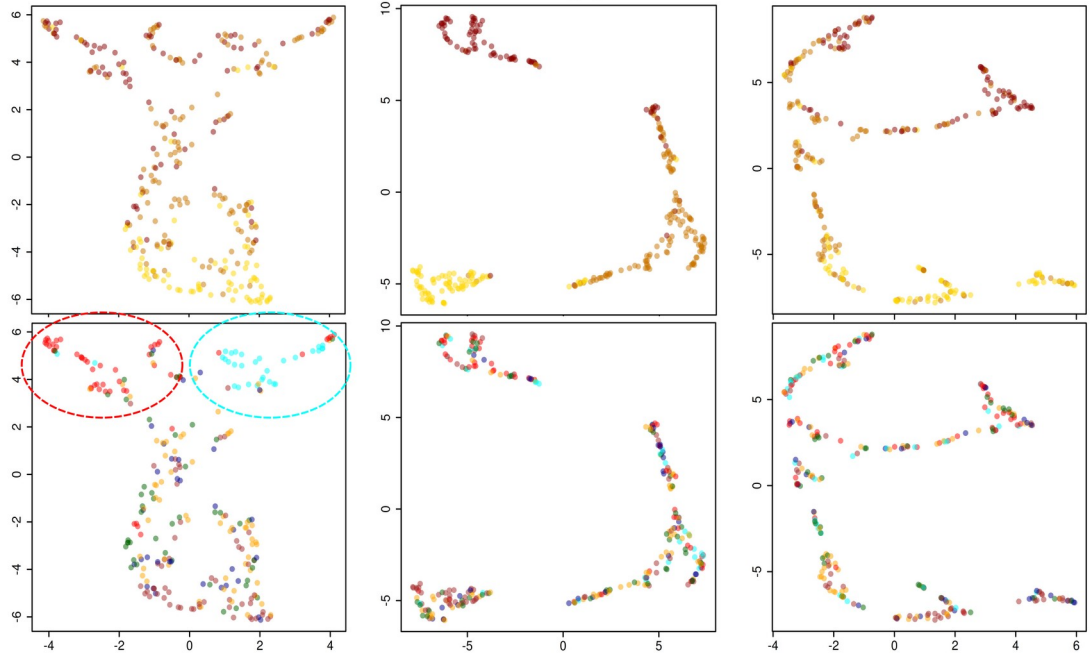




UMAP severity

UMAP Raverdy clusters

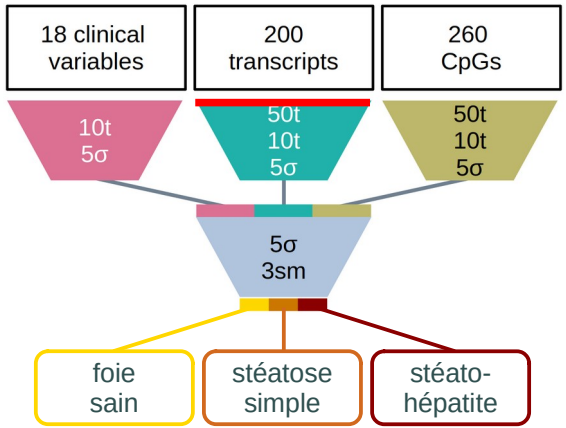


Clinical data: several populations
See Raverdy *et al. Nat Med* 30:3624–3633 (2024)

RNA-seq: recognises the 3 conditions

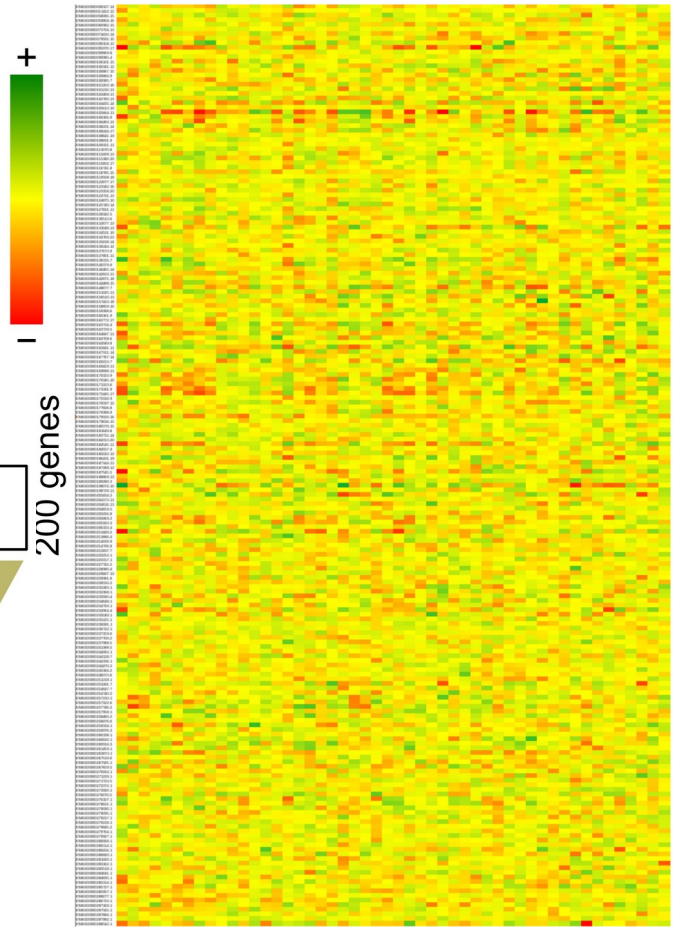
Methylation: continuum of severity

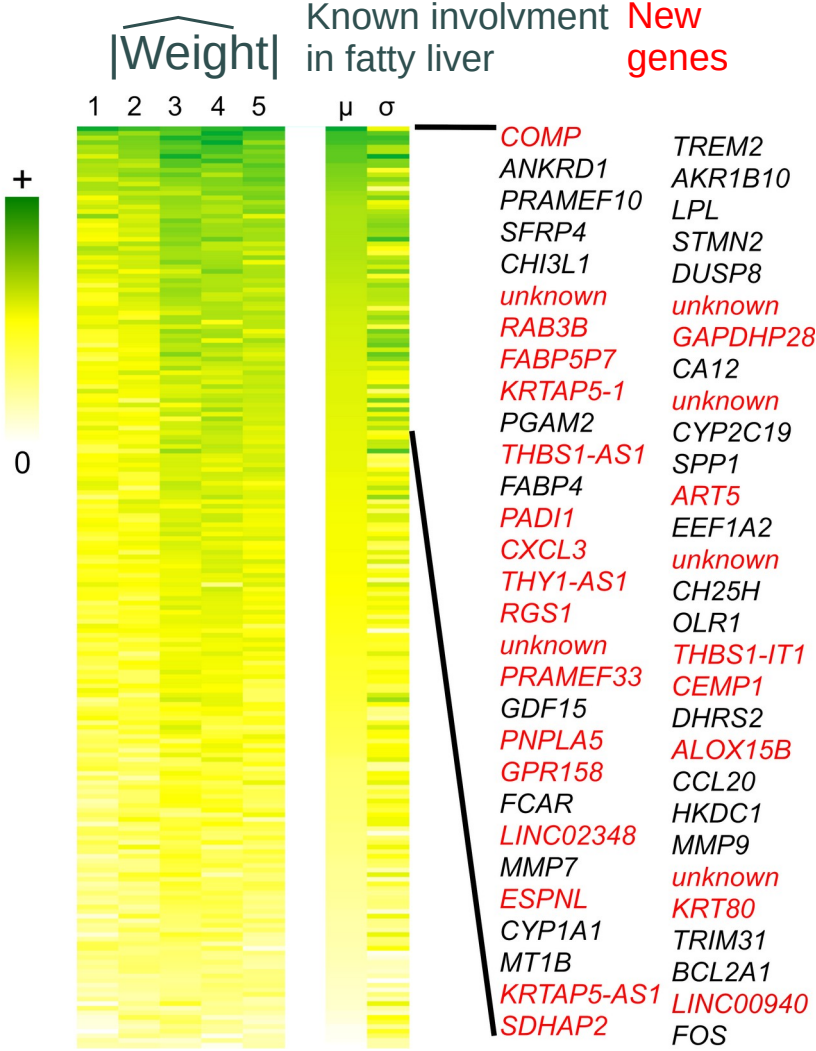
# What are the genes taken into account by the models?




"synaptic weights"

50 neurons



# The Transformer

# The paper that changed everything: the Transfomer

## Attention Is All You Need

**Ashish Vaswani*** Google Brain avaswani@google.com
**Noam Shazeer*** Google Brain noam@google.com
**Niki Parmar*** Google Research nikip@google.com
**Jakob Uszkoreit*** Google Research usz@google.com
**Llion Jones*** Google Research llion@google.com
**Aidan N. Gomez*** † University of Toronto aidan@cs.toronto.edu
**Łukasz Kaiser*** Google Brain lukaszkaiser@google.com
**Illia Polosukhin*** ‡ illia.polosukhin@gmail.com

### Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.

### 1 Introduction

Recurrent neural networks, long short-term memory [12] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and transduction problems such as language modeling and machine translation [29, 2, 5]. Numerous efforts have since continued to push the boundaries of recurrent language models and encoder-decoder architectures [31, 21, 13].

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

‡Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

# The paper that changed everything: the Transfomer

## Attention Is All You Need

Cool title


**Ashish Vaswani*** Google Brain avaswani@google.com

**Noam Shazeer*** Google Brain noam@google.com

**Niki Parmar*** Google Research nikip@google.com

**Jakob Uszkoreit*** Google Research usz@google.com

**Llion Jones*** Google Research llion@google.com

**Aidan N. Gomez*** † University of Toronto aidan@cs.toronto.edu

**Łukasz Kaiser*** Google Brain lukaszkaiser@google.com

**Illia Polosukhin*** ‡ illia.polosukhin@gmail.com



### Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.


All authors equal

### 1 Introduction

Recurrent neural networks, long short-term memory [12] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and

*Equal contribution. Listing order is random.

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

‡Work performed while at Google Research.

Never published in a journal


31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.


Cited... 203326 times as of 12 November 2025!

# The paper that changed everything: the Transfomer

## Attention Is All You Need


**Ashish Vaswani*** Google Brain avaswani@google.com

**Noam Shazeer*** Google Brain noam@google.com

**Niki Parmar*** Google Research nikip@google.com

**Jakob Uszkoreit*** Google Research usz@google.com

**Llion Jones*** Google Research llion@google.com

**Aidan N. Gomez*** † University of Toronto aidan@cs.toronto.edu

**Łukasz Kaiser*** Google Brain lukaszkaiser@google.com

**Illia Polosukhin*** ‡ illia.polosukhin@gmail.com



### Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.


### 1 Introduction

Recurrent neural networks, long short-term memory [12] and gated recurrent [7] neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and transduction problems such as language modeling and machine translation [29, 2, 5]. Numerous efforts have since continued to push the boundaries of recurrent language models and encoder-decoder architectures [31, 21, 13].

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

‡Work performed while at Google Research.


31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.


Claude

AlphaFold

# The Transformer: Memory + context = attention

Encoder

Decoder

Residual connections and layer normalization

Feed-forward network: after taking information from other tokens, take a moment to think and process this information

Feed-forward network: after taking information from other tokens, take a moment to think and process this information

Decoder-encoder attention: target token looks at the source

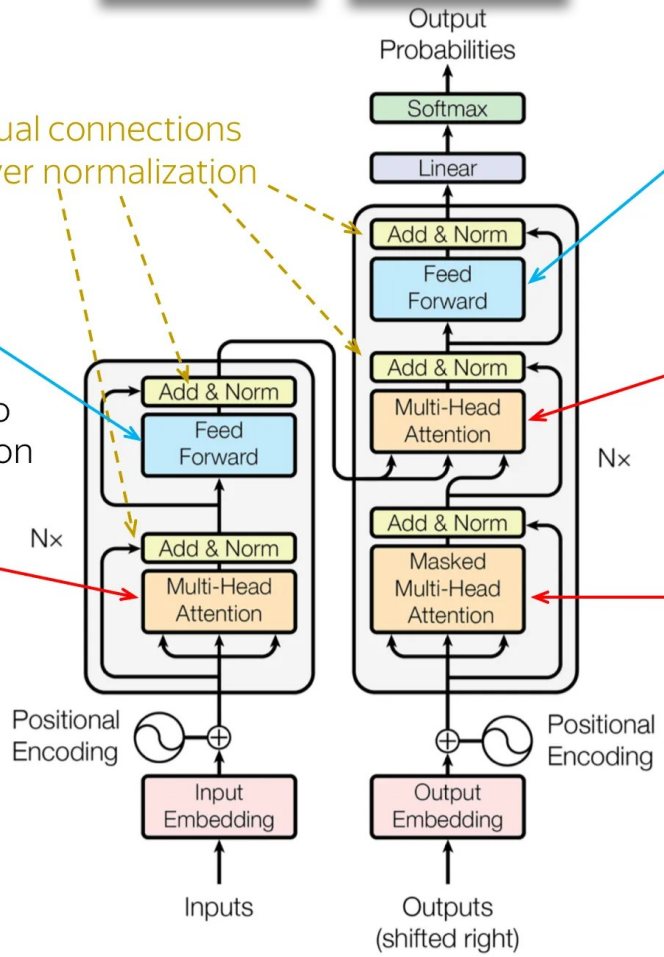
queries – from decoder states; keys and values from encoder states

Encoder self-attention: tokens look at each other

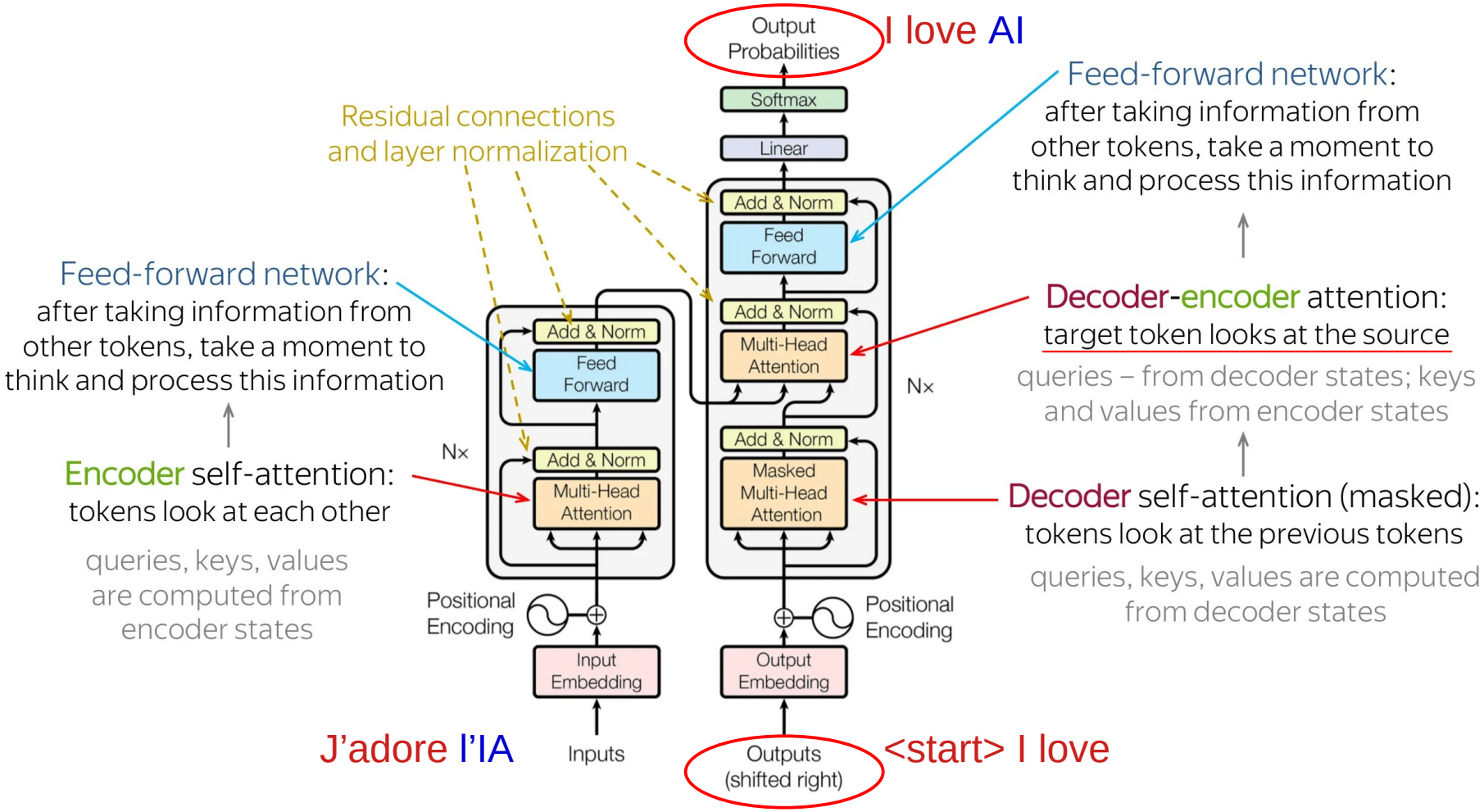
queries, keys, values are computed from encoder states

Decoder self-attention (masked): tokens look at the previous tokens

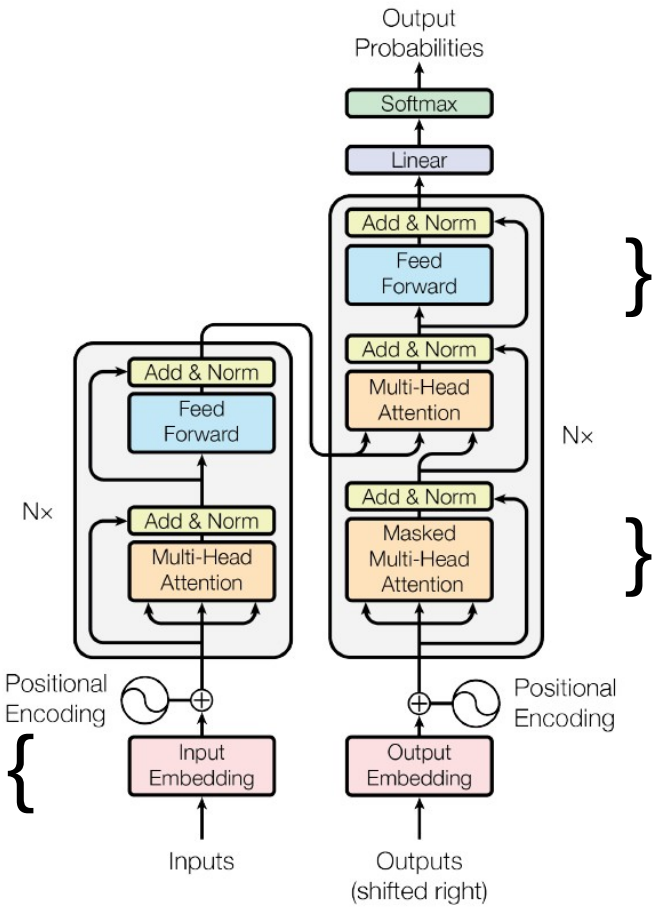
queries, keys, values are computed from decoder states



# The Transformer: Memory + context = attention



# Embedding, context and decision?




Takes a decision based on the attentions attached to each token
→ Foundational models;
To reuse a model, just replace the prediction head and retrain the feed forward blocks

Learned context. Provide attentions attached to each token in a sequence

Transformation of each element of a sequence (a token) into a vector representing its position in the structured space of possible. Think non-linear PCA

# Embeddings (“plongement”)

Values in reference frame A
**m** vectors from a dictionary of **n**
(i.e. **n** coordinates)

Embedding from a space with **n** dimensions
into a space of **o** dimensions

Values in reference frame B
**m’** vectors from a dictionary of **o**
(i.e. **o** coordinates)

# Embeddings (“plongement”)

Magic (PCA, MLP, CNN, ...)

## Dictionary (initial space)

| bunny | father | hamster | hutches | man | mother | rabbit | tractor | woman |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

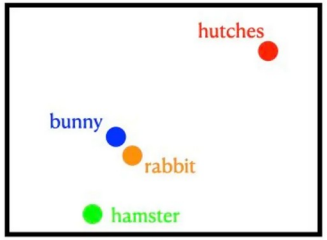
## Semantics (Embedding space)

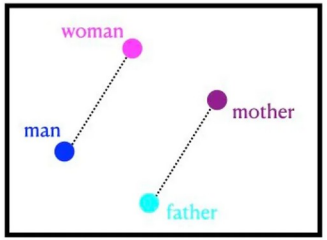
| Word | alive | mammal | rodent | bipedel | gender | plural |
|---|---|---|---|---|---|---|
| bunny | 0.9 | 0.7 | -0.1 | -0.1 | 0.2 | -0.3 |
| rabbit | 0.9 | 0.8 | -0.1 | -0.1 | 0.4 | -0.1 |
| hamster | 0.9 | 0.6 | 0.7 | -0.3 | -0.4 | -0.4 |
| hutches | -0.9 | -0.3 | -0.1 | -0.9 | 0.0 | 0.9 |
| mother | 0.9 | 0.1 | 0.1 | 0.2 | 0.8 | -0.7 |
| father | 0.9 | 0.2 | 0.1 | 0.2 | -0.8 | -0.7 |
| woman | 0.9 | 0.8 | -0.9 | 0.9 | 0.8 | -0.8 |
| man | 0.9 | 0.8 | -0.7 | 0.9 | -0.8 | -0.8 |

Vector Embedding

Dimensionality Reduction

Dimensionality




2D Visualization

Mikolov *et al* (2015) Linguistic Regularities in Continuous Space Word Representations In *Proceedings of NAACL-HLT,* pp 746–751, Atlanta, Georgia.

# Embeddings (“plongement”)

Magic (PCA, MLP, CNN, ...)

## Dictionary (initial space)

| bunny | father | hamster | hutches | man | mother | rabbit | tractor | woman |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

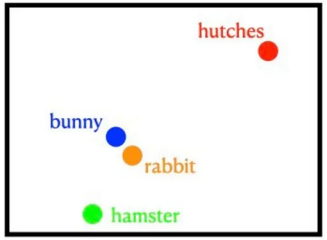
## Semantics (Embedding space)

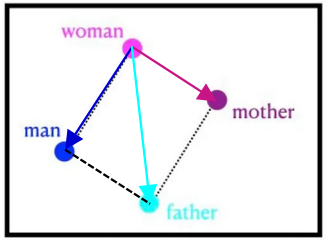
| Word | alive | mammal | rodent | bipedel | gender | plural |
|---|---|---|---|---|---|---|
| bunny | 0.9 | 0.7 | -0.1 | -0.1 | 0.2 | -0.3 |
| rabbit | 0.9 | 0.8 | -0.1 | -0.1 | 0.4 | -0.1 |
| hamster | 0.9 | 0.6 | 0.7 | -0.3 | -0.4 | -0.4 |
| hutches | -0.9 | -0.3 | -0.1 | -0.9 | 0.0 | 0.9 |
| mother | 0.9 | 0.1 | 0.1 | 0.2 | 0.8 | -0.7 |
| father | 0.9 | 0.2 | 0.1 | 0.2 | -0.8 | -0.7 |
| woman | 0.9 | 0.8 | -0.9 | 0.9 | 0.8 | -0.8 |
| man | 0.9 | 0.8 | -0.7 | 0.9 | -0.8 | -0.8 |

Vector Embedding

Dimensionality Reduction

Dimensionality

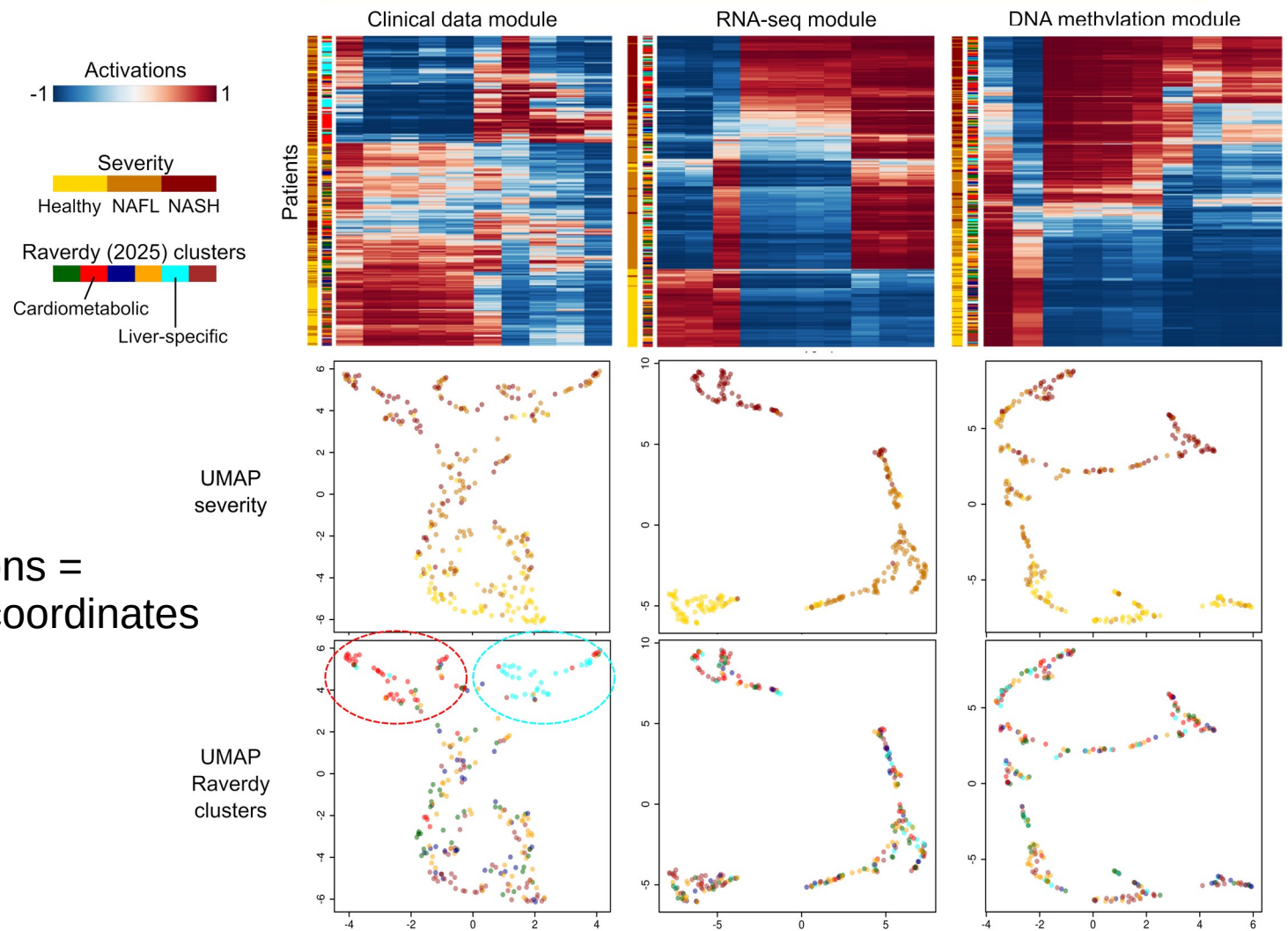



2D Visualization

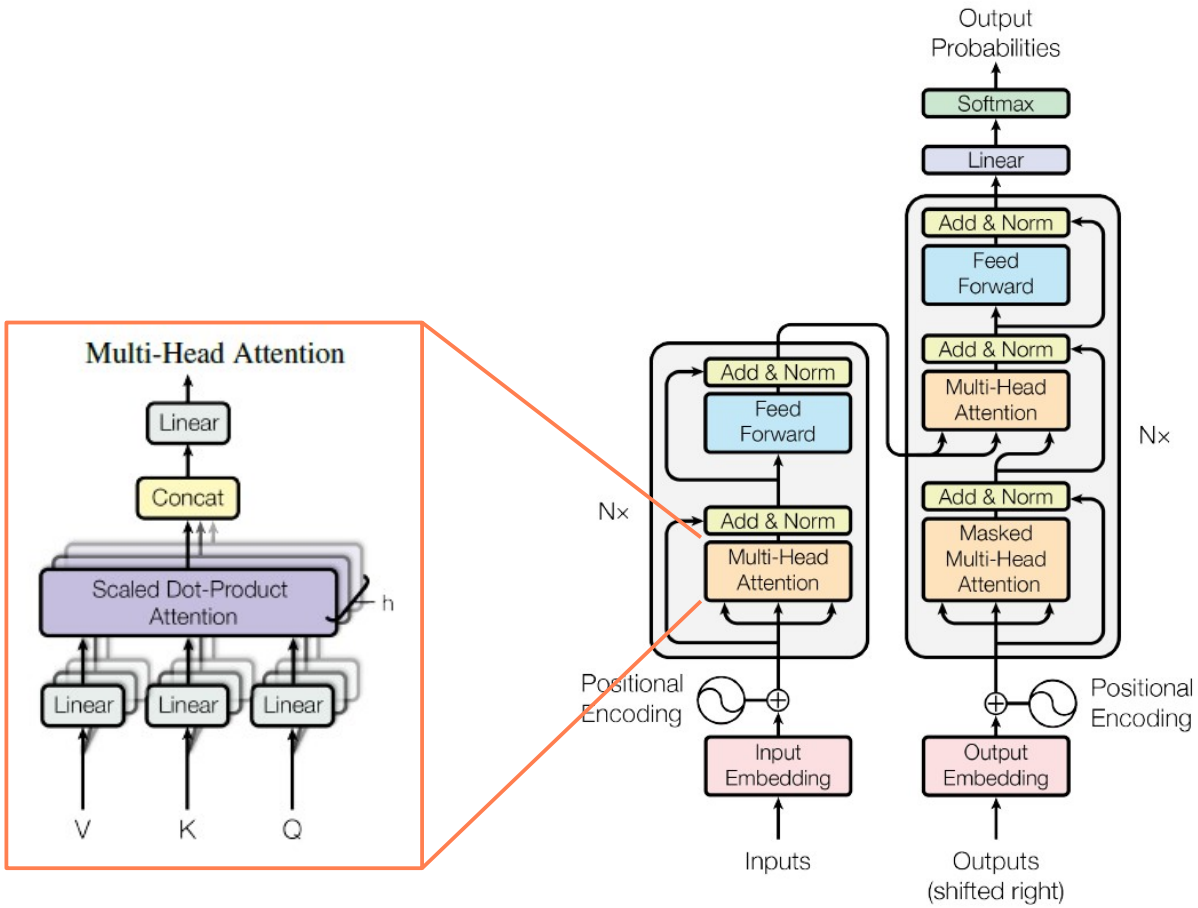
In the embedding space, the vector going from woman to father is equal to the vector going from woman to man plus the vector going from woman to mother, i.e. (replacing by the coordinates), father = mother – woman + man

# MLPs create embeddings (latent space coordinates)



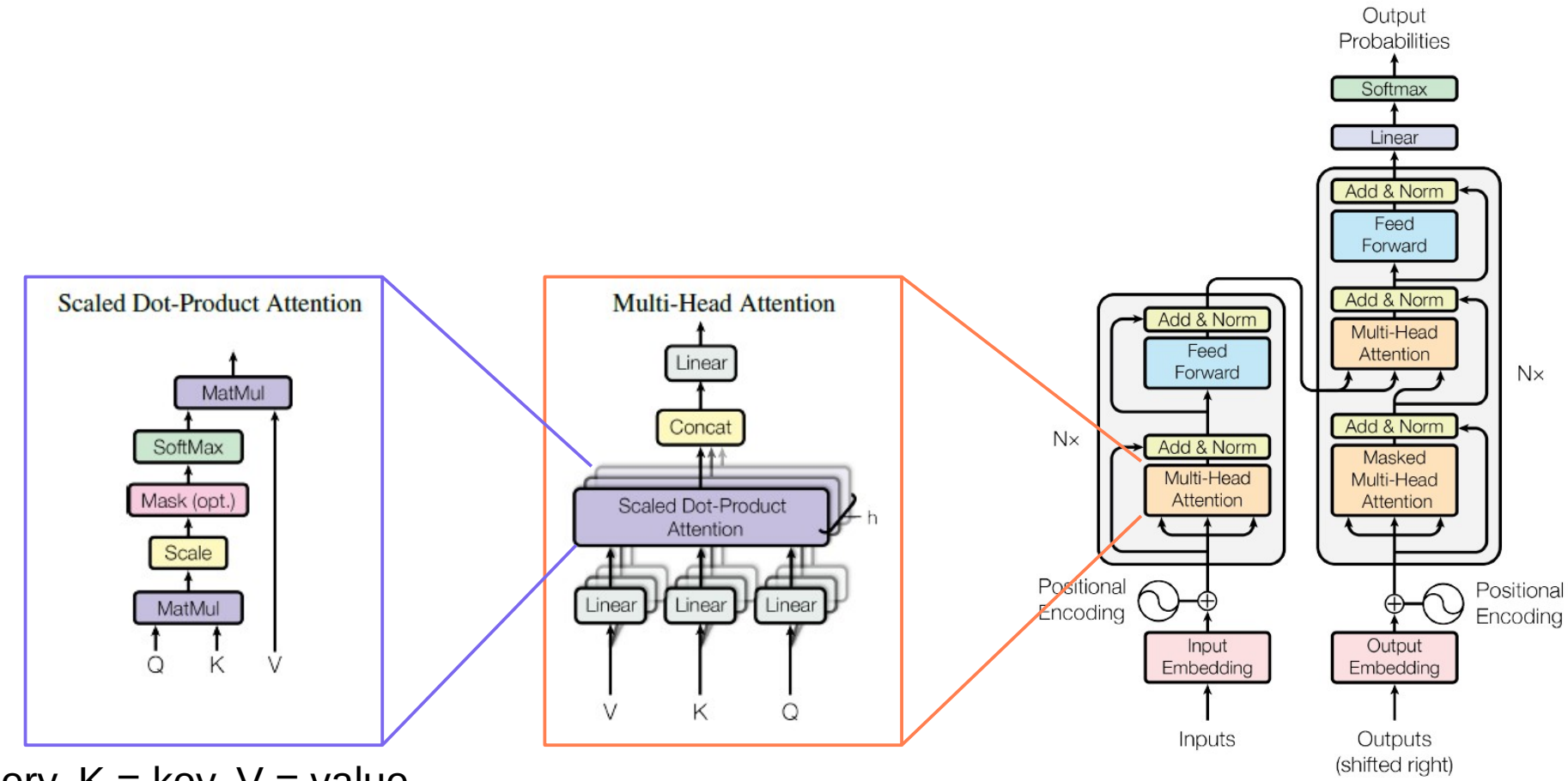

activations =
latent space coordinates

# Attention in the Transformer




Q = query, K = key, V = value

# Attention in the Transformer




Q = query, K = key, V = value

# A drama analogy?

3 The spotlight represents the intensity of the responses from Lady MacBeth to MacBeth's current line

1

The model learned the impact of MacBeth's speeches on others

His current line becomes **Q** and puts the spotlight on Lady MacBeth

2

The model learned Lady MacBeth' reactions to MacBeth's speeches His current line becomes Lady MacBeth's reactions **K**

4

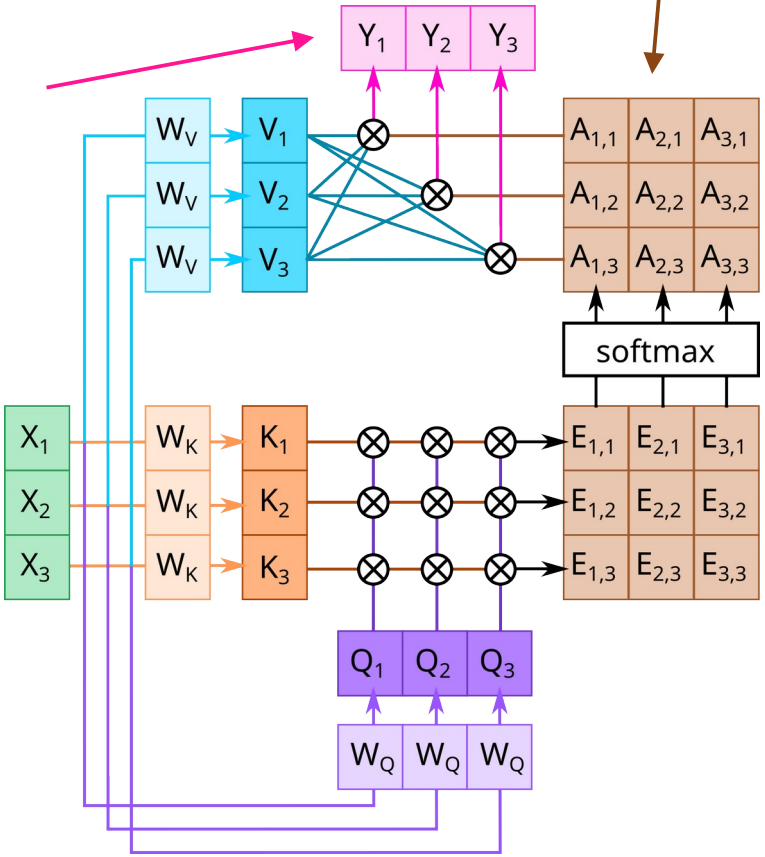
Lady MacBeth's *illuminated* emotional presence and cue send back reactions **V** to MacBeth

The public did not attract focus from MacBeth and do not affect the rest of his speech

# Attention in the Transformer

Attentions of all tokens on all tokens

Values associated with relevant attentions



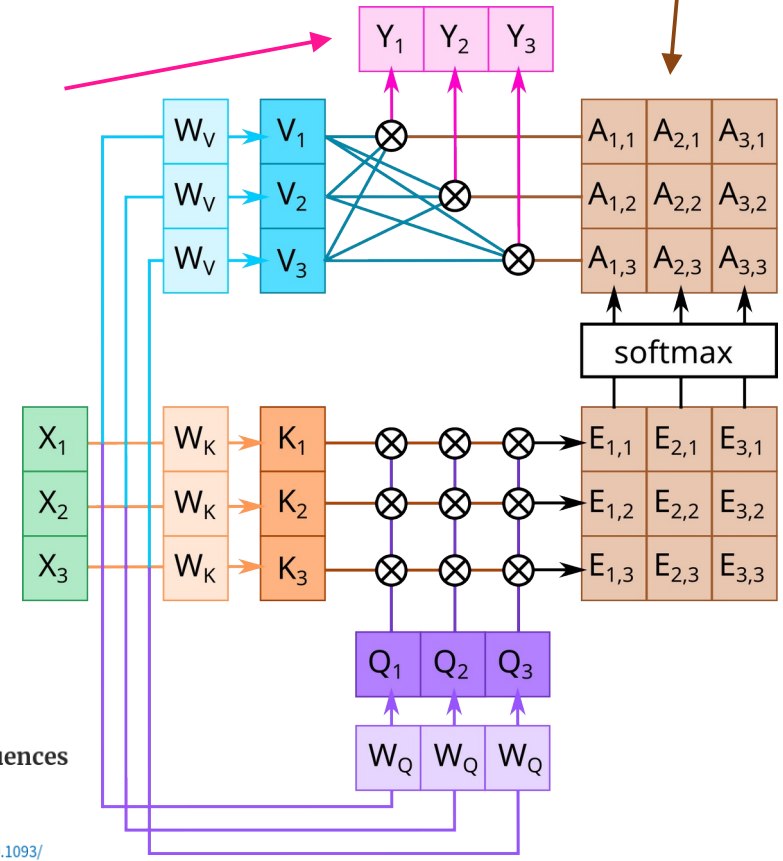

X is entered
$W_Q$, $W_K$, and $W_V$ are learned
Everything else is computed

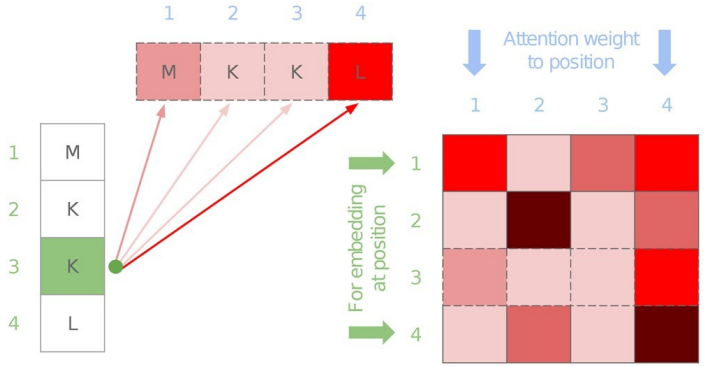
The dot product between Q and $K^T$ compute how aligned are the vectors encoding two tokens (~cosine similarity)

# Attention in the Transformer


Attentions of all tokens on all tokens

Values associated with relevant attentions




X is entered
$W_Q$, $W_K$, and $W_V$ are learned
Everything else is computed

The dot product between Q and $K^T$ compute how aligned are the vectors encoding two tokens (~cosine similarity)

**Predicting enzymatic function of protein sequences with attention**

Nicolas Buton, François Coste, Yann Le Cunff

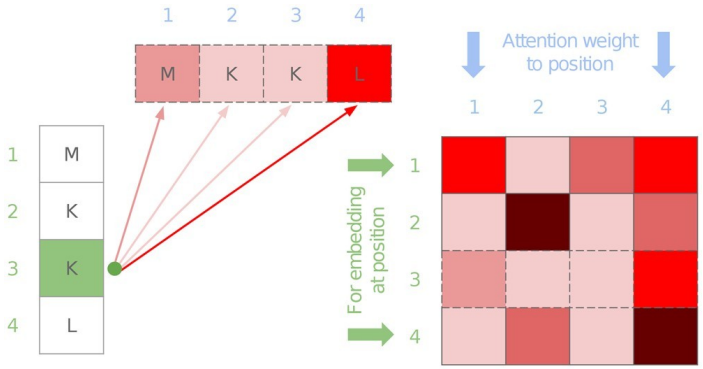
*Bioinformatics*, Volume 39, Issue 10, October 2023, btad620, https://doi.org/10.1093/bioinformatics/btad620

# EnzBERT: amino-acids as tokens

**Predicting enzymatic function of protein sequences with attention**

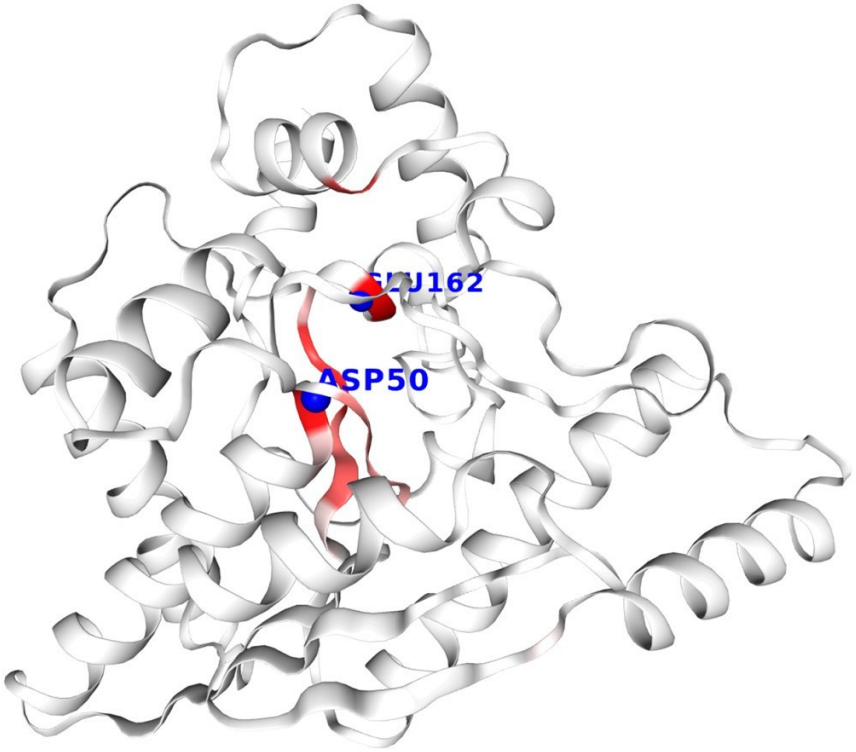
Nicolas Buton, François Coste, Yann Le Cunff

*Bioinformatics*, Volume 39, Issue 10, October 2023, btad620, https://doi.org/10.1093/bioinformatics/btad620

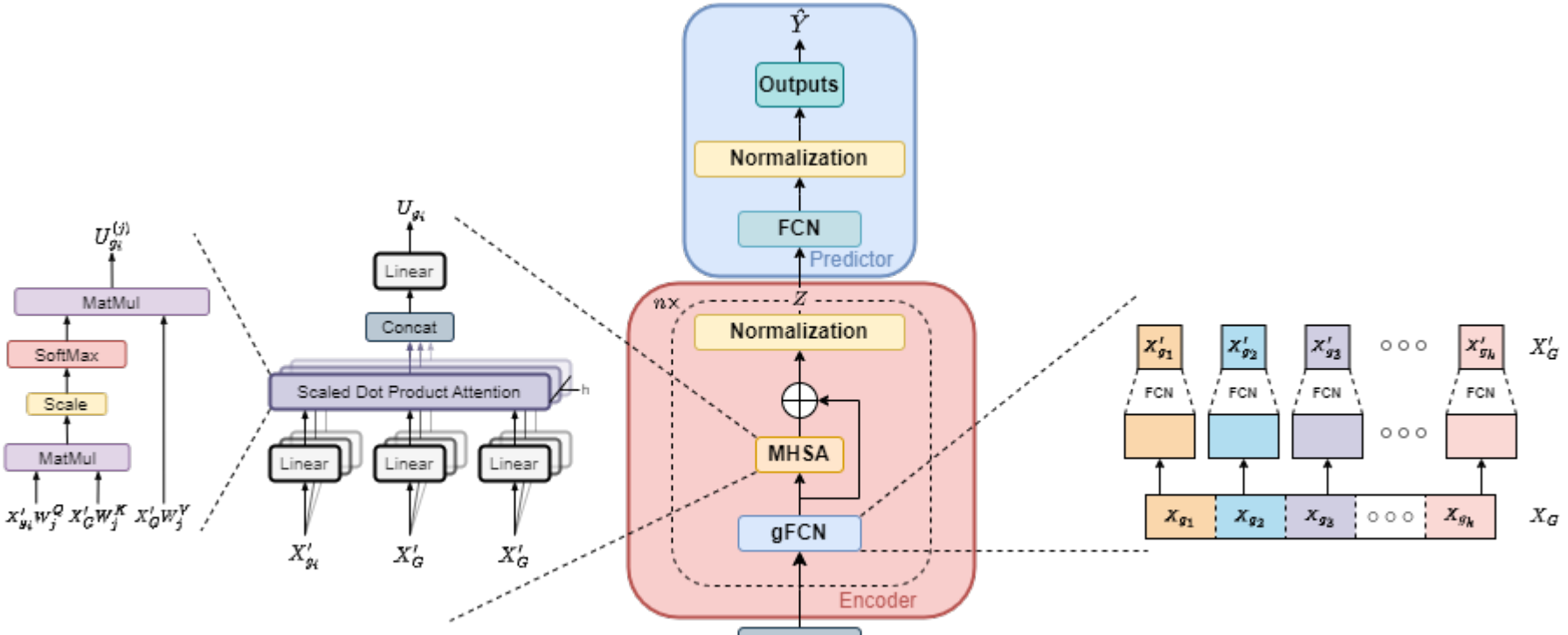



Aggregated attention
for each token (amino acid)

**Nh(3)-dependent nad(+) synthetase**




# AttOmics: Omics values as tokens

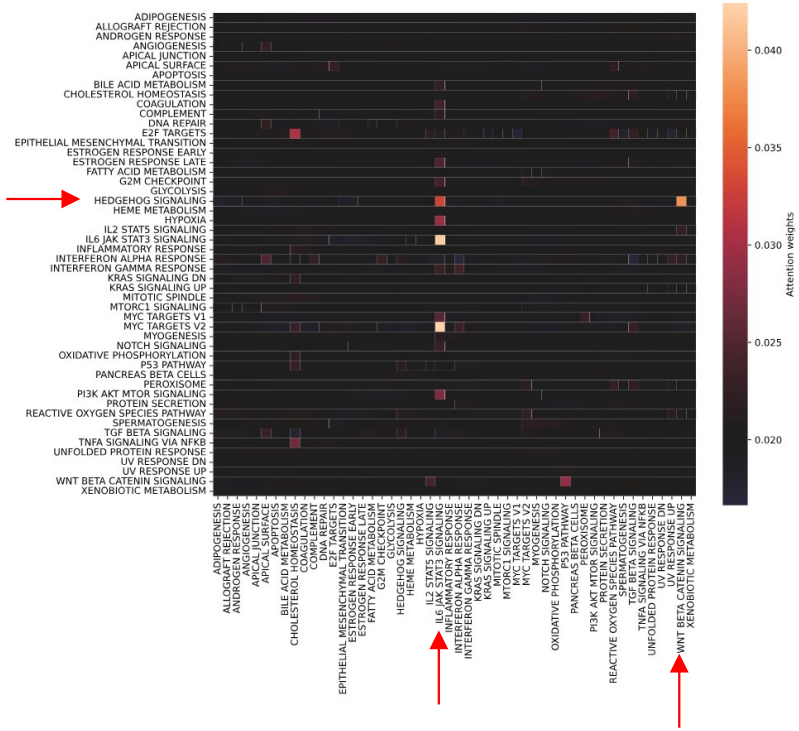



Random, GO BP, MSigDB hallmarks, Clustering

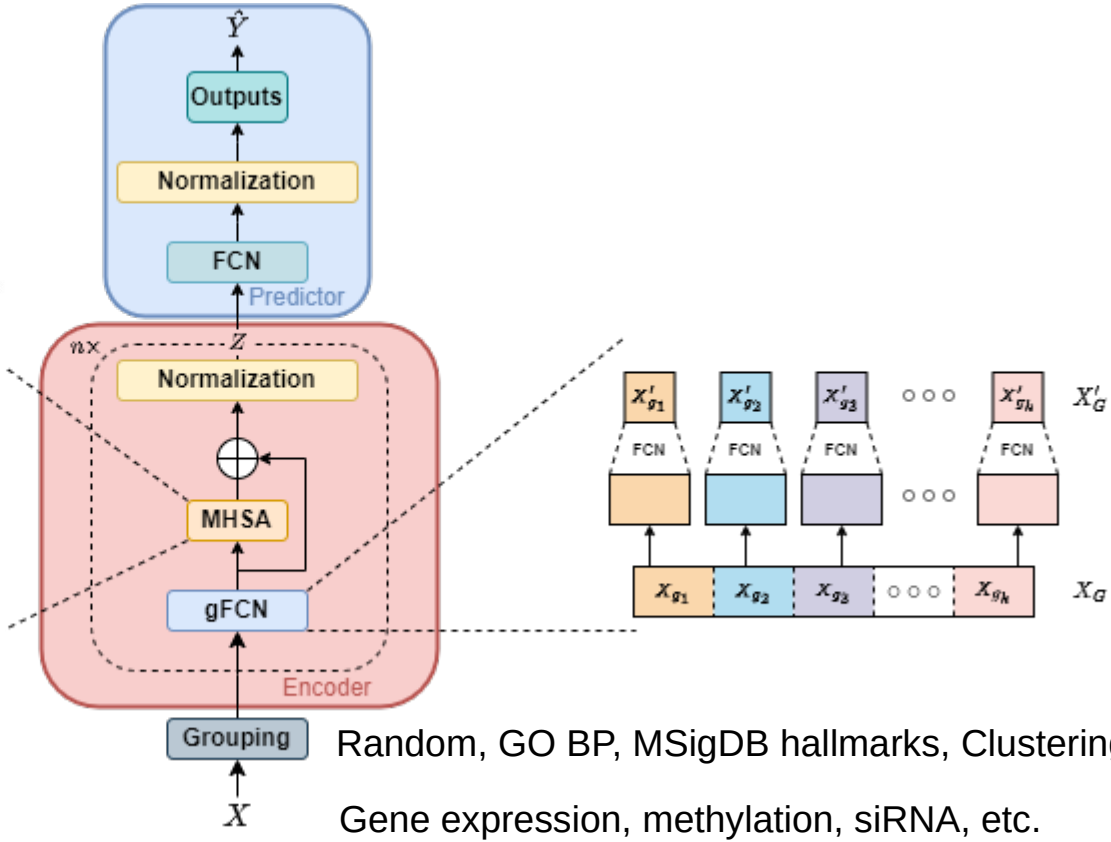
Gene expression, methylation, siRNA, etc.
e.g., TCGA cervical cancers


Beaude, A., Rafiee Vahid, M., Augé, F., Zehraoui, F., & Hanczar, B. (2023).
AttOmics: attention-based architecture for diagnosis and prognosis from omics data. *Bioinformatics*, 39(Supplement_1), i94-i102.

# AttOmics: Omics values as tokens




Pathways affected in cervical cancer

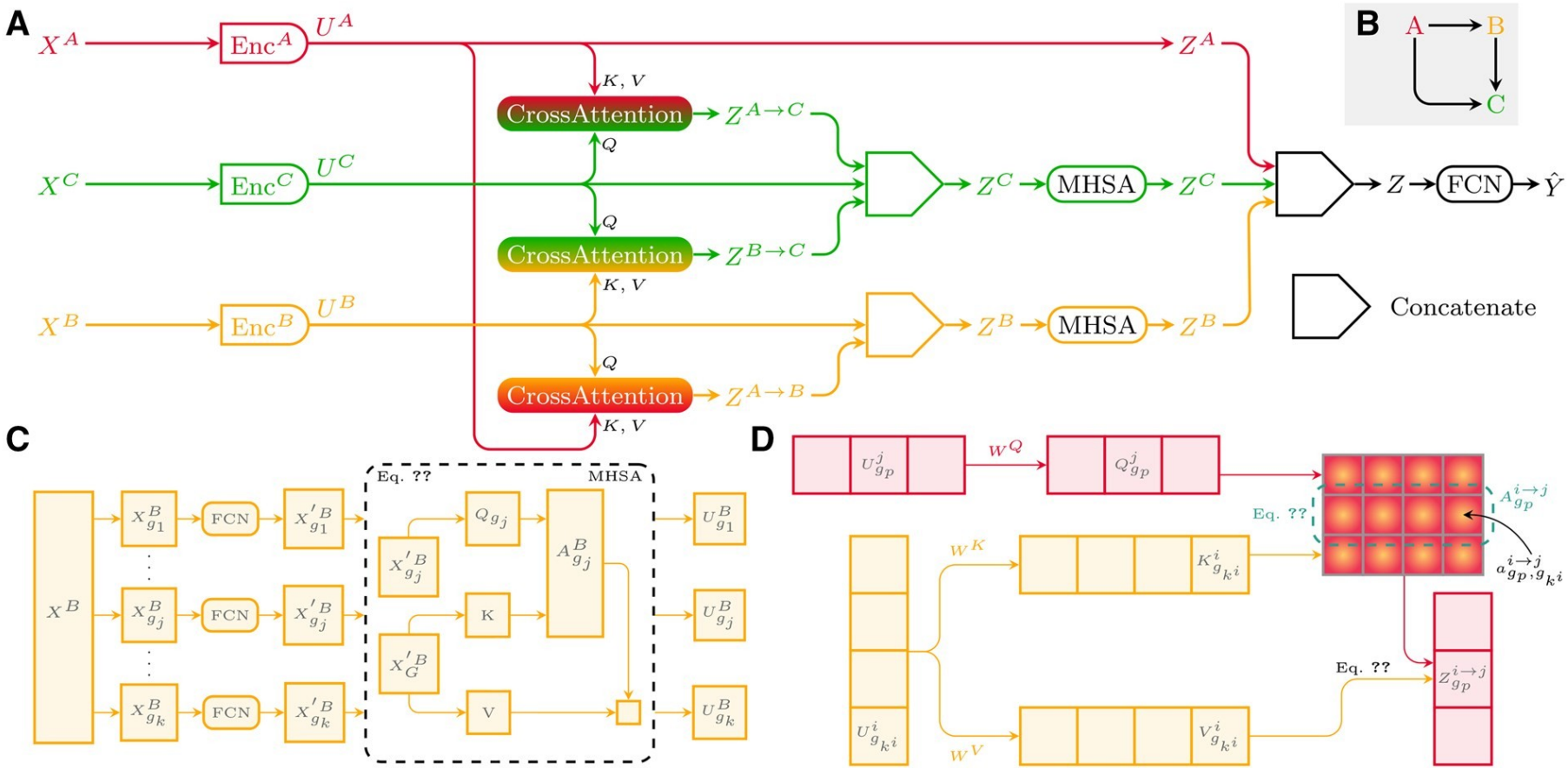



Random, GO BP, MSigDB hallmarks, Clustering

Gene expression, methylation, siRNA, etc.
e.g., TCGA cervical cancers

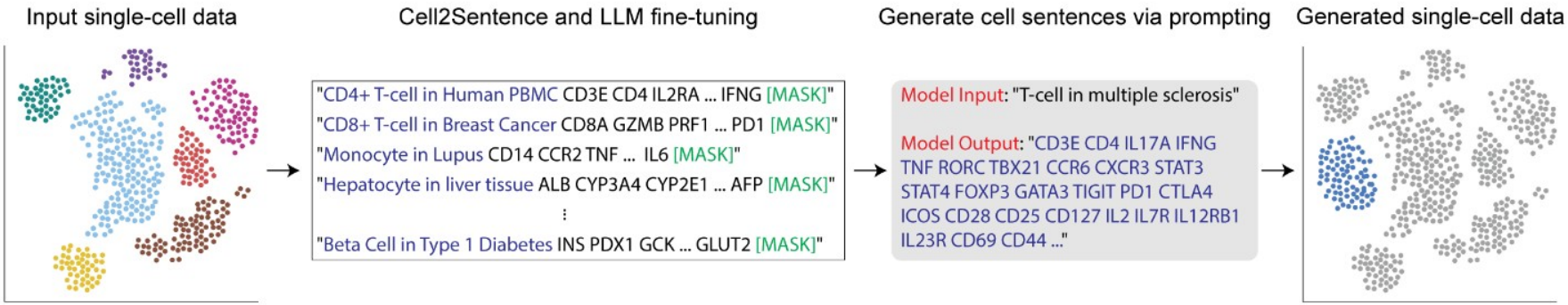

Beaude, A., Rafiee Vahid, M., Augé, F., Zehraoui, F., & Hanczar, B. (2023).
AttOmics: attention-based architecture for diagnosis and prognosis from omics data. *Bioinformatics*, 39(Supplement_1), i94-i102.

# CrossAttomics: Feeding the model all omics at once



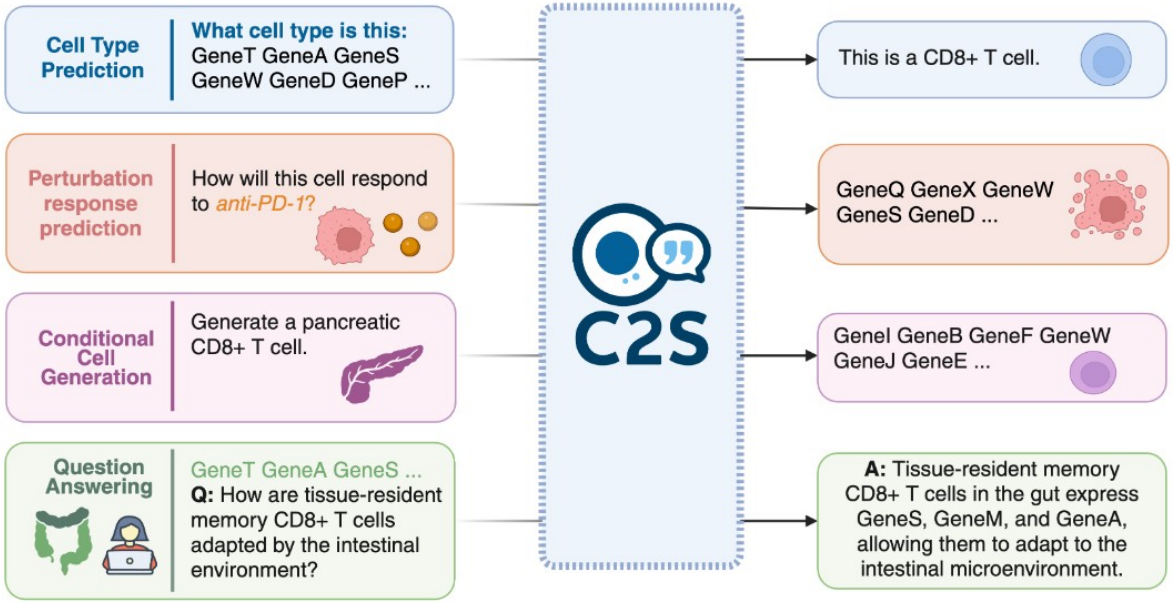

Beaude, A., Augé, F., Zehraoui, F., & Hanczar, B. (2025).
CrossAttOmics: multiomics data integration with cross-attention. *Bioinformatics*, 41 (6): btaf30239.

# Cell2Sentence: gene names as token



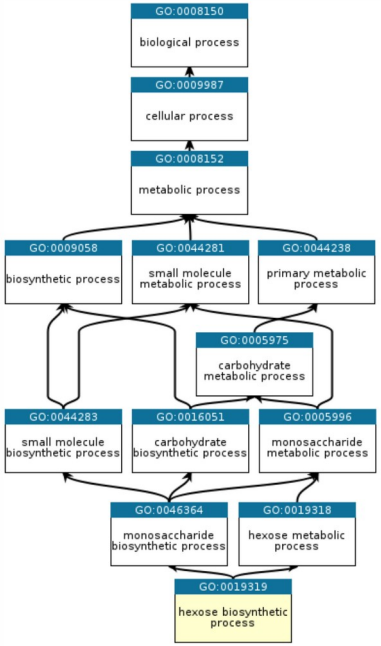

Levine *et al* (2024). Cell2Sentence: Teaching Large Language Models the Language of Biology. *BioRxiv* https://doi.org/10.1101/2023.09.11.557287

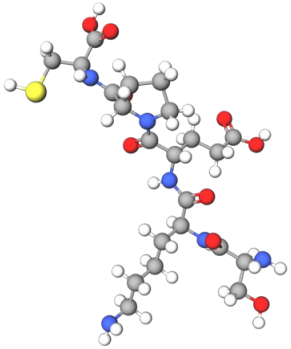
Rizvi *et al* (2025). Scaling large language models for next-generation single-cell analysis. *BioRxiv* https://doi.org/10.1101/2025.04.14.648850

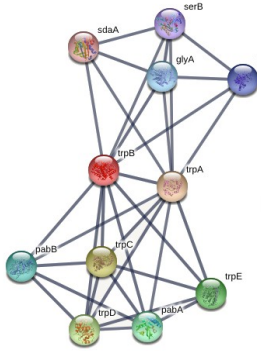


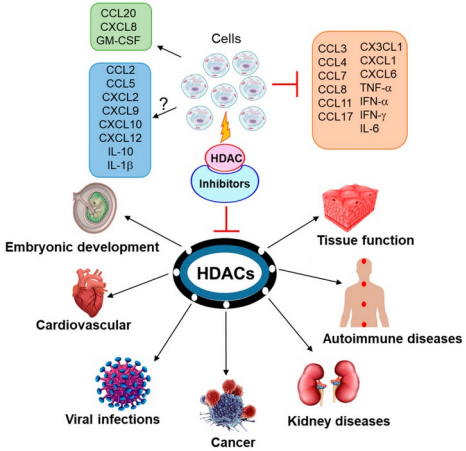
# Graph neural networks

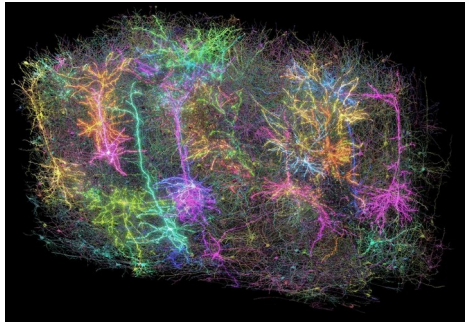
# Most biological knowledge comes as graphs

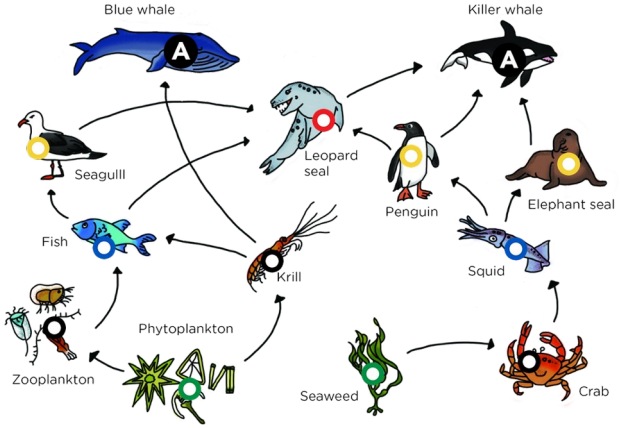


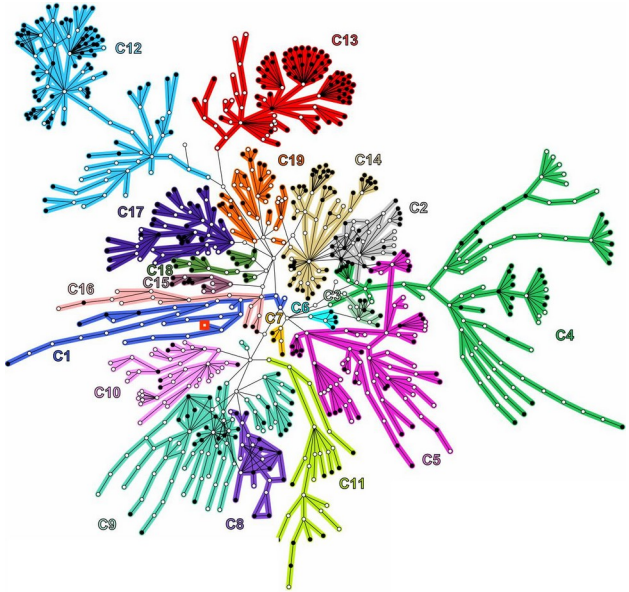



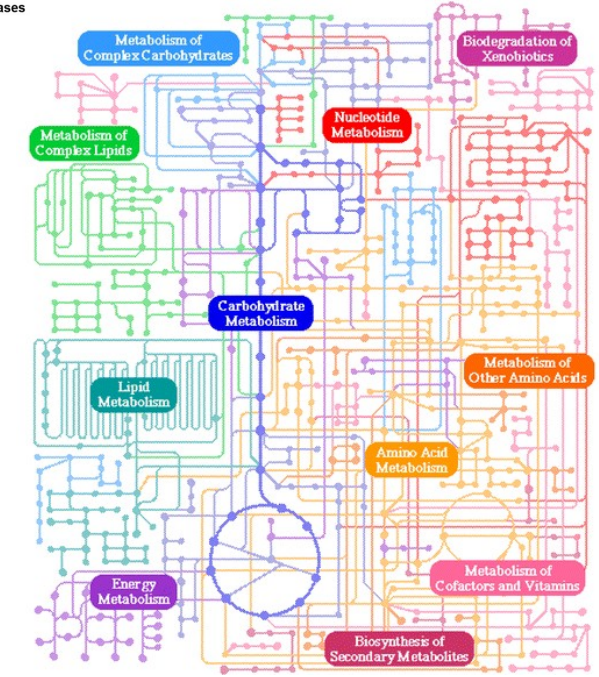


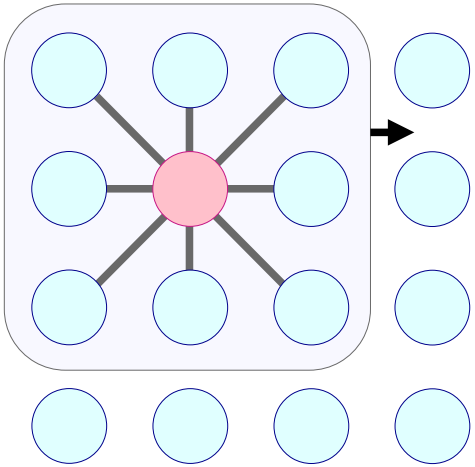


# Graph Neural Networks (GNNs)

## Convolutional Neural Networks
(image recognition)

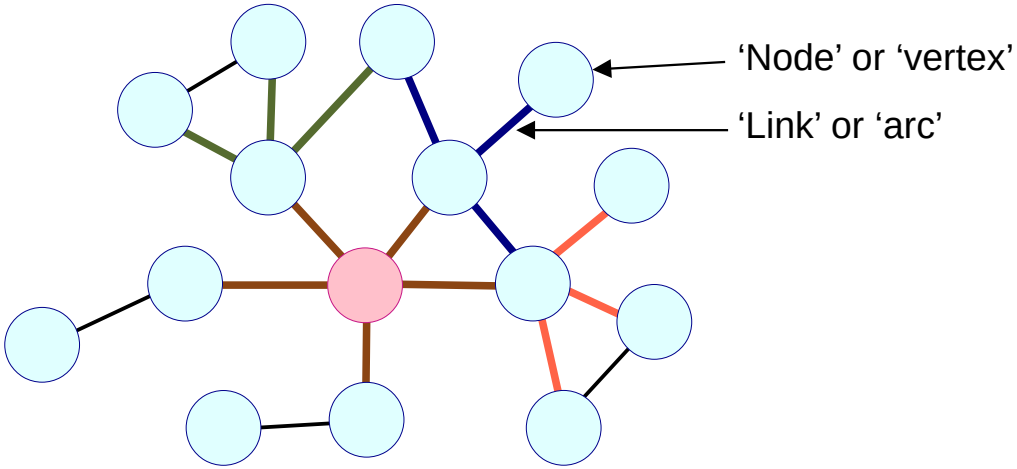


Regular grid (same number of neighbours)
Homogeneous kernels

IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 20, NO. 1, JANUARY 2009 61

## The Graph Neural Network Model

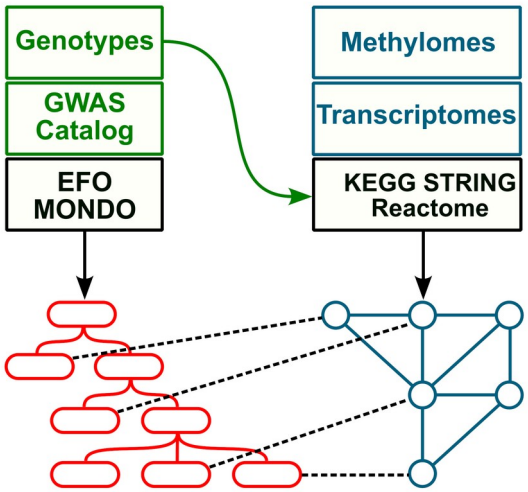
Franco Scarselli, Marco Gori, *Fellow, IEEE*, Ah Chung Tsoi, Markus Hagenbuchner, *Member, IEEE*, and Gabriele Monfardini



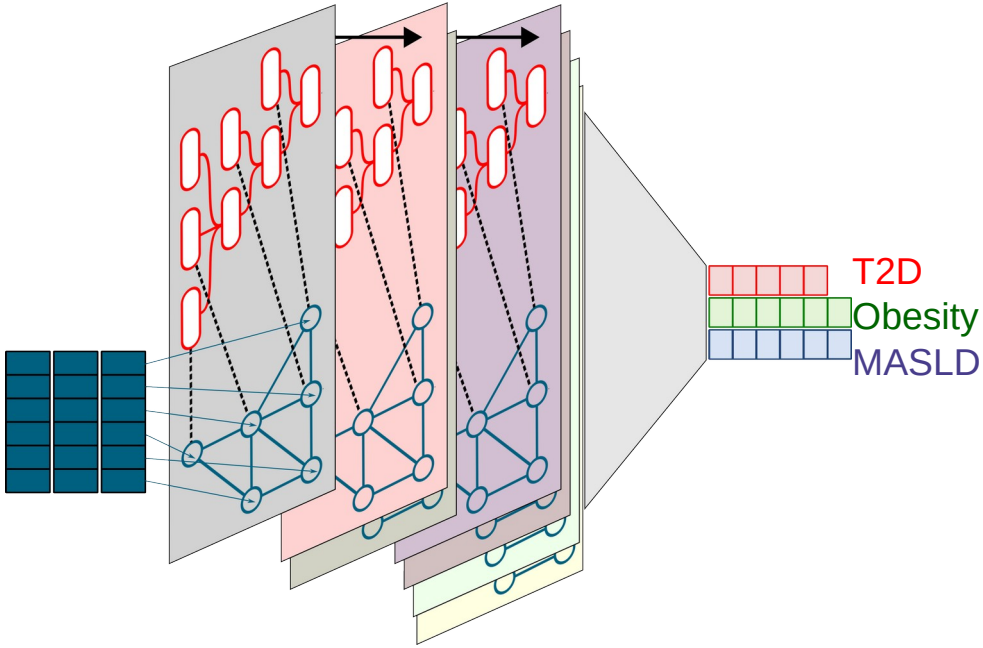

Any number of neighbours
Information passed from neighbours depends on contexts and positions.

# GNN can be heterogeneous

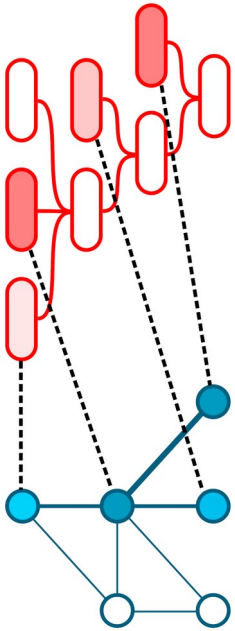
Building



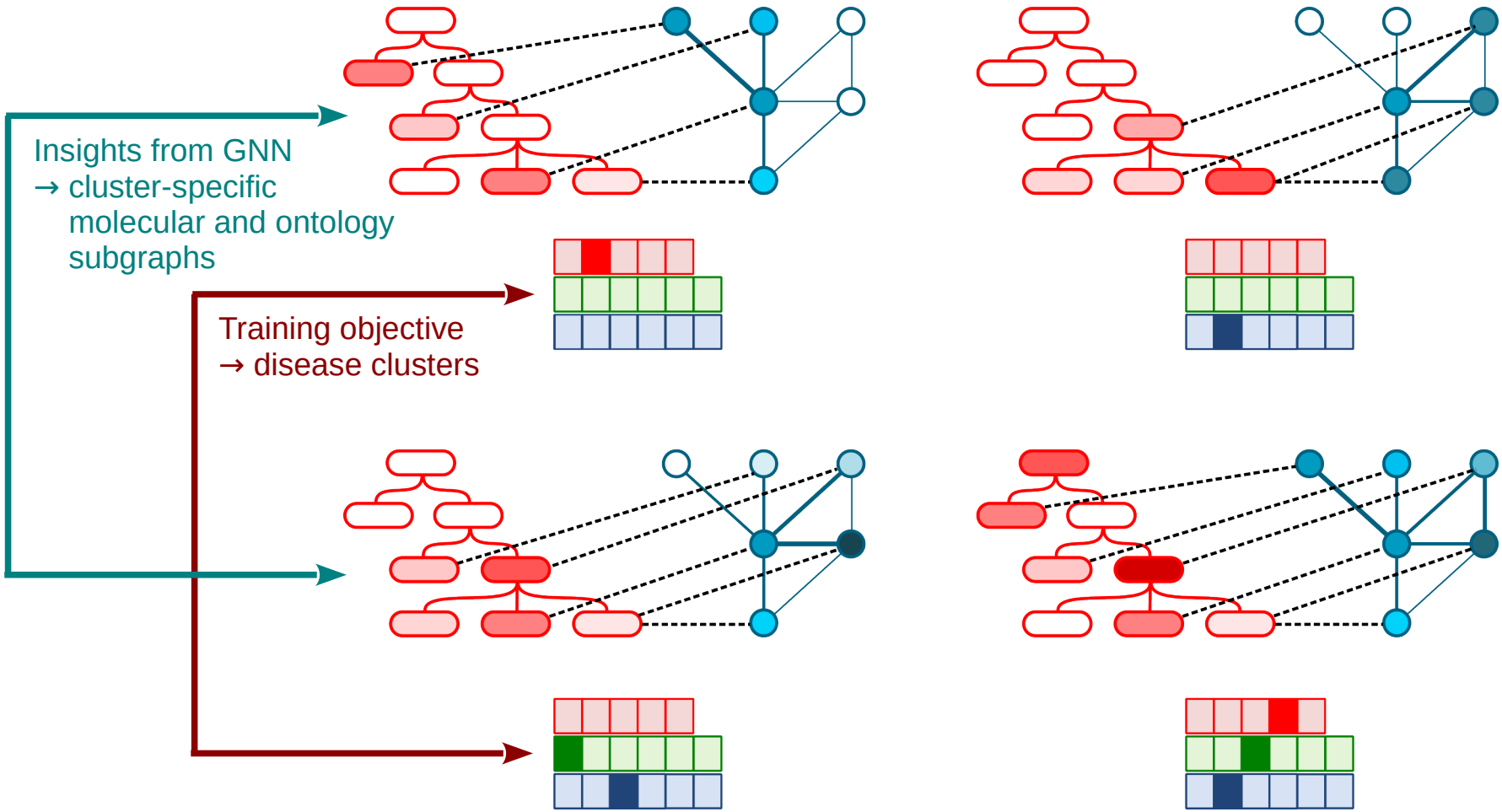

Training




Explanation

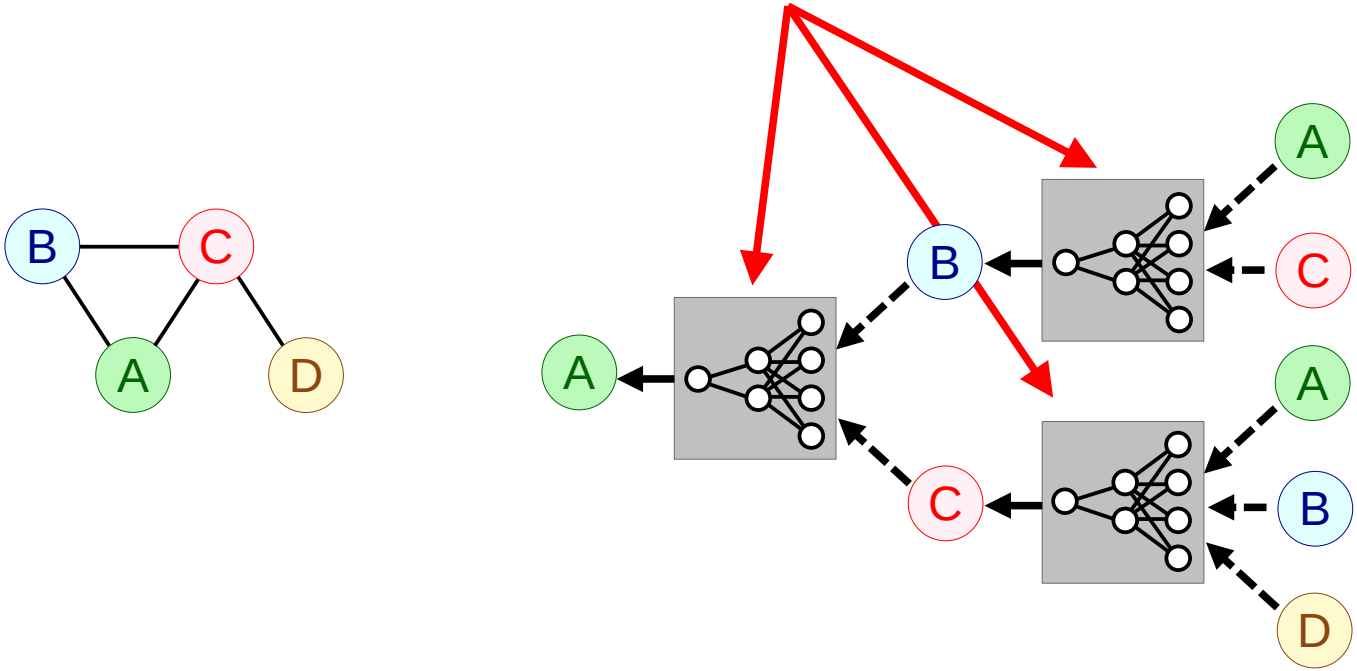


# GNN insights can be subgraphs



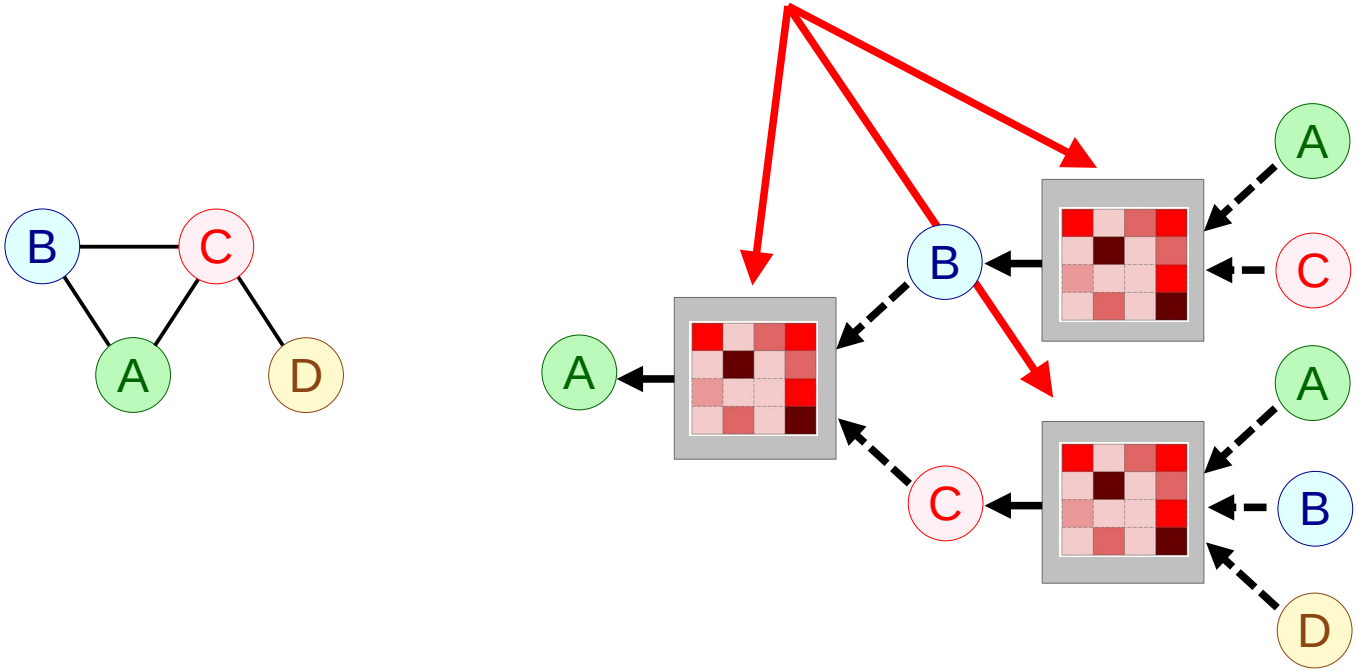
# Many different ways to update GNNs

Can be message passing (MLP)



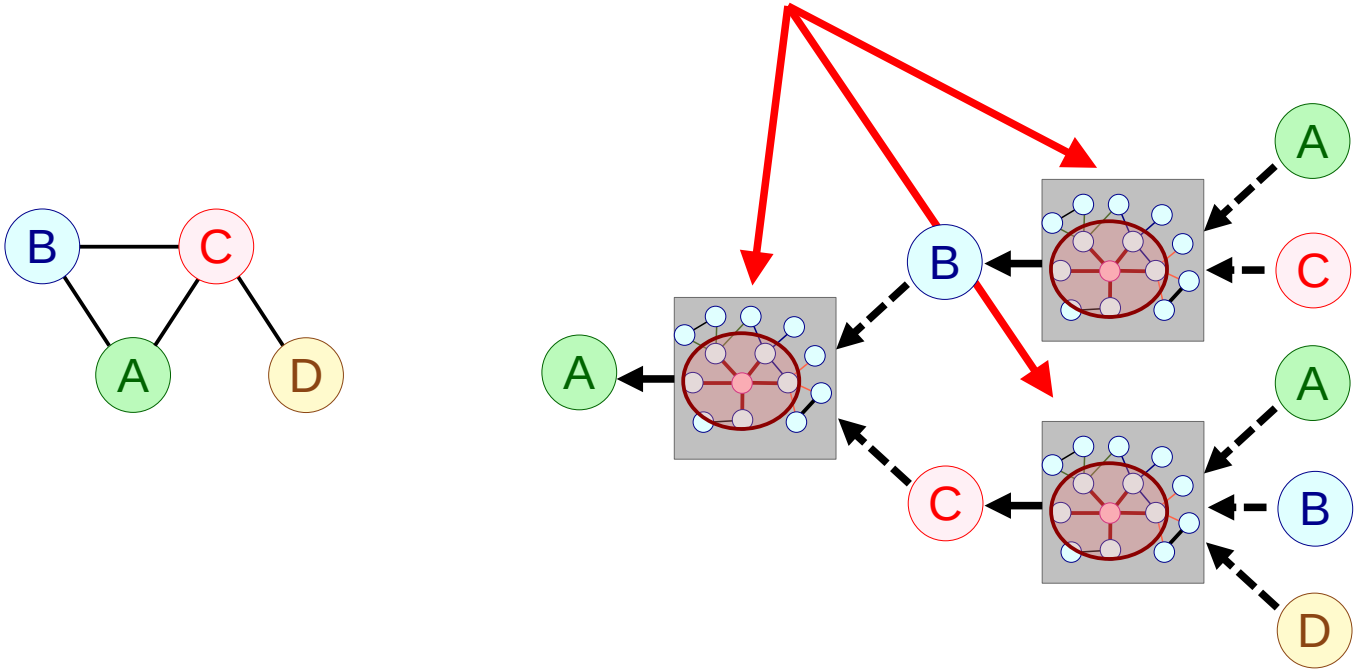
# Many different ways to update GNNs

Can be message passing (MLP),
attention-based

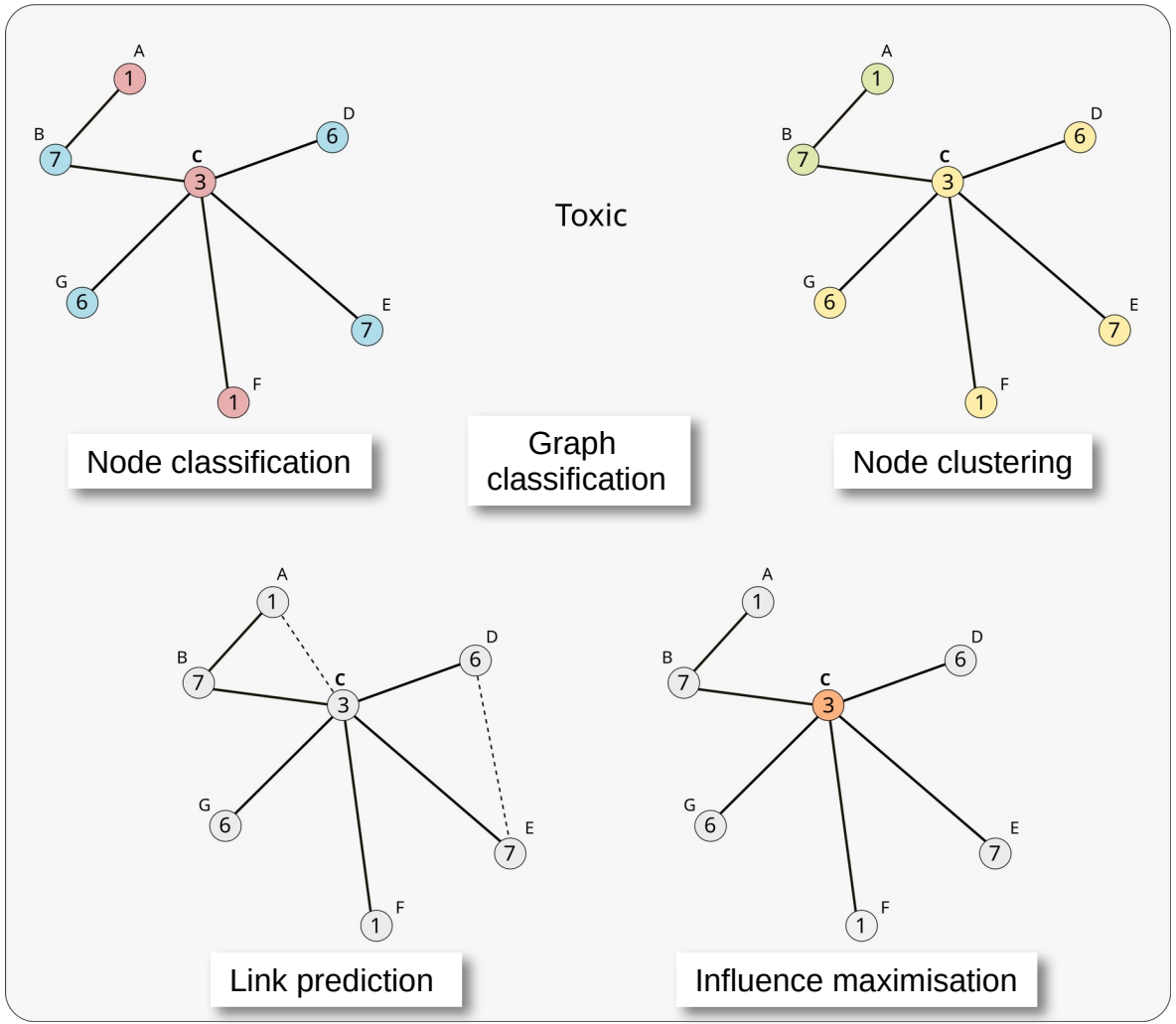


# Many different ways to update GNNs

Can be message passing (MLP),
attention-based, convolutions, etc.



# What can we do with GNN?




Source: Understanding Convolutions on Graphs
https://distill.pub/2021/understanding-gnns/

See also: A Gentle Introduction to Graph Neural Networks
https://distill.pub/2021/gnn-intro/

Both by Google Research teams

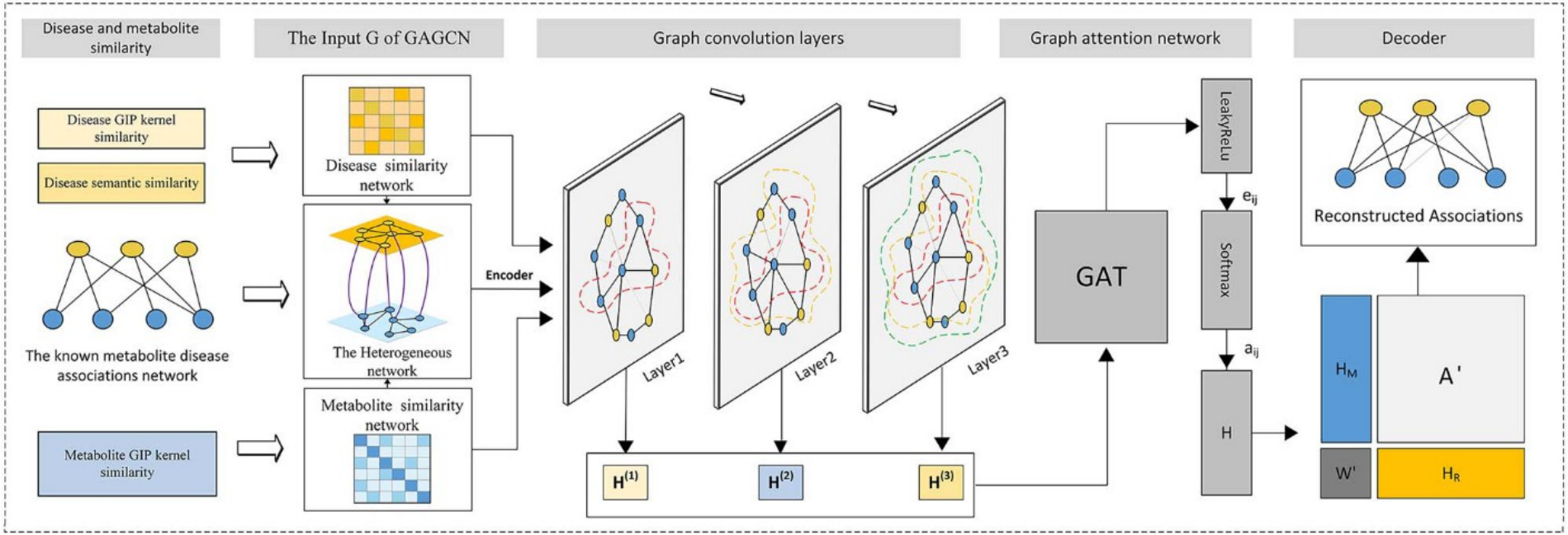
# Metabolites-disease association

OXFORD

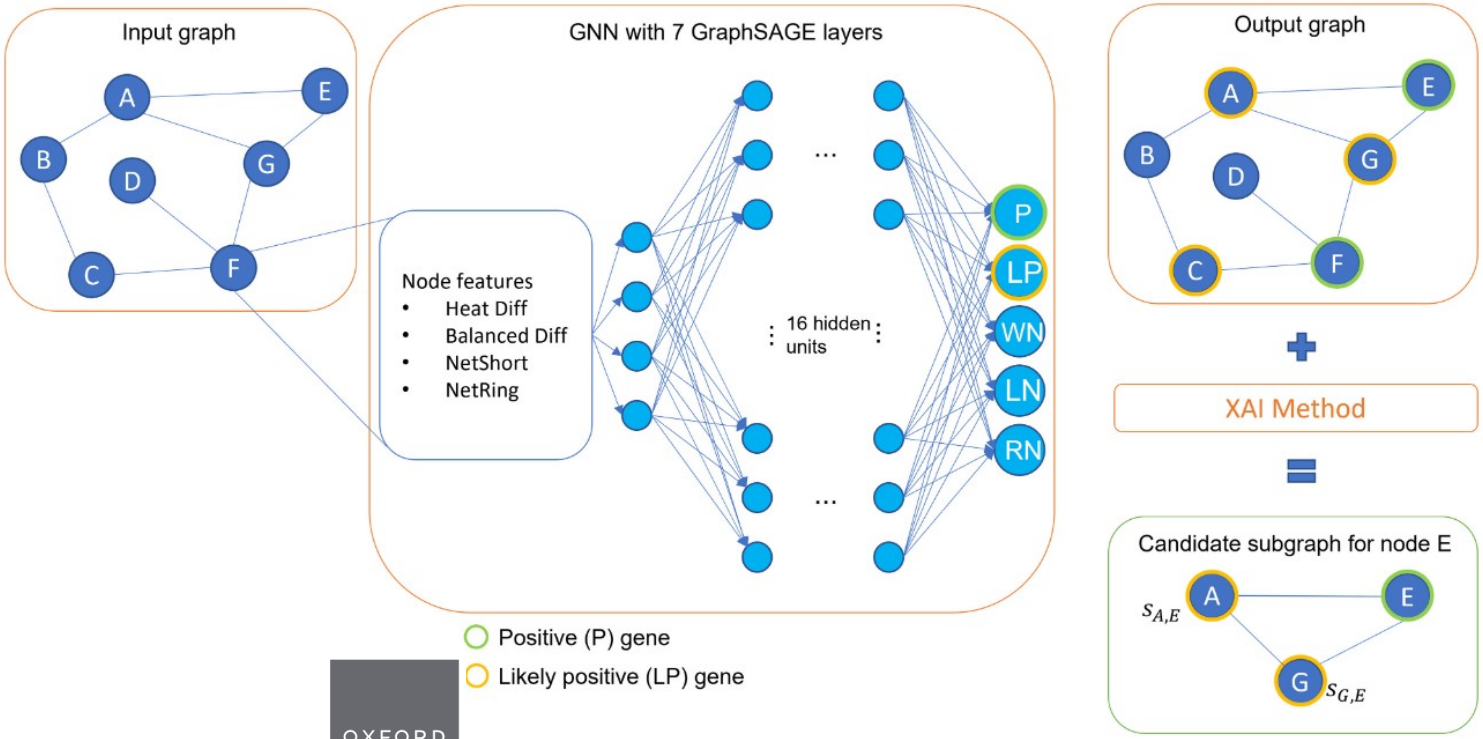
*Briefings in Bioinformatics*, 2022, **23**(4), 1–11
https://doi.org/10.1093/bib/bbac266
Advance access publication date: 12 July 2022
**Problem Solving Protocol**

## A deep learning method for predicting metabolite–disease associations via graph neural network

Feiyue Sun, Jianqiang Sun and Qi Zhao



# Gene-disease associations




*Bioinformatics*, 2023, **39(8)**, btad482
https://doi.org/10.1093/bioinformatics/btad482
Advance access publication 2 August 2023
**Original Paper**

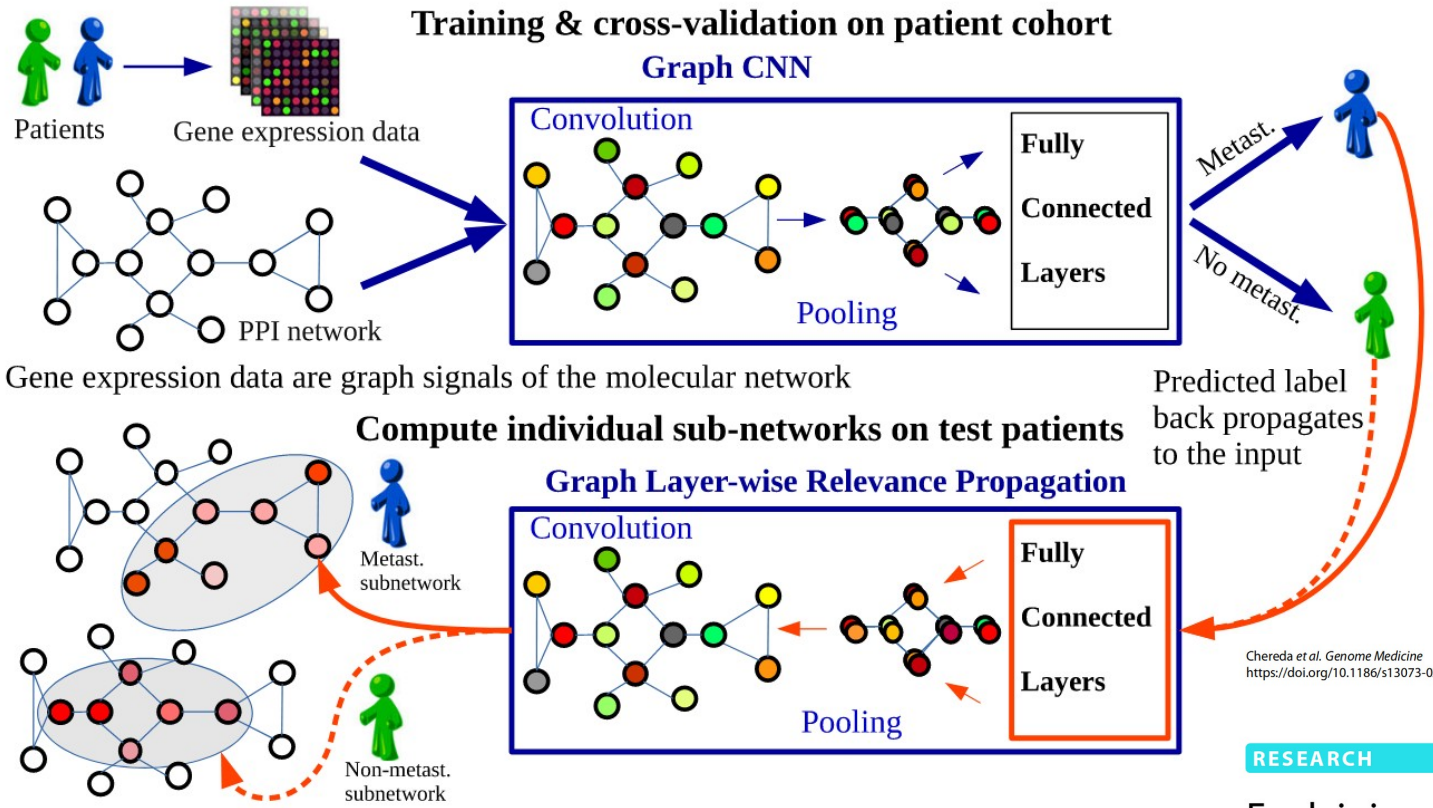
OXFORD

Systems biology

## XGDAG: explainable gene–disease associations via graph neural networks

**Andrea Mastropietro *, Gianluca De Carlo , Aris Anagnostopoulos**

# Explaining predictions




Chereda *et al. Genome Medicine* (2021) 13:42
https://doi.org/10.1186/s13073-021-00845-7

Genome Medicine

RESEARCH Open Access

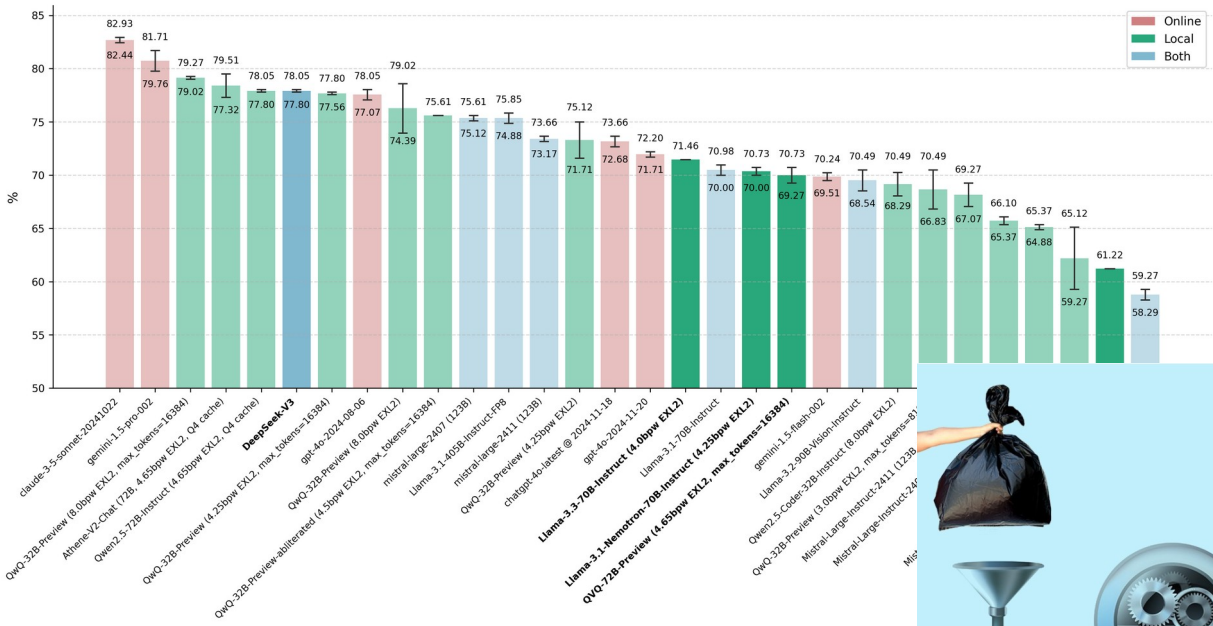
## Explaining decisions of graph convolutional neural networks: patient-specific molecular subnetworks responsible for metastasis prediction in breast cancer

Hryhorii Chereda[1], Annalen Bleckmann[2], Kerstin Menck[2], Júlia Perera-Bel[3], Philip Stegmaier[4], Florian Auer[5], Frank Kramer[5], Andreas Leha[6] and Tim Beißbarth[1,7*]

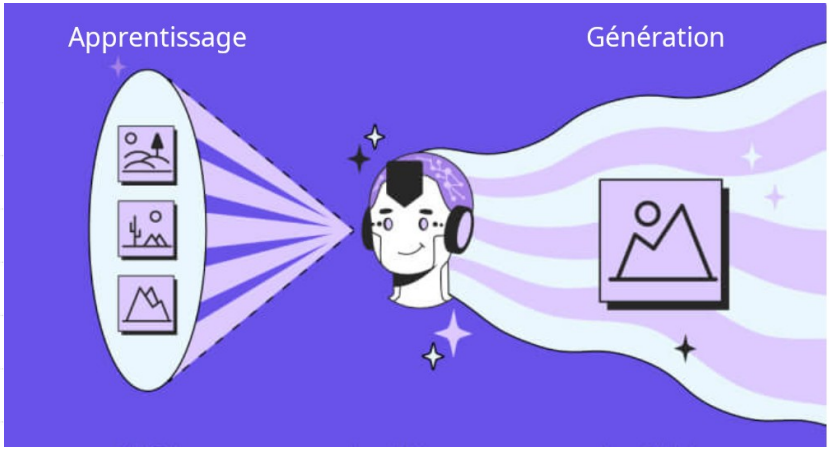
# Some warnings

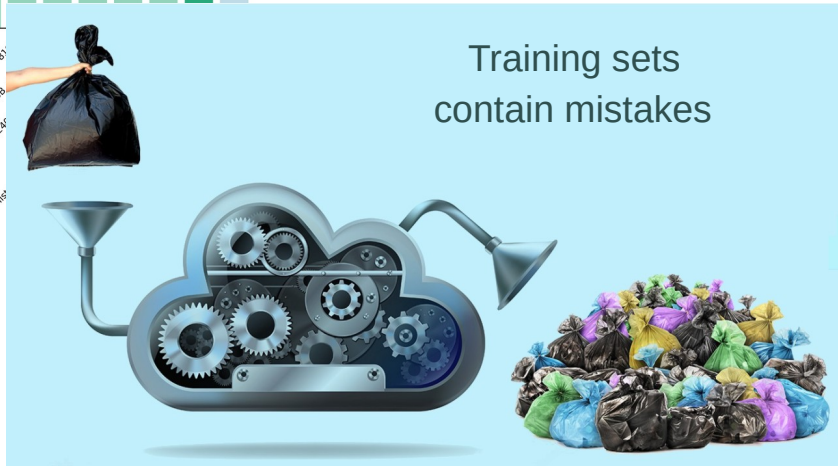
# AI is not infallible

AI models are not perfect;
training sets are incomplete,
biased, contextualised → poor generalisation



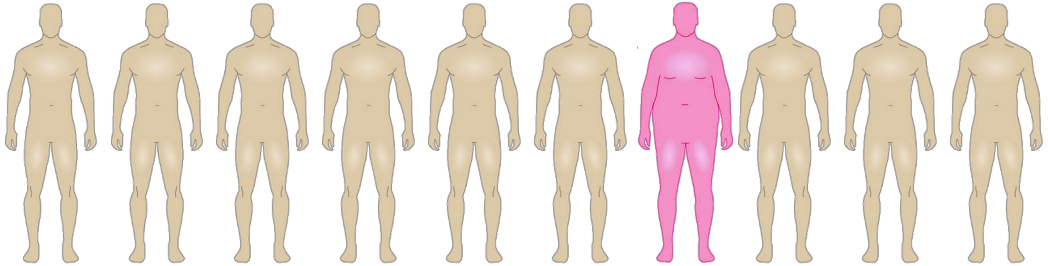

Generative AI systems hallucinate;
a consequence of their very versatility.



# AI can be biased

The learning sets are unbalanced
(How can obesity be easily predicted with 90% accuracy?)



Learning sets are not representative
of the real-world context

# Equity of access to healthcare

Models may not be available for all

Using models can be costly

Models may require data that is difficult to collect

Model usage and interpretation of results may depend on context, e.g. practitioners

AI is a fantastic tool that improves on existing approaches and opens up new avenues for the prediction, screening, diagnosis, and treatment of various conditions.

However, it must be used in a controlled and careful manner...

...just like all other healthcare tools!

PreciDIAB
National Center for
Precision Diabetic Medicine

Egid
European
Genomic
Institute
for Diabetes

Région
Hauts-de-France

UMR 1283 8199
Functional (Epi)genomics and Mechanisms of Type 2 Diabetes
and Related Diseases