

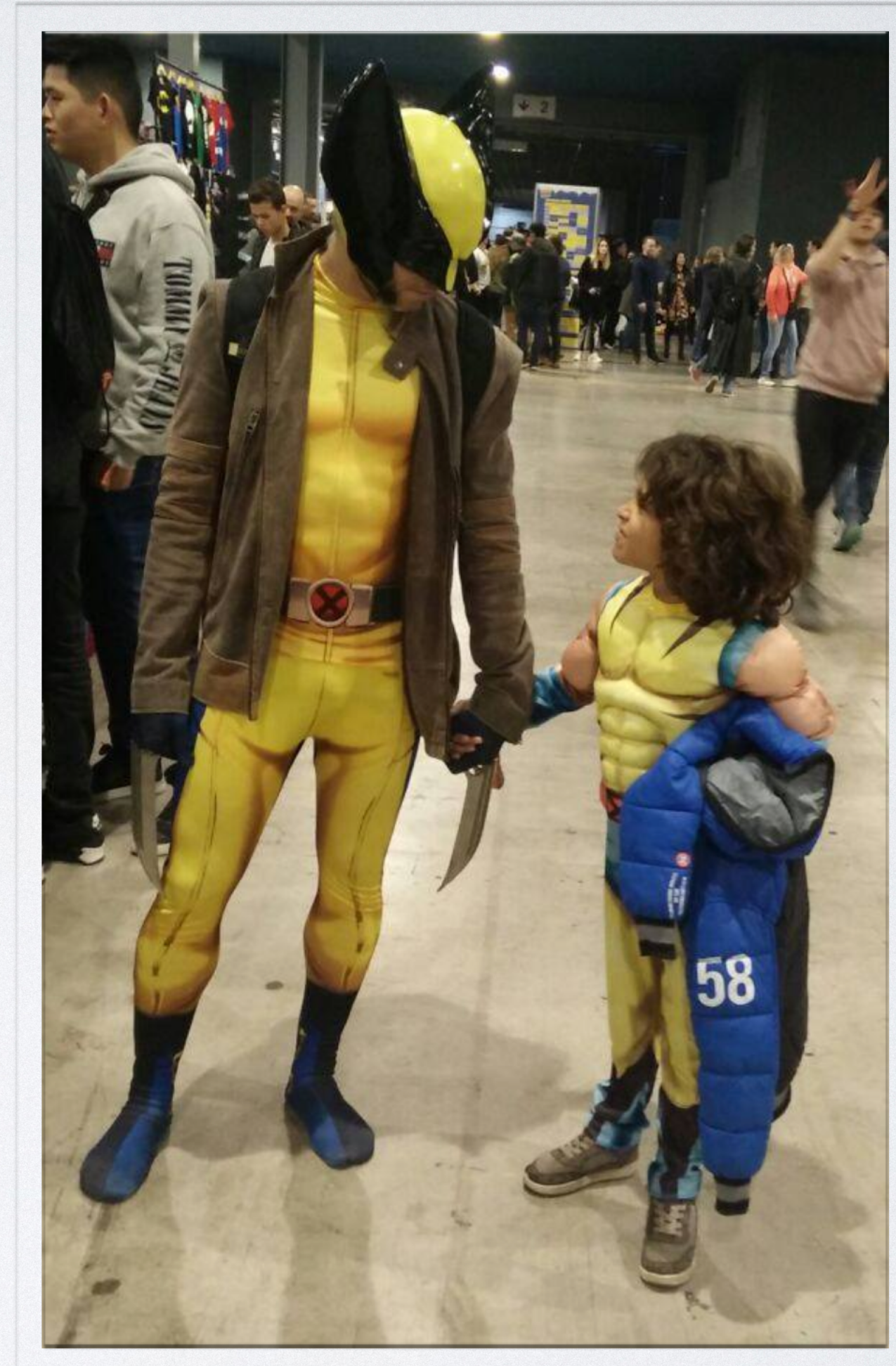


SineReLU

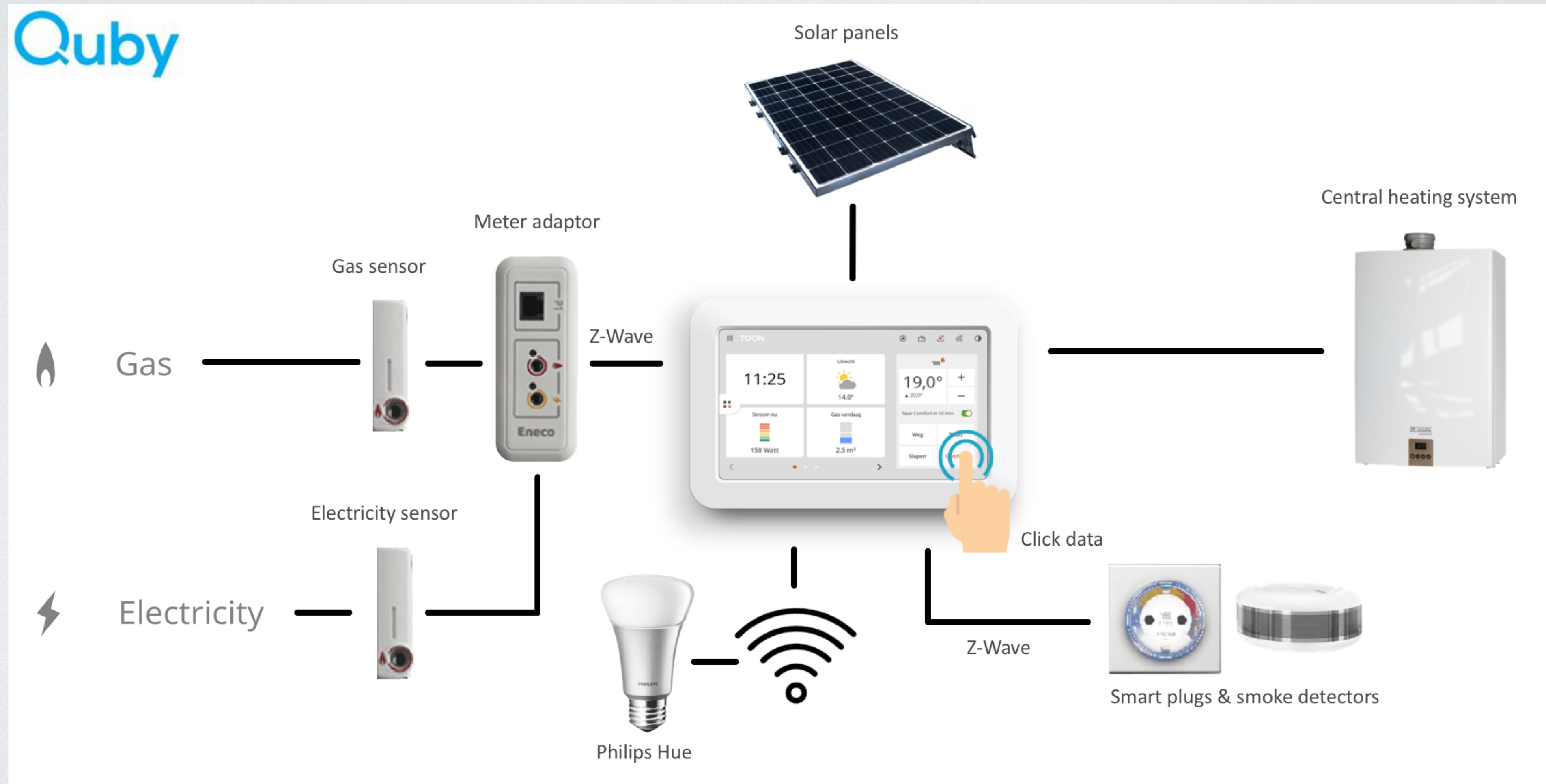
An Alternative to the ReLU Activation Function

Wilder Rodrigues

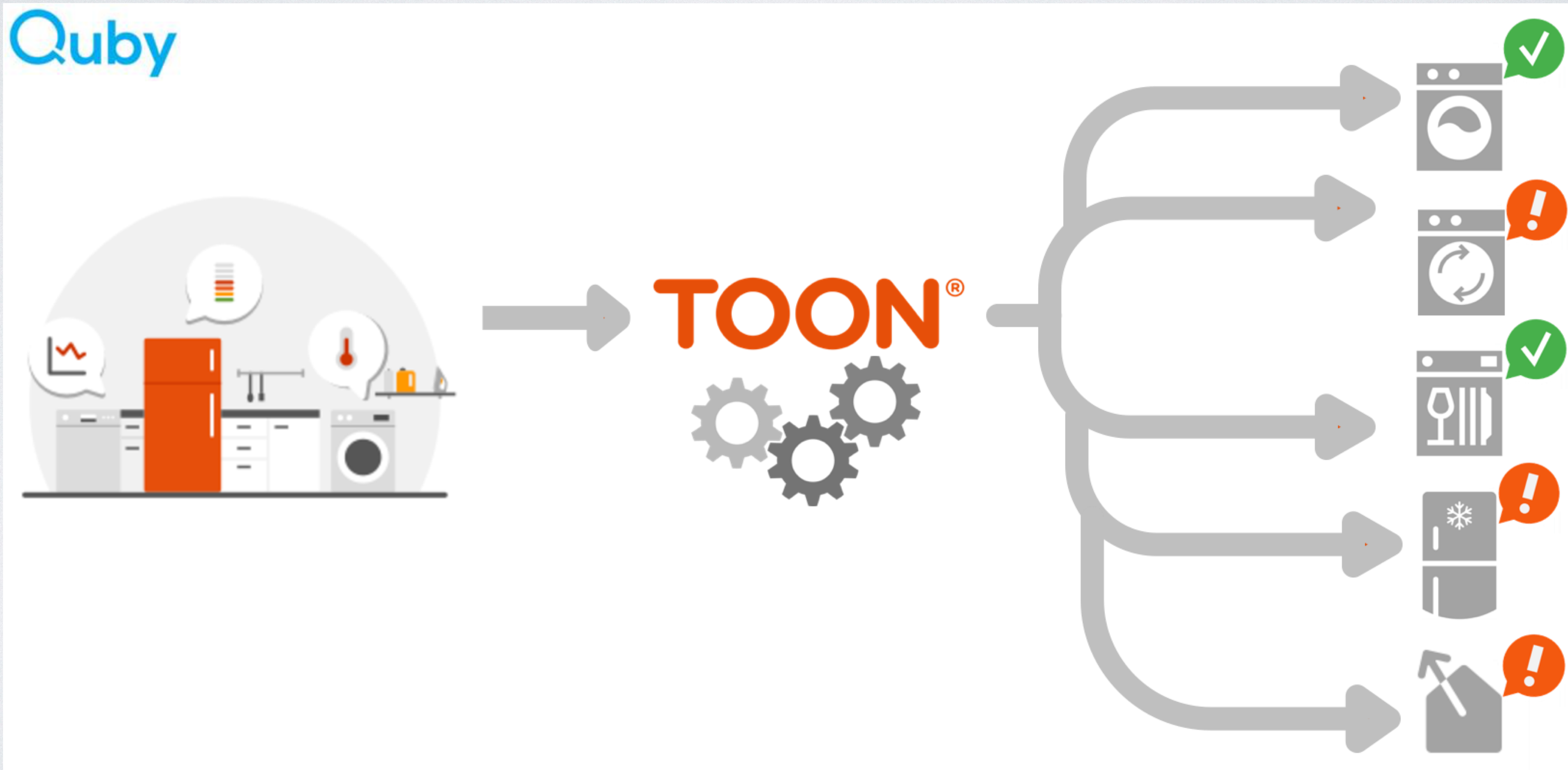
- Machine Learning Engineer @QubyEnergy;
- City.AI Ambassador;
- IBM Watson AI XPRIZE contestant;
- Guest attendee at the AI for Good Summit at the UN in 2017/2018.
- Kaggle;
- X-Men geek; and
- Family man and father of 3.



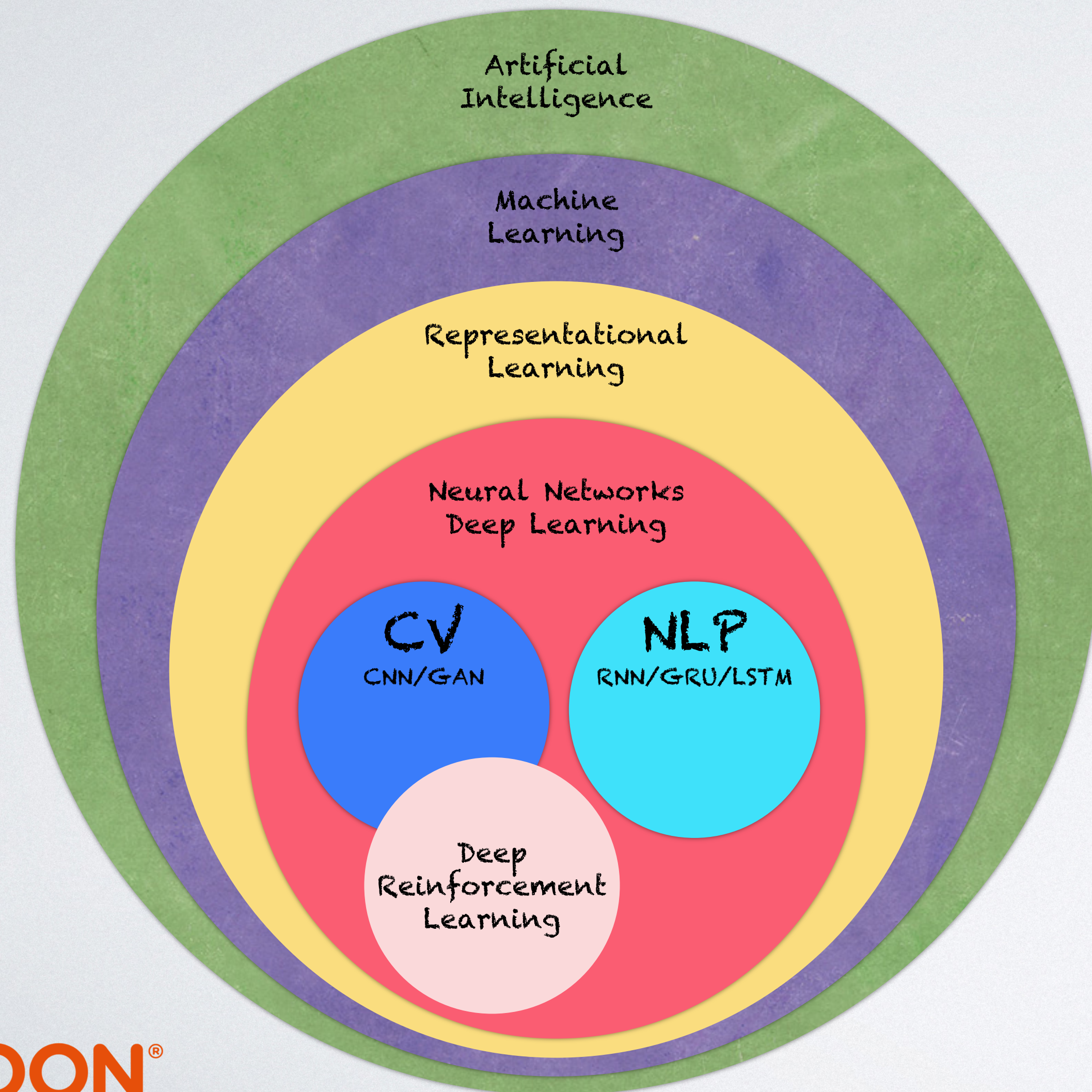
A Word from our Sponsors



A Word from our Sponsors



Inspiration



Make computers do things that require intelligence when done by humans.

Make machines find patterns without explicitly programming them to do so.

Raw data is provided and the engineering of the features is done automatically.

By far the most important way of Representation Learning; vastly used.

Aims to give a machine the same ability we have.

“Say ‘friend’ and enter.”. All about interactions between computers and humans, make a machine understand what is written or spoken.

Gets closer to the popular perception of AI as a system to replicate the human cognitive capacity.

How did I get here?



**NOT SURE IF HEAD IS ABOUT TO
EXPLODE FROM CONFUSION**

**OR BRAIN IS GETTING BIGGER FROM
KNOWLEDGE**

quickmeme.com

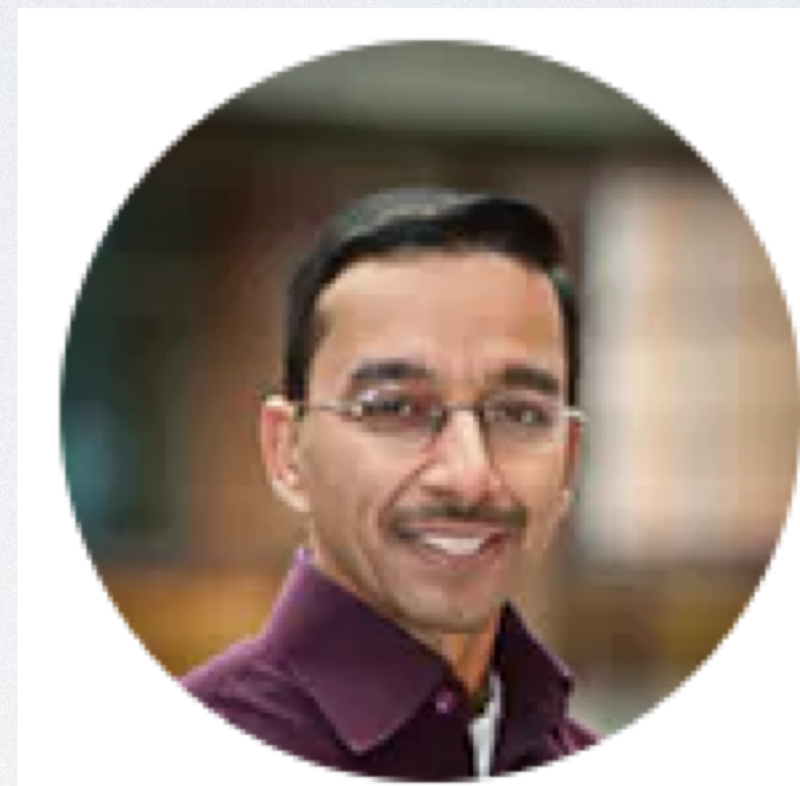
Learning from the best



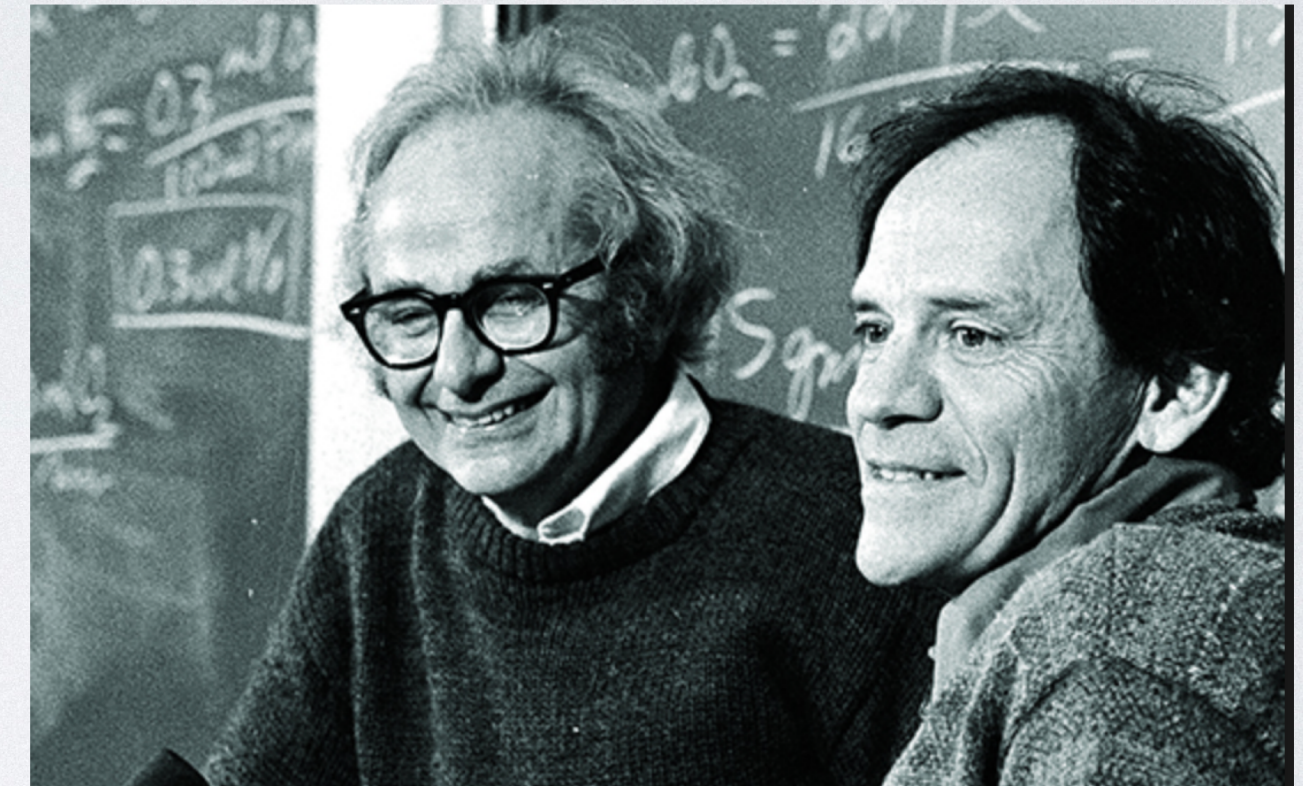
Andrew Ng



Adrienne Fairhall

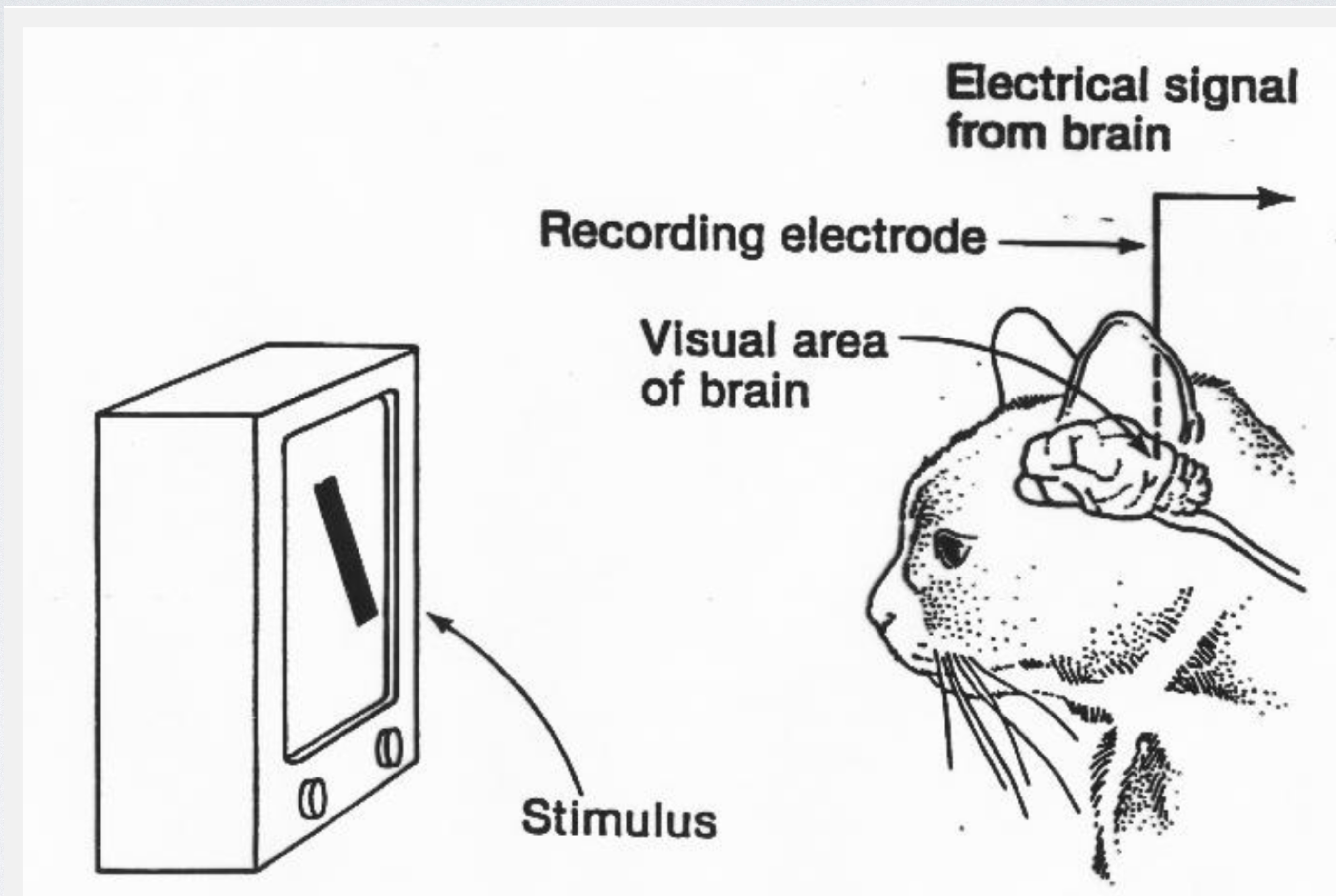


Rajesh Rao



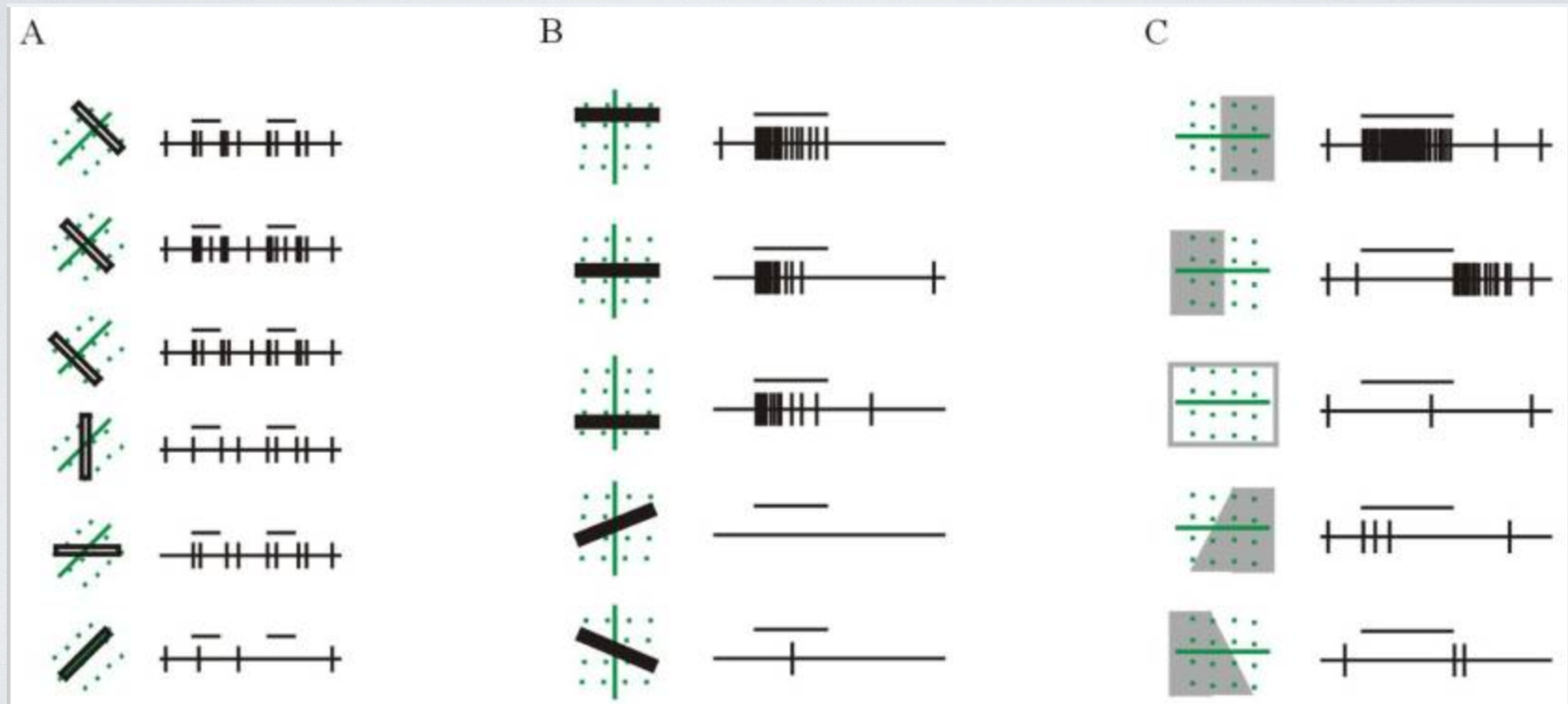
David Hubel and Torsten Wiesel

Inside the Brain... of a Cat.



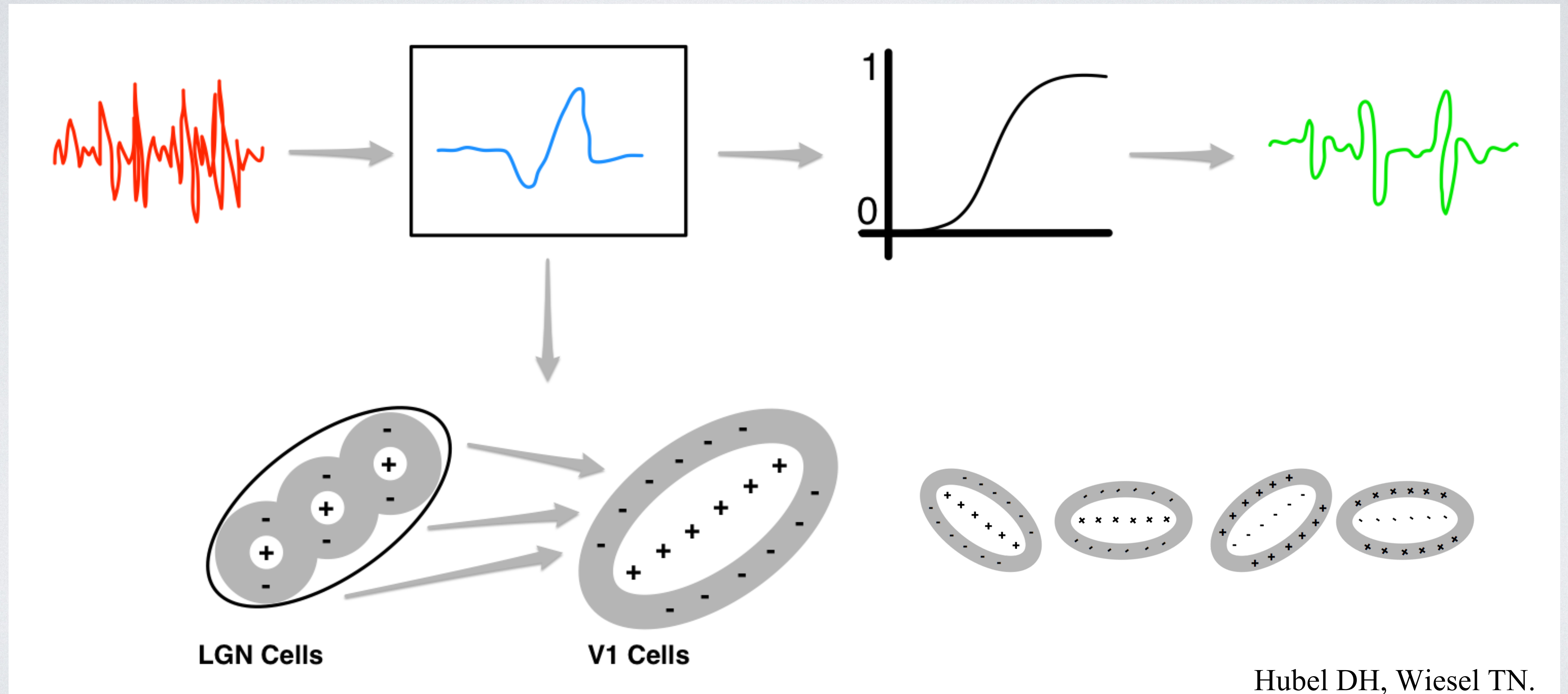
Hubel and Wiesel Experiment

Any Similarities?



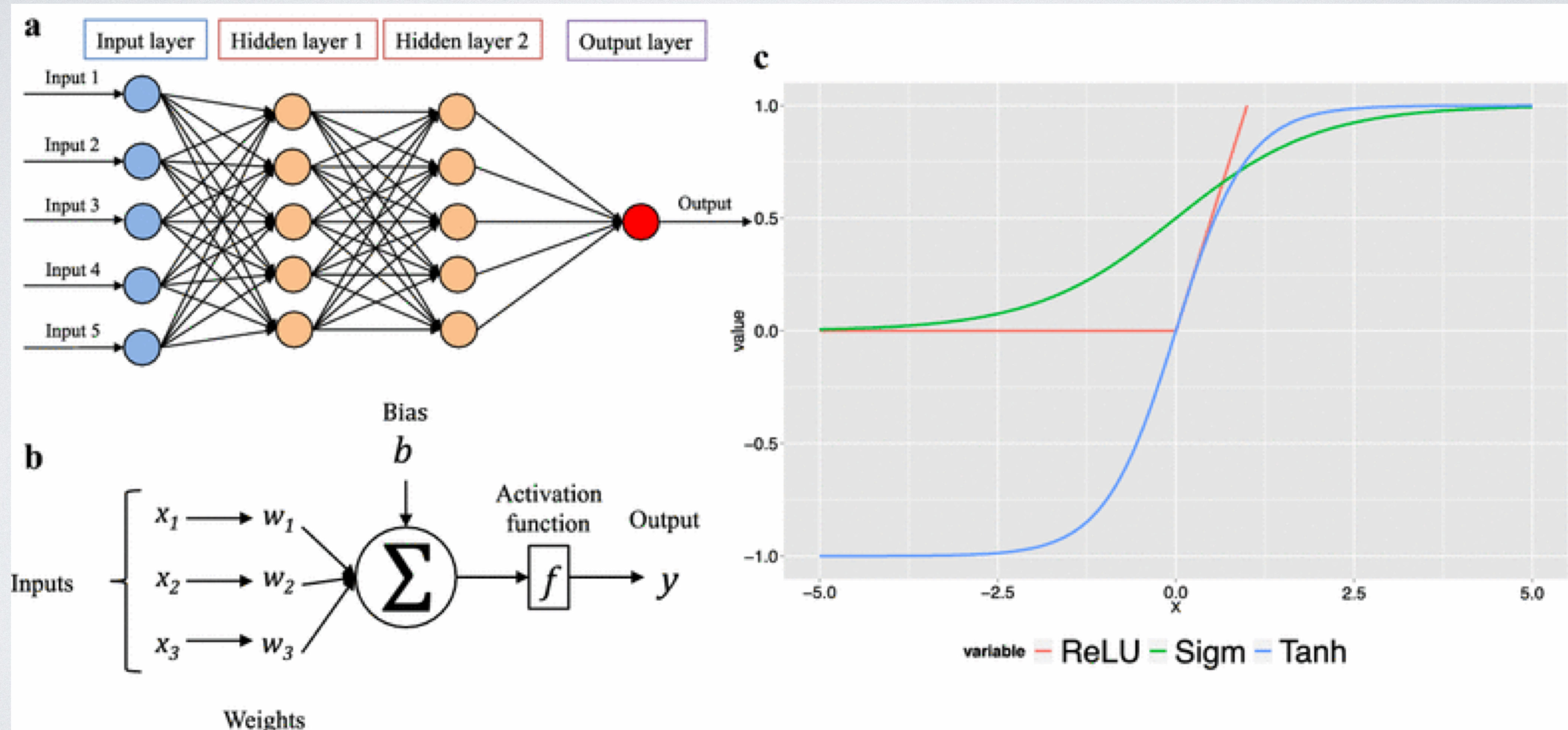
Hubel DH, Wiesel TN.

Lateral Geniculate Nucleus and V1 Fields





Activation Functions

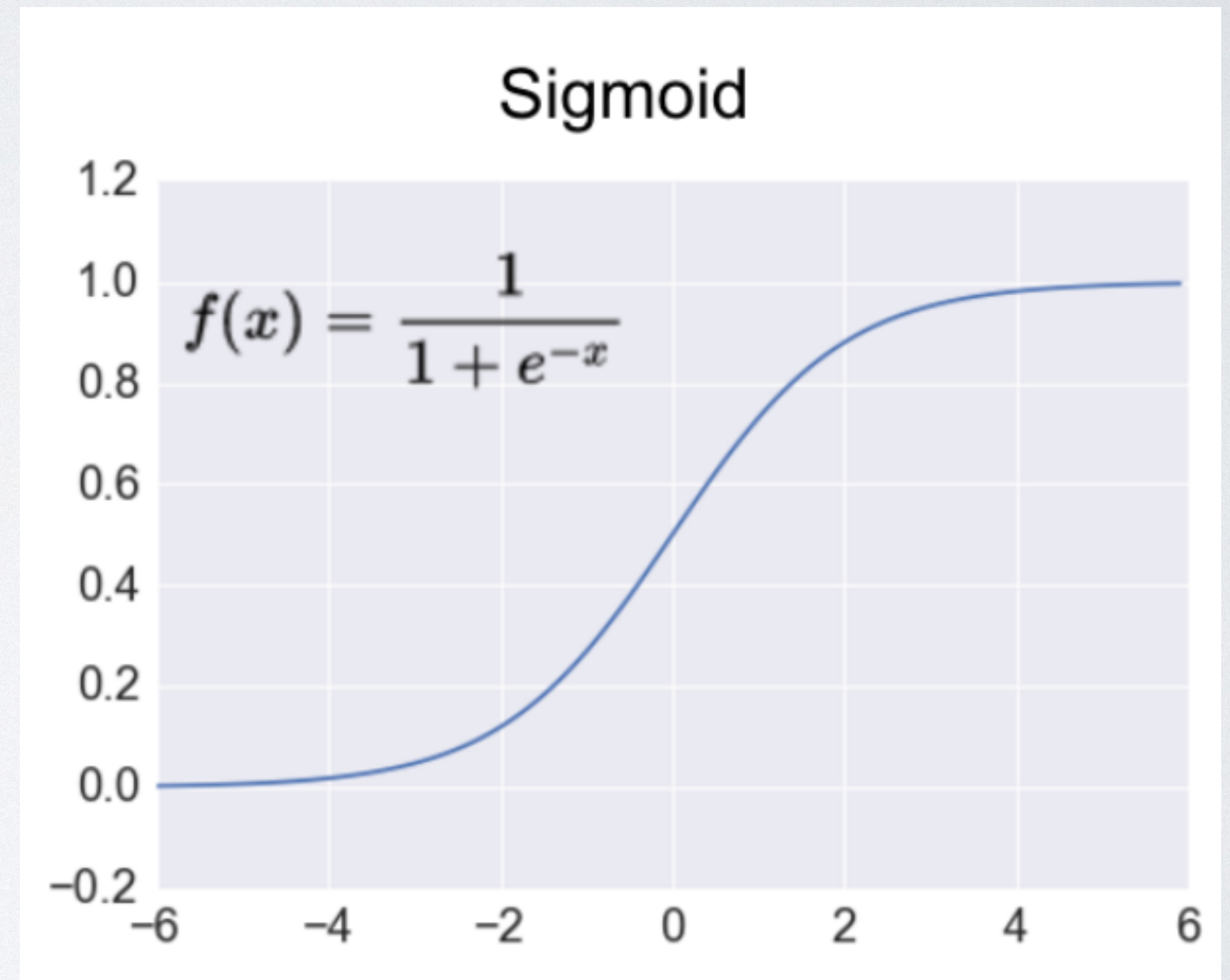


Why Another Activation Function?



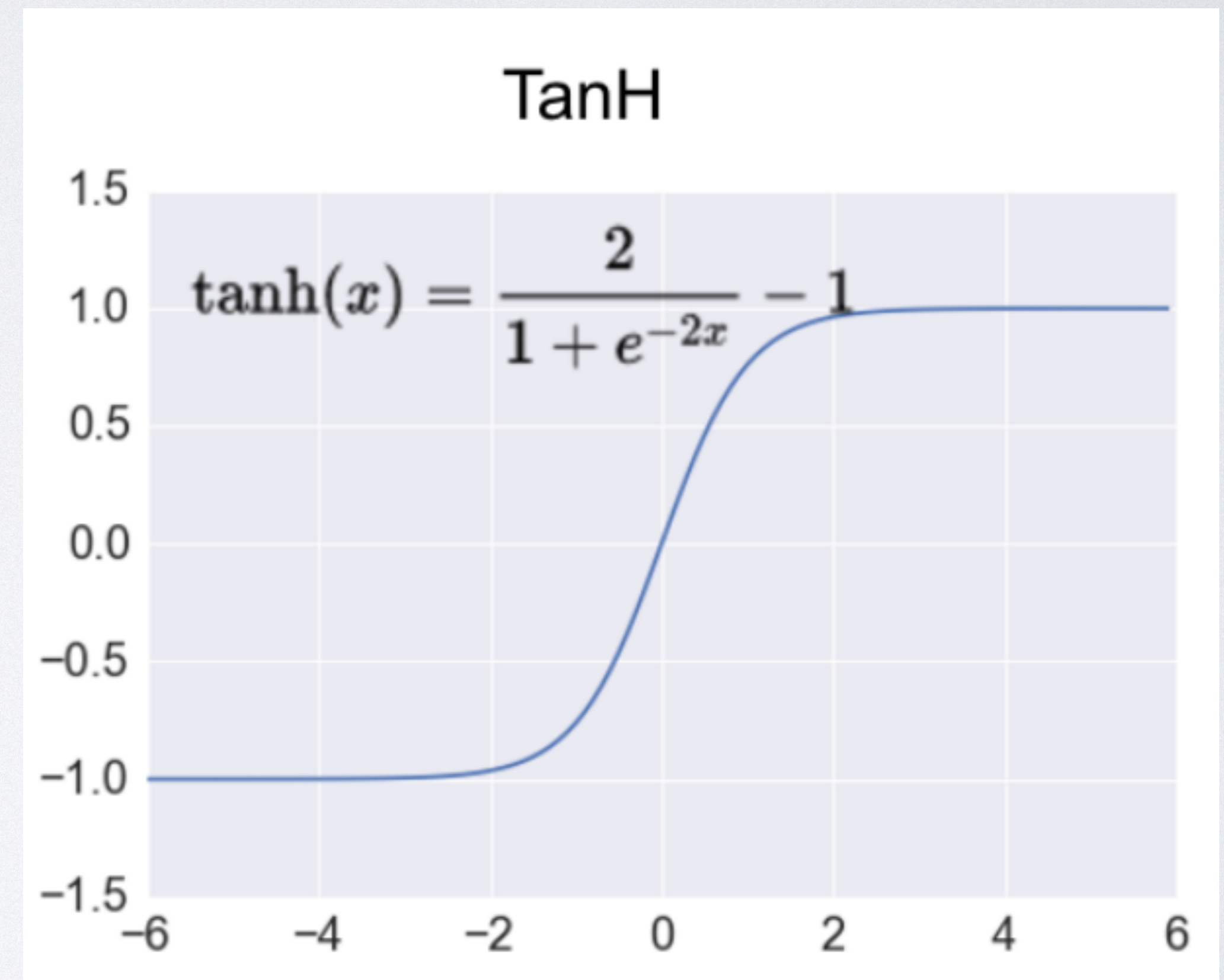
Activation Functions

- Sigmoid
 - Vanishing/Exploding Gradients.



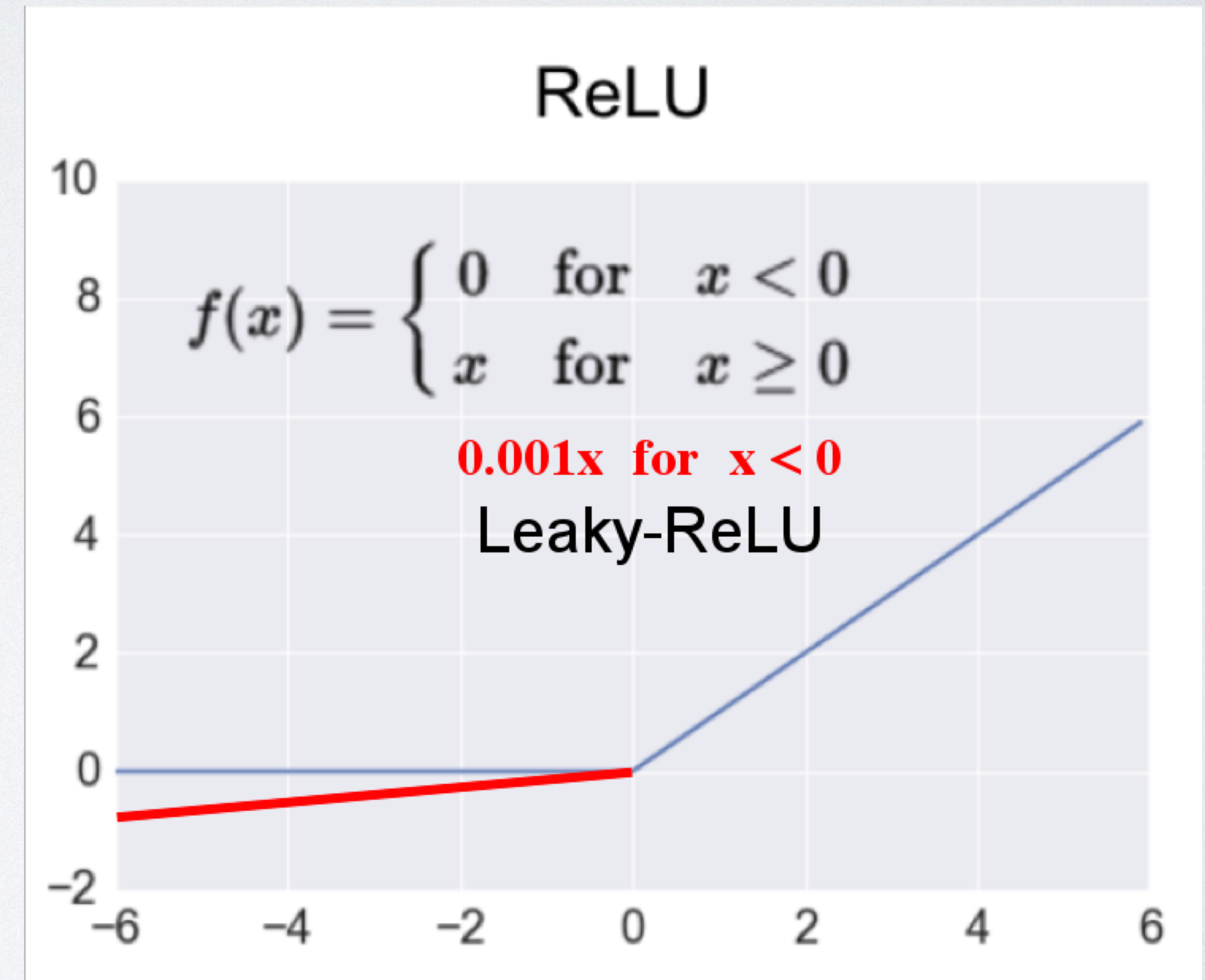
Activation Functions

- Tanh
 - Vanishing/Exploding Gradients.



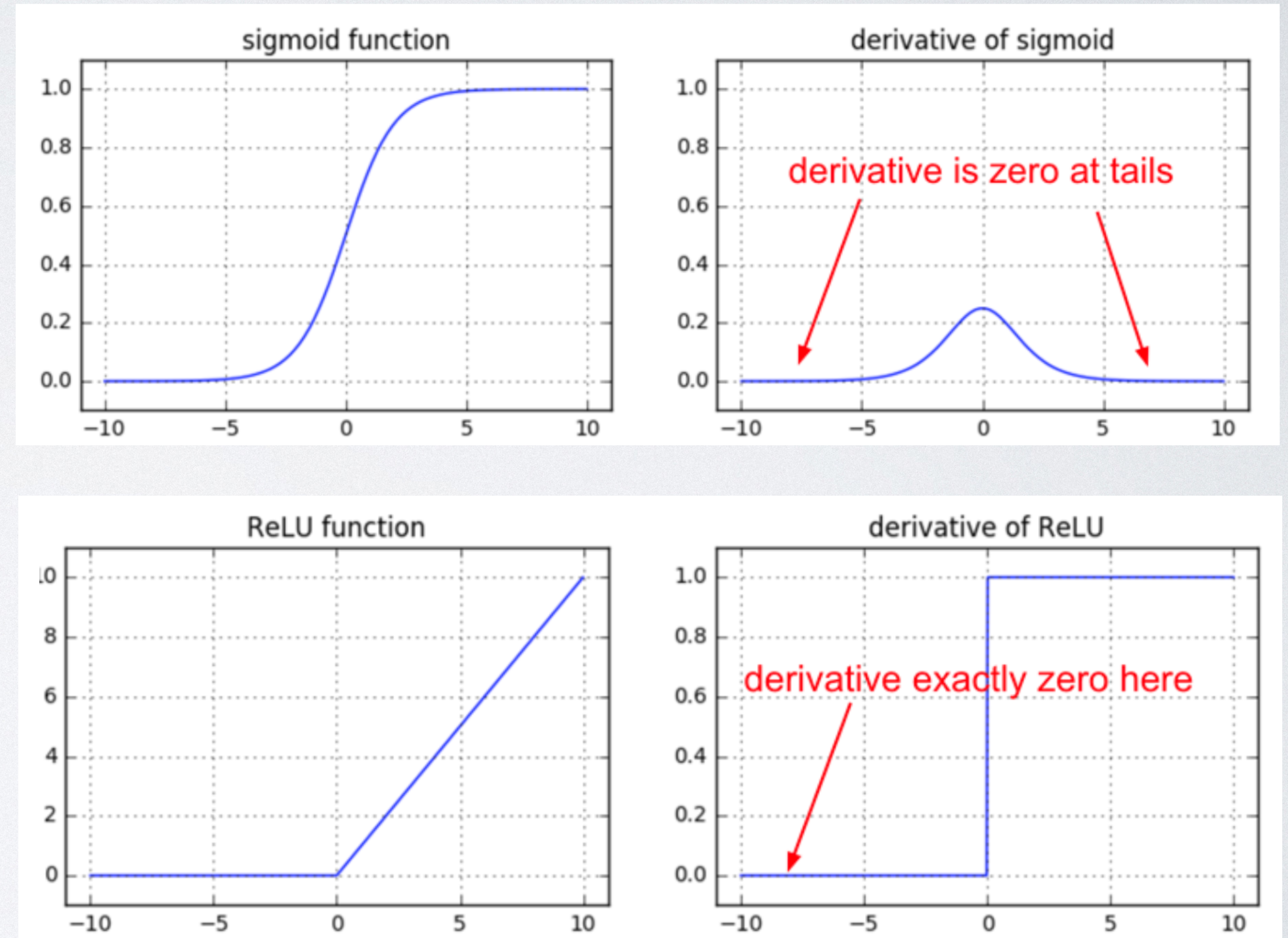
Activation Functions

- ReLU
 - Dying ReLU



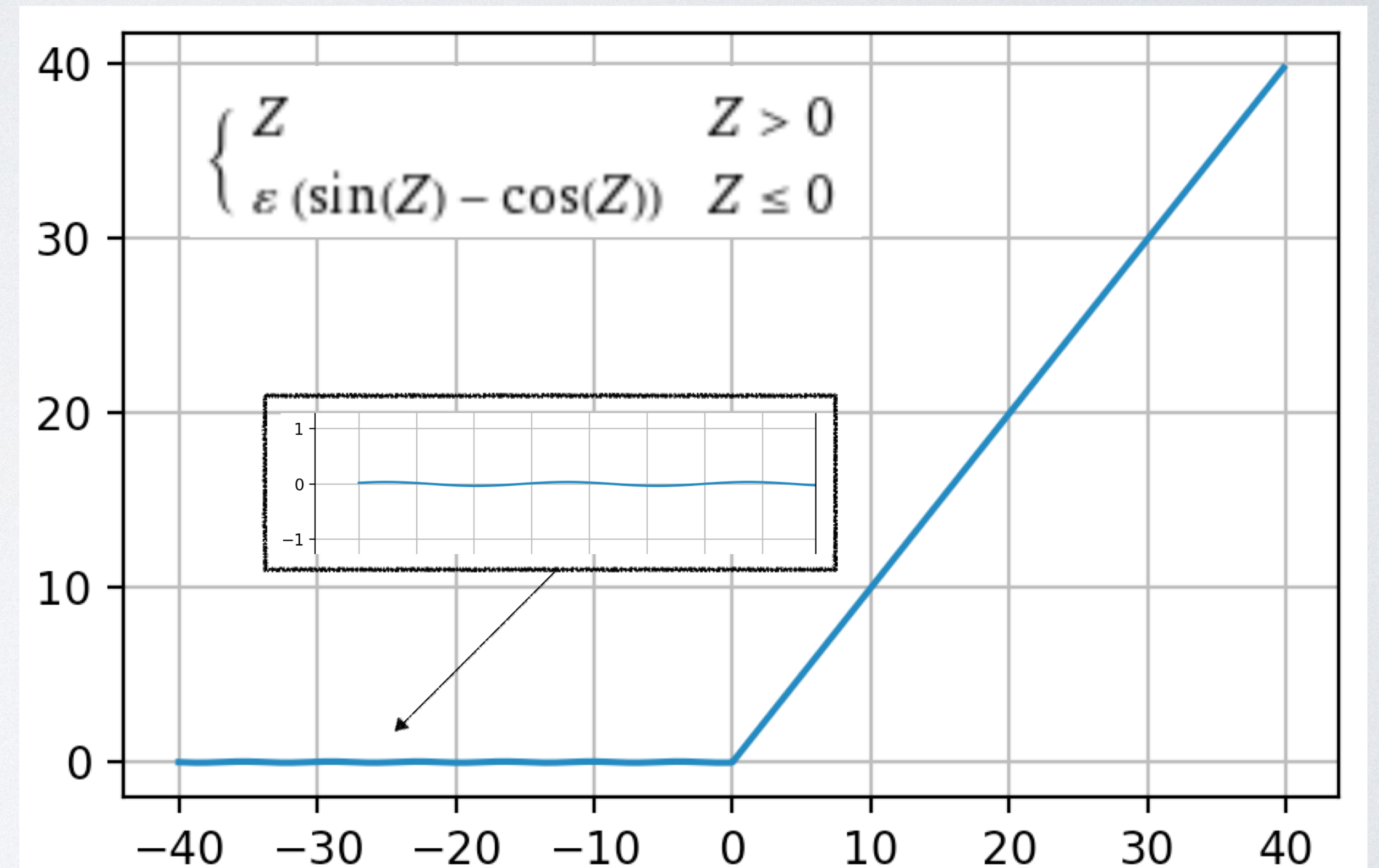
Forward Pass & Back Propagation

- Sloppiness with the weight initialisation can saturate non-linearities and cause the learning process to stop.
- Or, when using ReLU and neurons get clamped to zero in the forward pass, then the weights will get zero gradient.



Even if it is not true, it is well conceived.

- The epsilon hyper-parameter is used to control the amplitude of the wave.
- It mitigates the Dying ReLU issue by making the function differentiable.





How is it with the SineReLU?

$$\cos(0)$$

$$\sin(0)$$

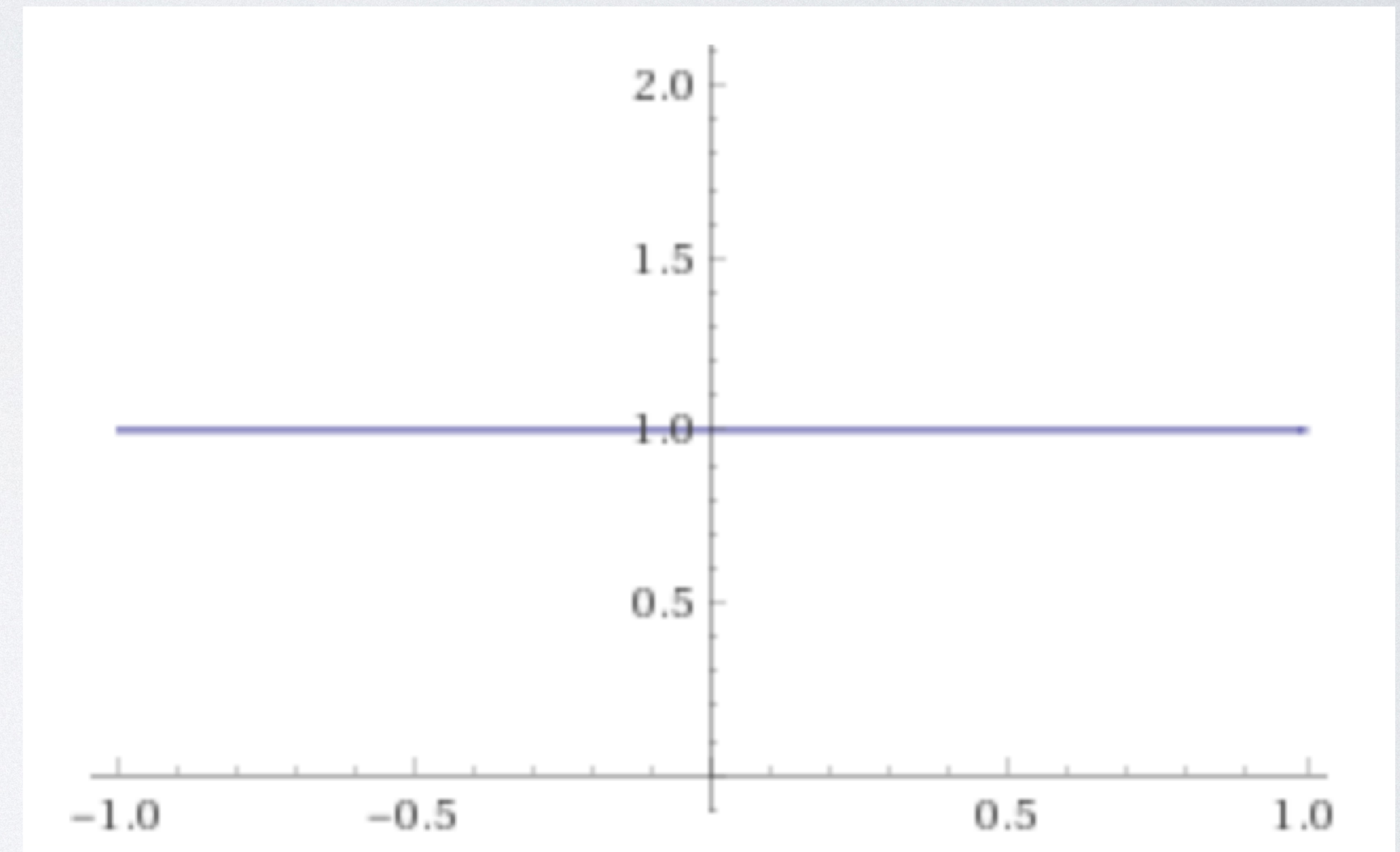
$$\cos(0) = \frac{\text{adjacent}}{\text{hypotenuse}}$$

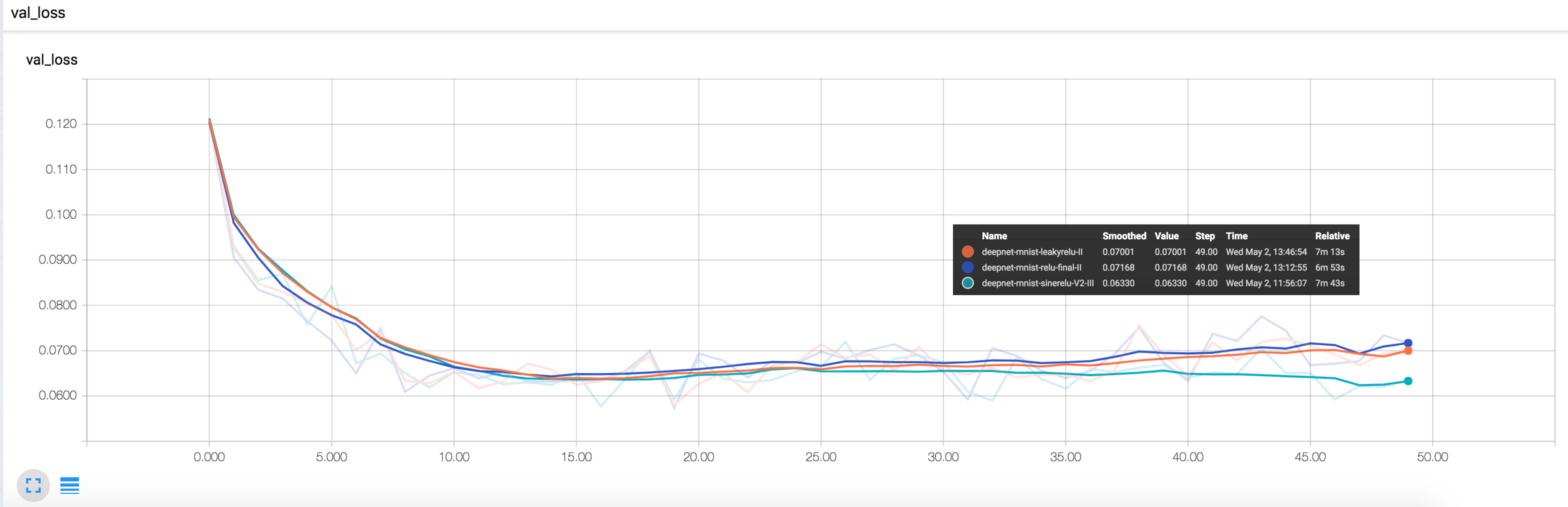
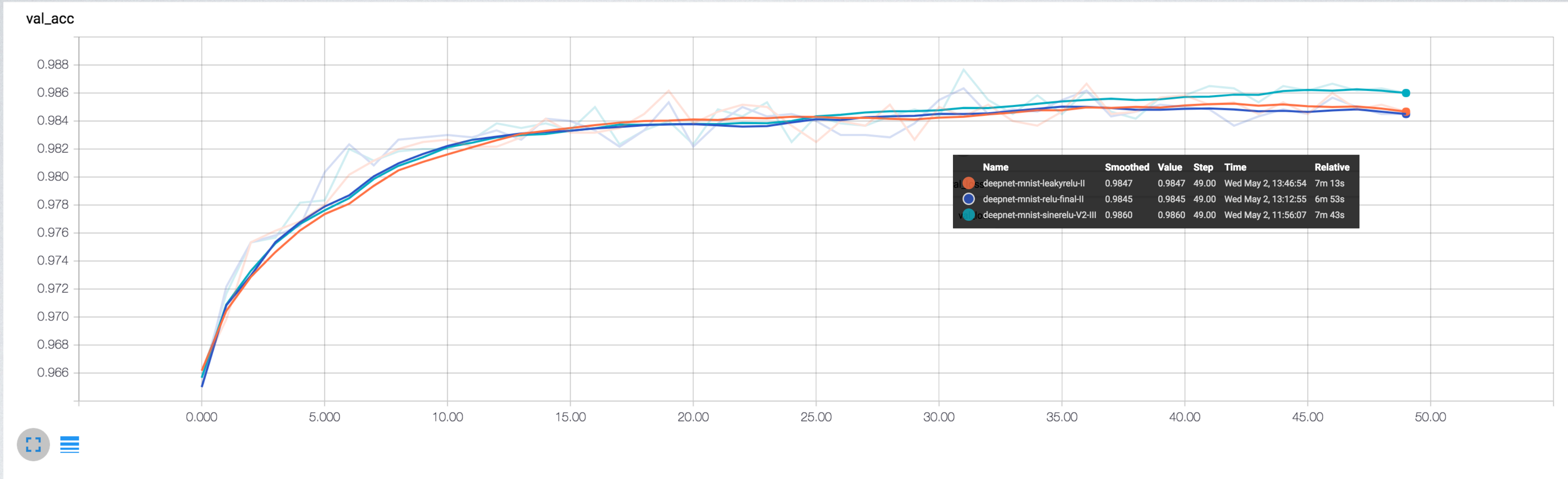
+

$$\sin(0) = \frac{\text{opposite}}{\text{hypotenuse}}$$

$$\cos(0) = \frac{1}{1}$$

$$\sin(0) = \frac{0}{1}$$





Dataset	Function	Network	Final Loss	Final Accuracy	STD Loss	AUC
MNIST	SineReLU	DNN	0.0765	98,33%	0.053755	—
		CNN	0.0197	99,50%	0.036901	—
MNIST	ReLU	DNN	0.0823	98,29%	0.057369	—
		CNN	0.0203	99,39%	0.045921	—
IMDB	SineReLU	DNN	0.3268	85,90%	0.176337	93,54 %
		CNN	0.2868	87,83%	0.123844	95,09 %
IMDB	ReLU	DNN	0.3265	85,77%	0.177140	93,54 %
		CNN	0.4135	87,57%	0.163340	94,85 %

Where to find it?



Keras

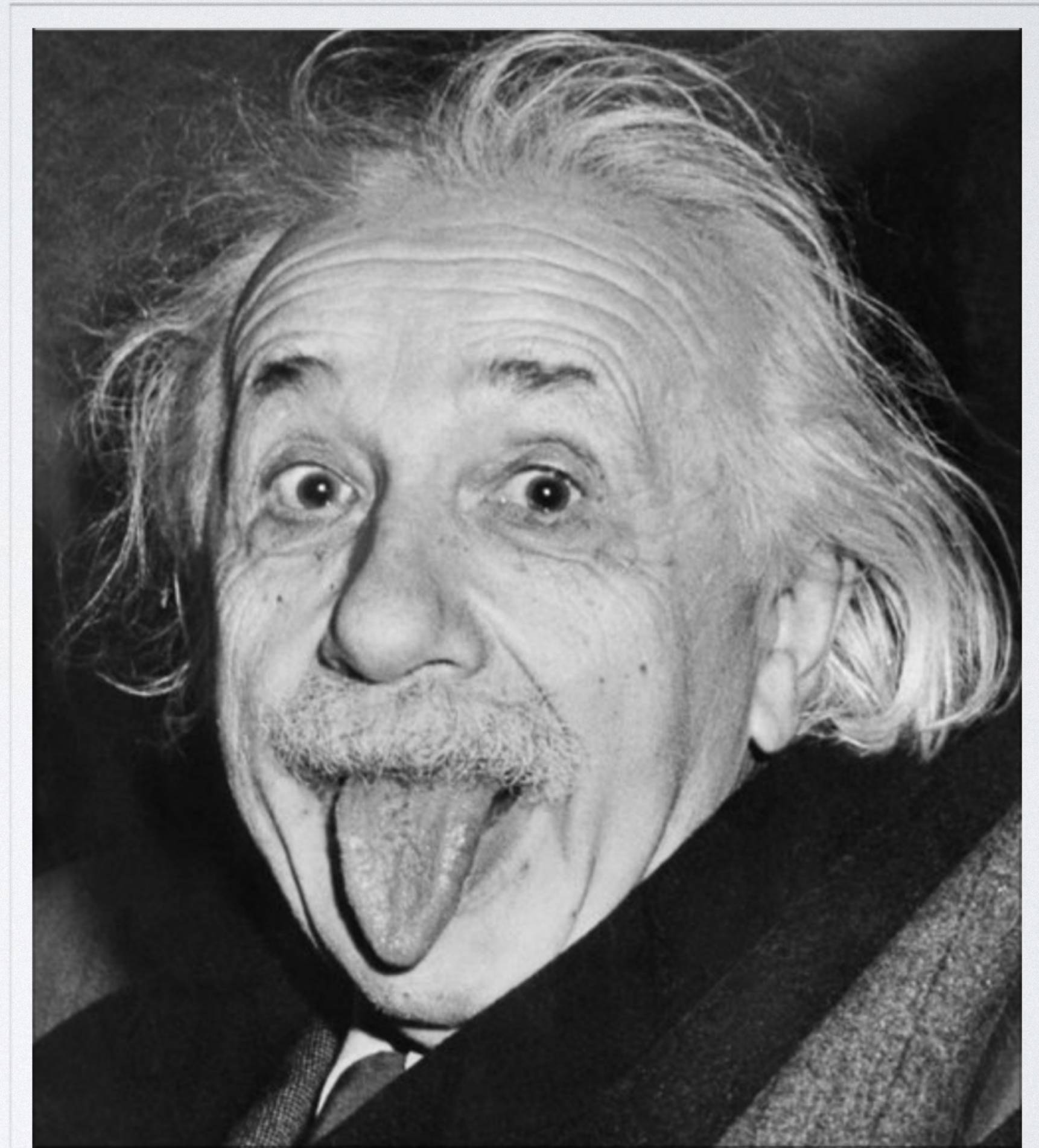
So far, who is interested?



In the middle of difficulty lies opportunity.

- *Not in the academia;*
- *Not in the industry;*
- *Not in a startup applying AI;*
- *And not giving up...*

Machine Learning Engineer at Quby!



References

- Hubel DH, Wiesel TN. Integrative action in the cat's lateral geniculate body. Journal of Physiology (London) 1961
- <https://medium.com/@karpathy/yes-you-should-understand-backprop-e2f06eab496b>
- <https://jcheminf.springeropen.com/articles/10.1186/s13321-017-0226-y>
- https://github.com/keras-team/keras-contrib/tree/master/keras_contrib
- <https://github.com/ekholabs/DLink>

