A background image featuring a person wearing a virtual reality headset. The image is heavily overlaid with a blue-toned circuit board pattern, suggesting themes of technology, AI, and machine learning.

DR. SHIRIN GLANDER

Explaining Keras Image Classification Models with Lime

Data Scientist @ codecentric AG

About me

Nerd-topia

Machine Learning /
Computer Science

Bioinformatics /
Statistics



lab rat



Postdoc

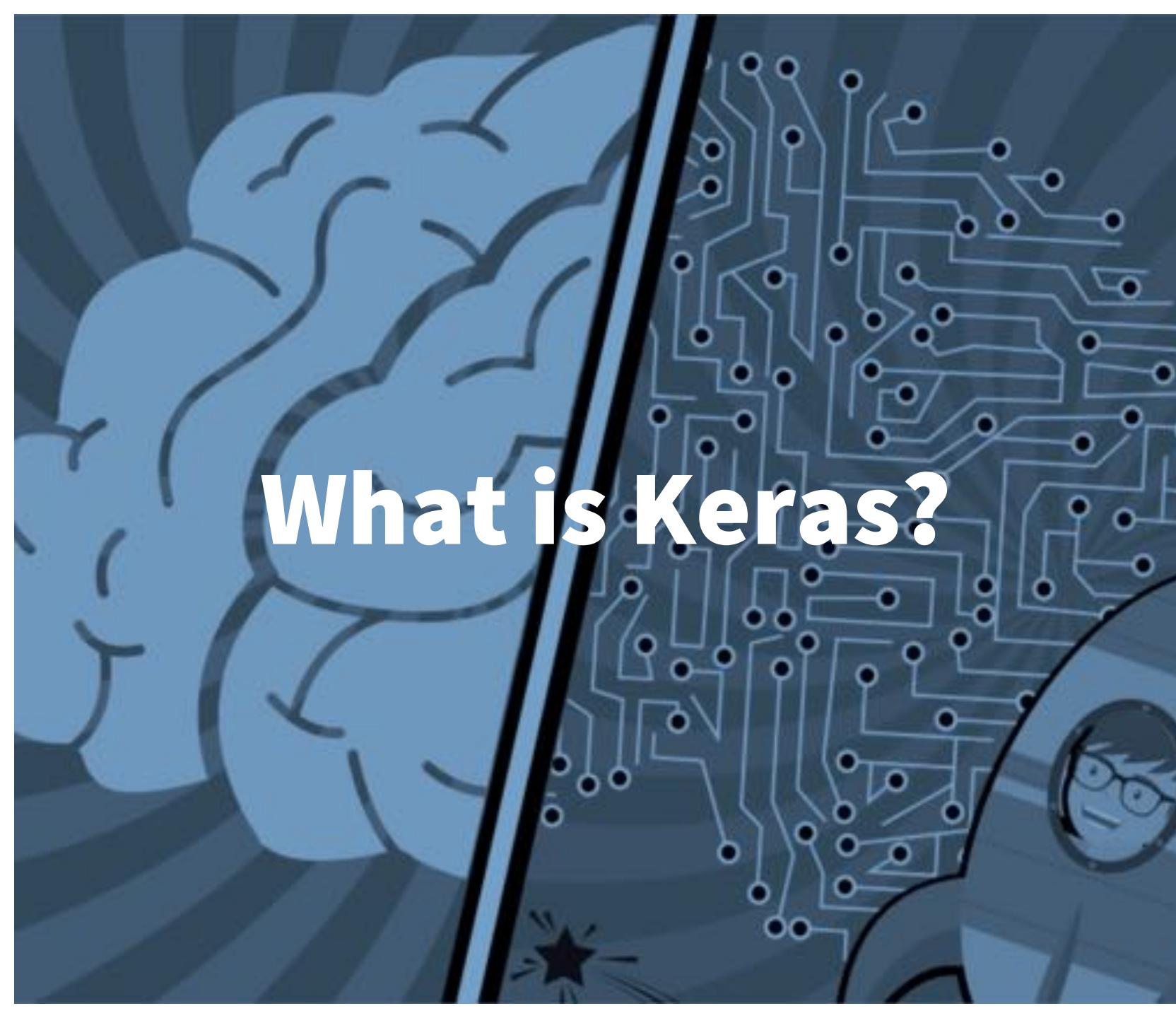


Blog

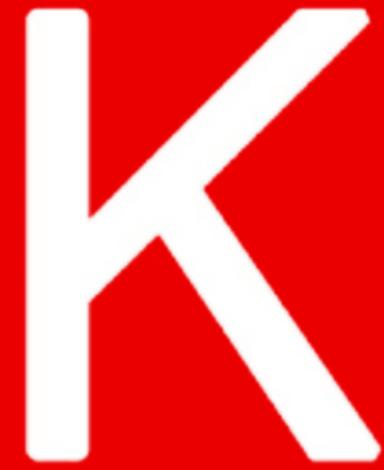


Overview - what you'll hear today

- 1 What is Keras?
- 2 Deep Learning & Neural Nets in a nutshell
- 3 Building an image classifier with Keras
- 4 Explaining image classification models with LIME



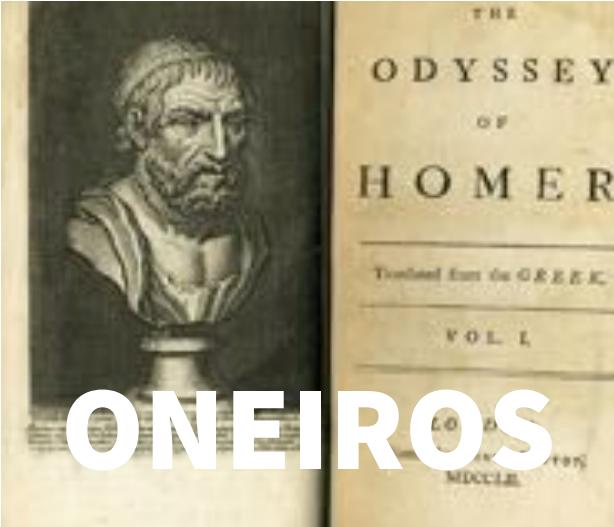
What is Keras?



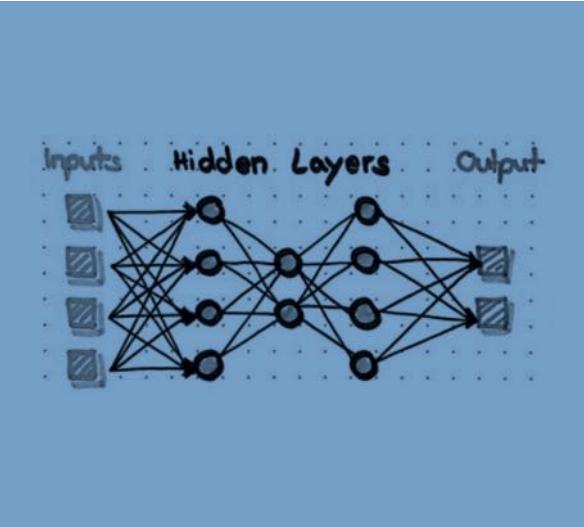
An Open-source DL Framework



A deep learning library



ONEIROS

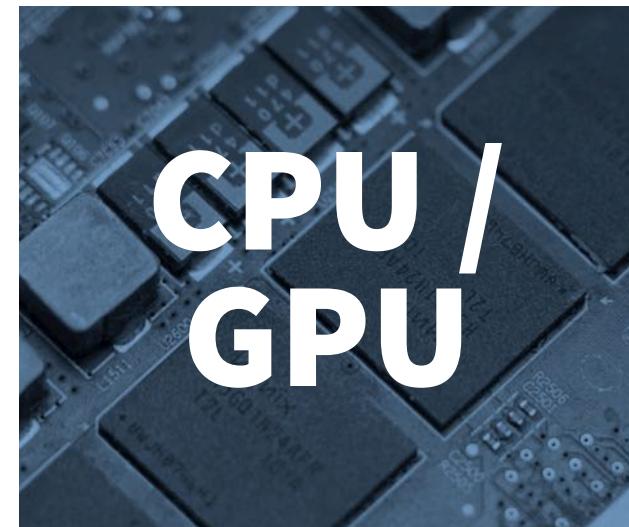


Layers



"Oneiroi are beyond our unravelling -
who can be sure what tale they tell?
Not all that men look for comes to pass.
Two gates there are that give passage to fleeting Oneiroi;
one is made of horn, one of ivory.
The Oneiroi that pass through sown ivory are deceitful,
bearing a message that will not be fulfilled;
those that come out through polished horn have truth behind them,
to be accomplished for men who see them."

Homer, Odyssey 19. 562 ff (Shewring translation).



CPU / GPU

Developed by François Chollet

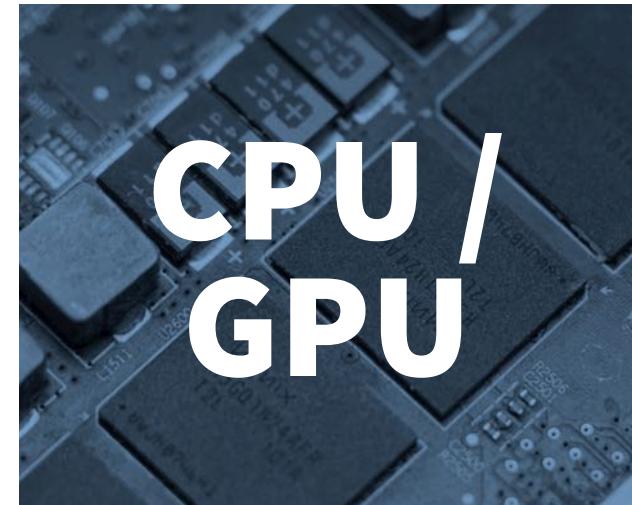
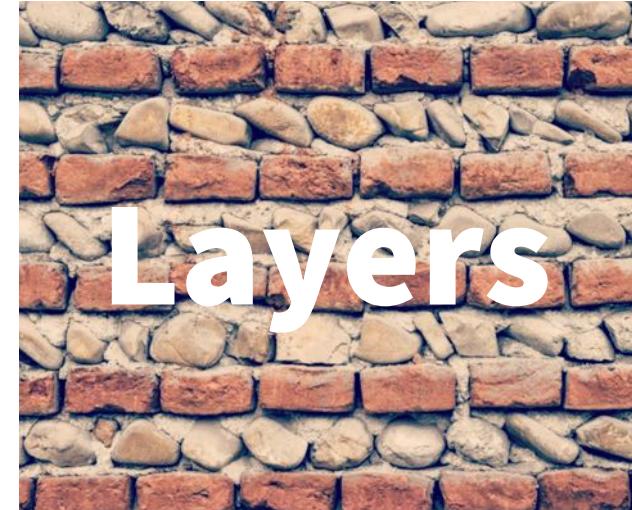
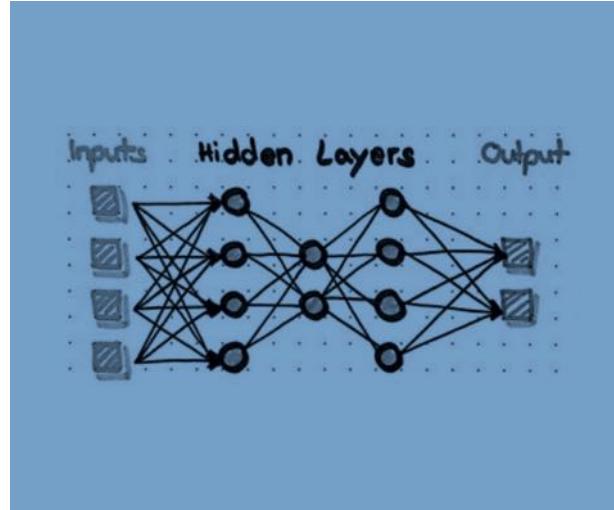
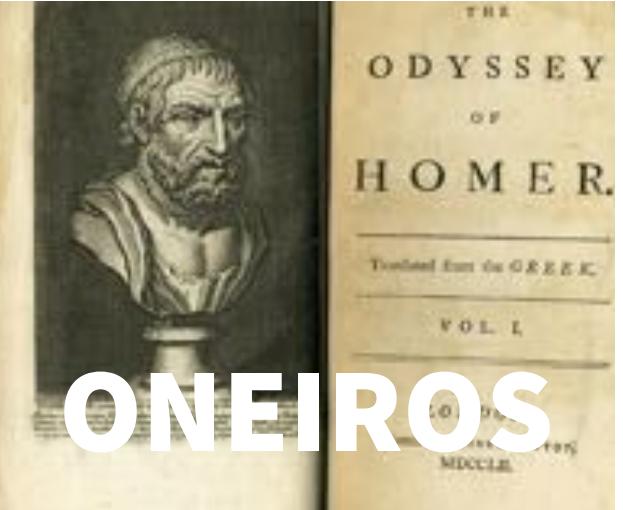
written in Python



A deep learning library

"Oneiroi are beyond our unravelling -
who can be sure what tale they tell?
Not all that men look for comes to pass.
Two gates there are that give passage to fleeting Oneiroi;
one is made of horn, one of ivory.
The Oneiroi that pass through sawn ivory are deceitful,
bearing a message that will not be fulfilled;
those that come out through polished horn have truth behind them,
to be accomplished for men who see them."

Homer, Odyssey 19. 562 ff (Shewring translation).



Check Out Keras Cheat Sheet

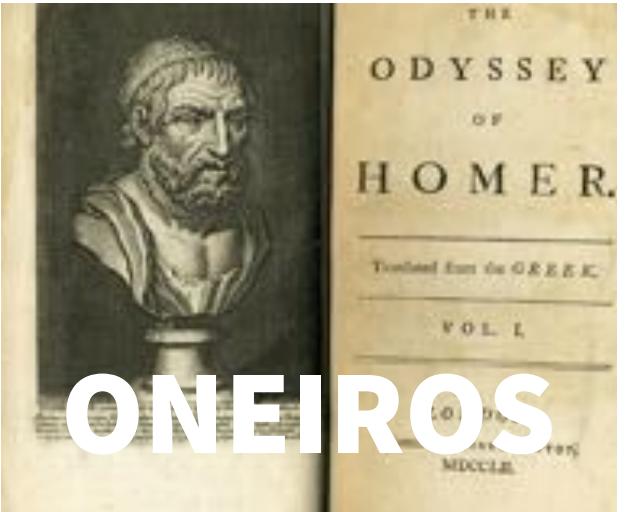


Originally from ONEIROS project

Open-Ended Neuro-Electronic Intelligent Robot OperatingSystem

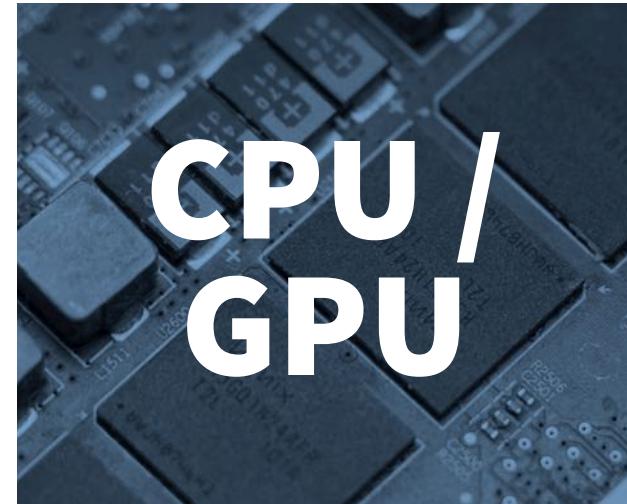
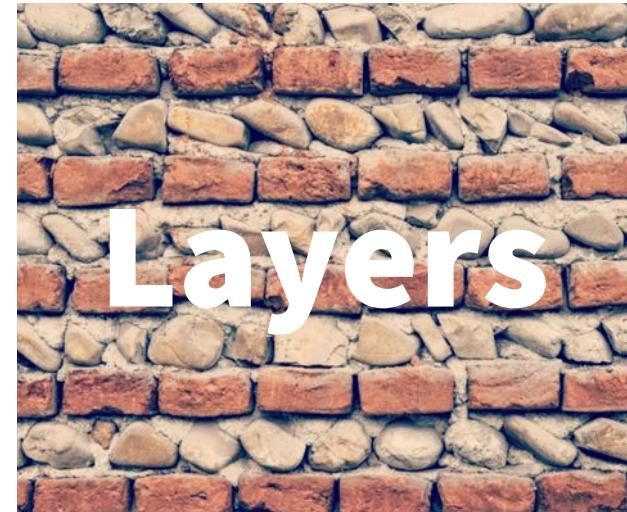
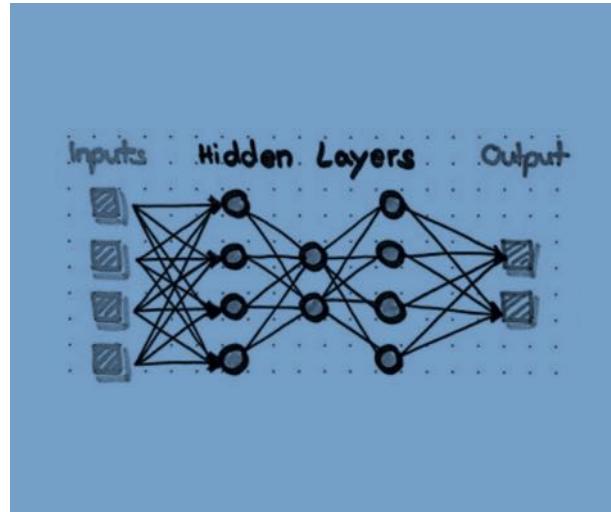
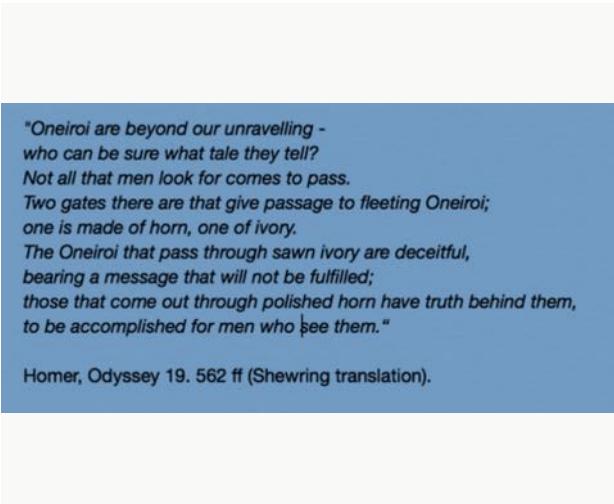


A deep learning library



"Oneiroi are beyond our unravelling -
who can be sure what tale they tell?
Not all that men look for comes to pass.
Two gates there are that give passage to fleeting Oneiroi;
one is made of horn, one of ivory.
The Oneiroi that pass through sawn ivory are deceitful,
bearing a message that will not be fulfilled;
those that come out through polished horn have truth behind them,
to be accomplished for men who see them."

Homer, Odyssey 19. 562 ff (Shewring translation).

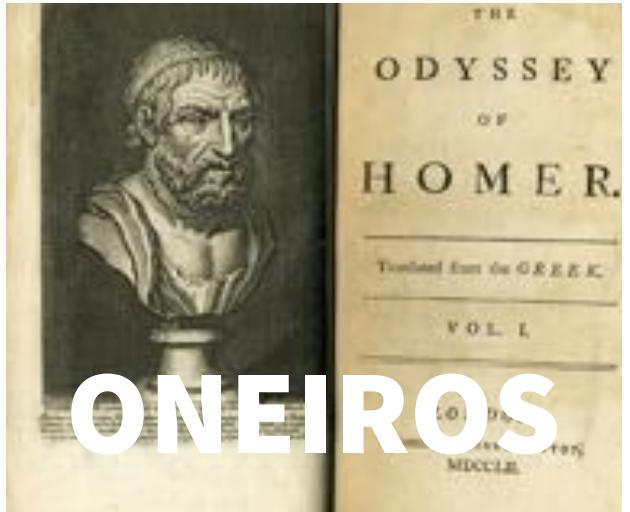


Originally from ONEIROS project

Keras means “horn” in Greek and comes from the Odyssey where it stands for true dreams

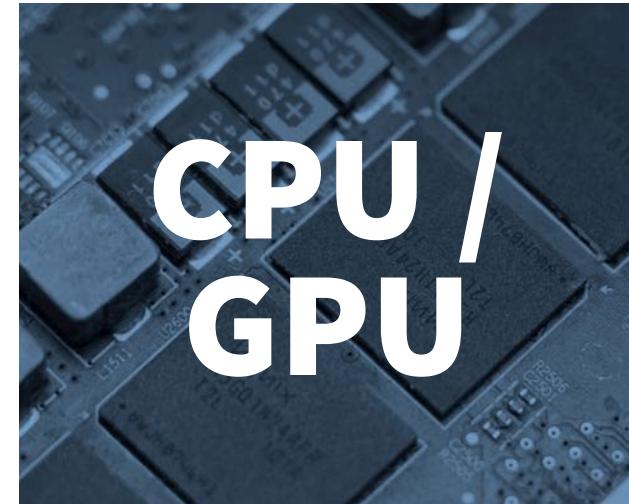
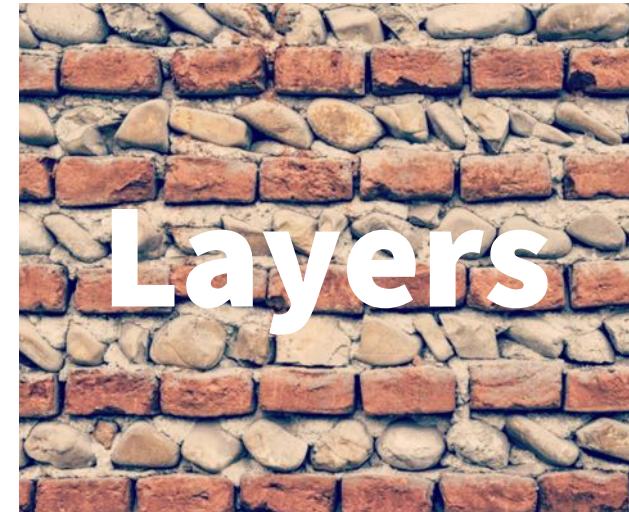
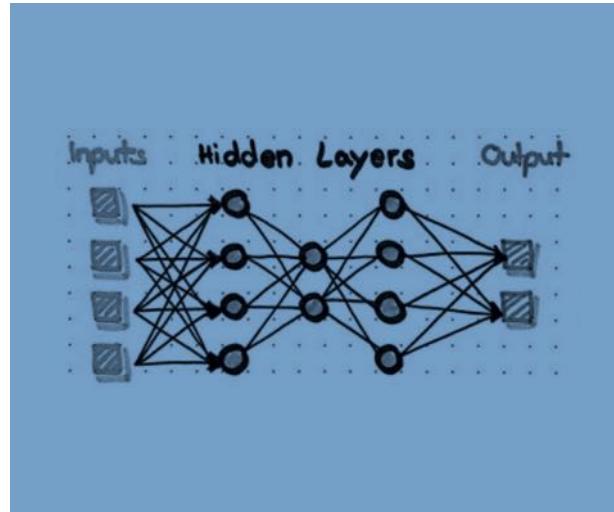
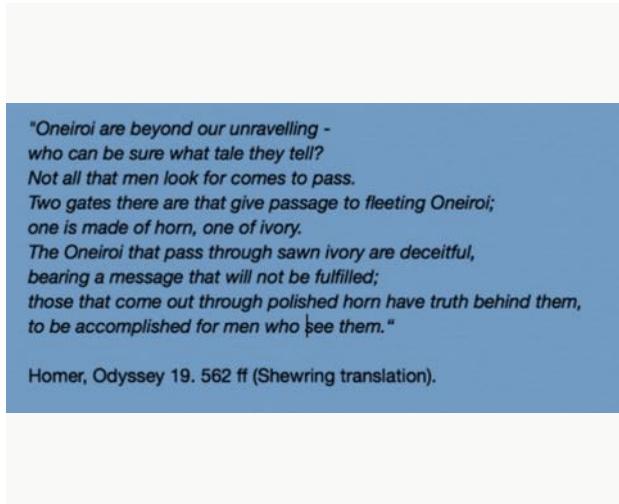


A deep learning library



*"Oneiroi are beyond our unravelling -
who can be sure what tale they tell?
Not all that men look for comes to pass.
Two gates there are that give passage to fleeting Oneiroi;
one is made of horn, one of ivory.
The Oneiroi that pass through sawn ivory are deceitful,
bearing a message that will not be fulfilled;
those that come out through polished horn have truth behind them,
to be accomplished for men who see them."*

Homer, Odyssey 19. 562 ff (Shewring translation).

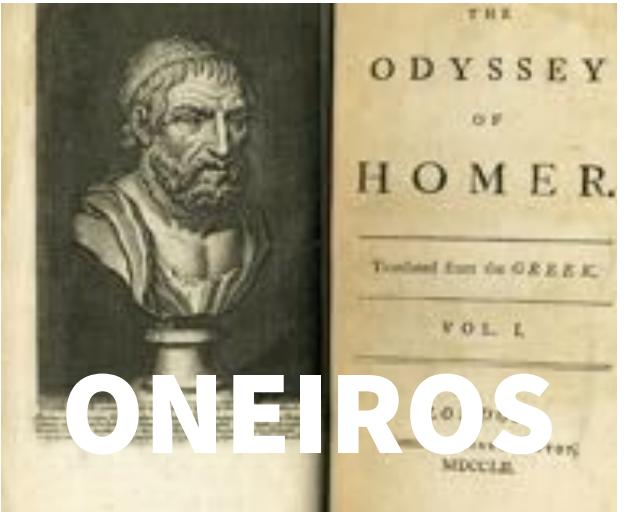


Lets you design neural nets

minimalistic, efficient, highly flexible, framework - independent



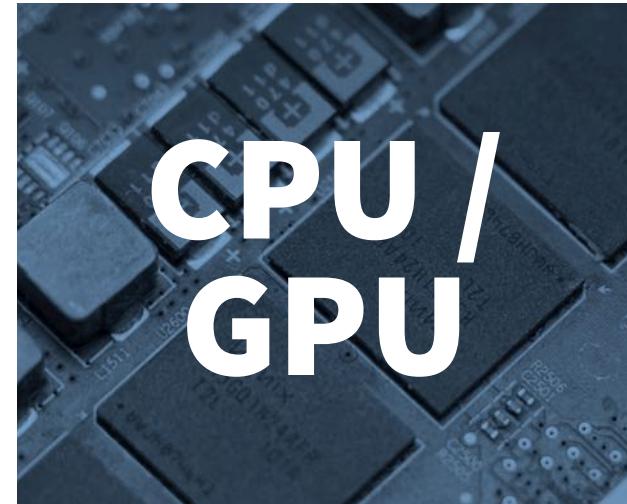
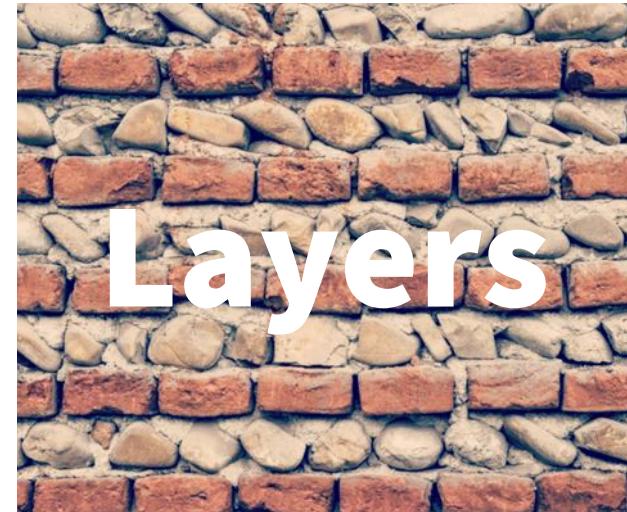
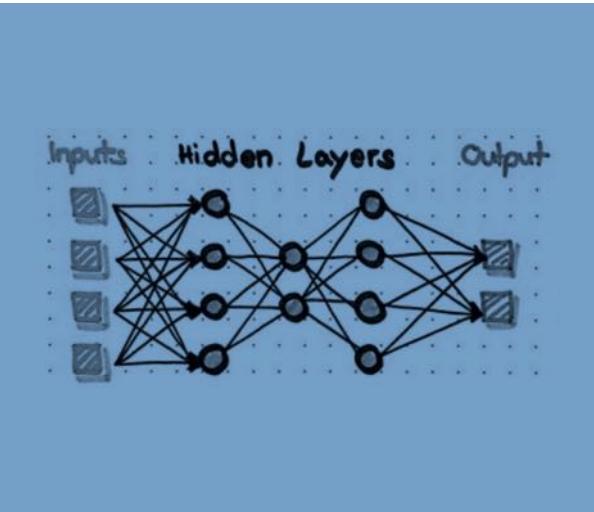
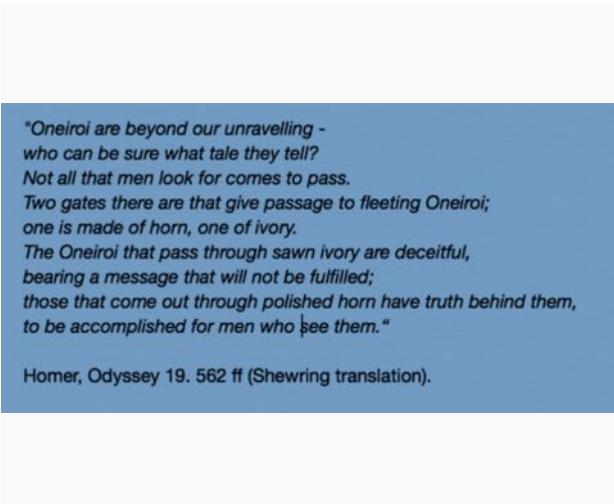
A deep learning library



ONEIROS

"Oneiroi are beyond our unravelling -
who can be sure what tale they tell?
Not all that men look for comes to pass.
Two gates there are that give passage to fleeting Oneiroi;
one is made of horn, one of ivory.
The Oneiroi that pass through sawn ivory are deceitful,
bearing a message that will not be fulfilled;
those that come out through polished horn have truth behind them,
to be accomplished for men who see them."

Homer, Odyssey 19. 562 ff (Shewring translation).

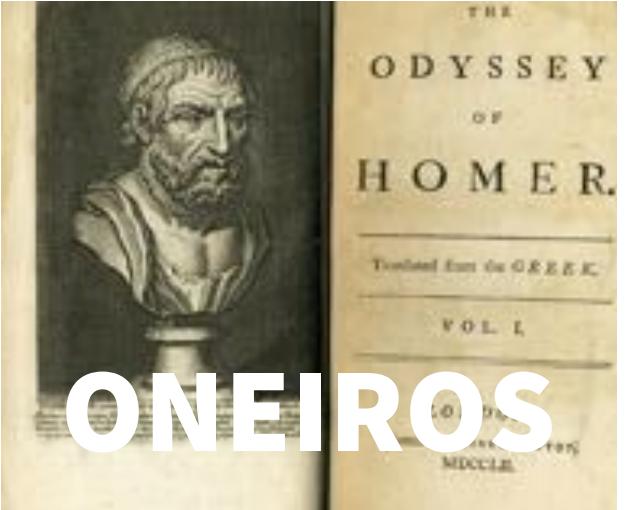


Lets you design neural nets

based on a layer system, like LEGO

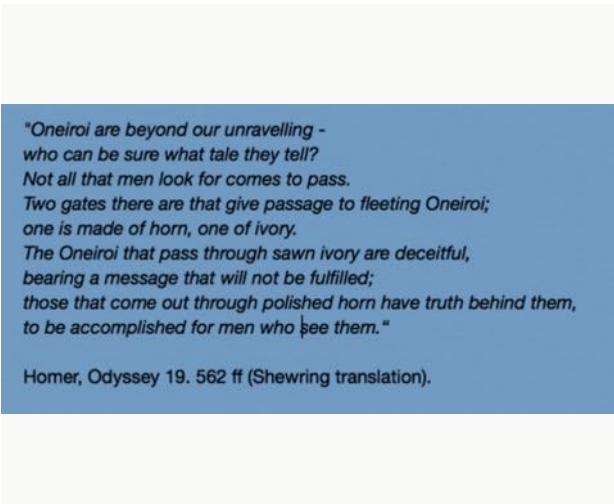
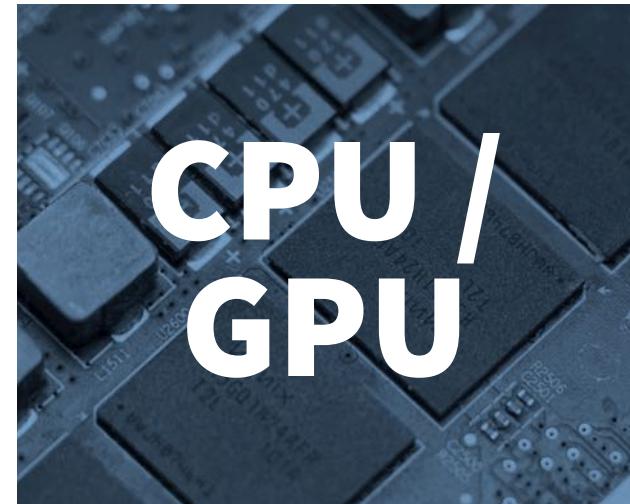
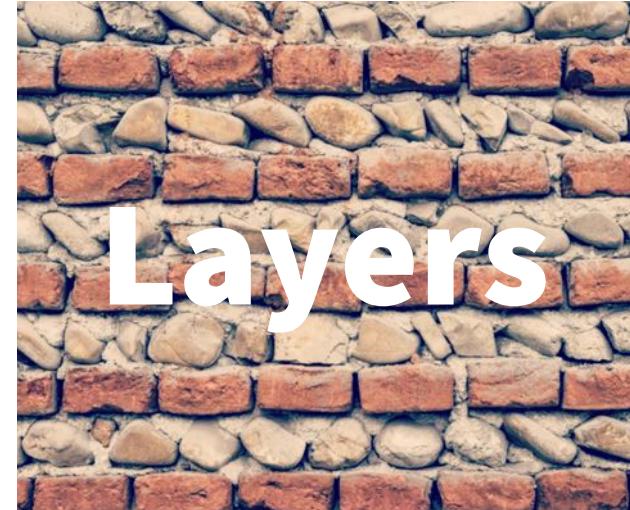
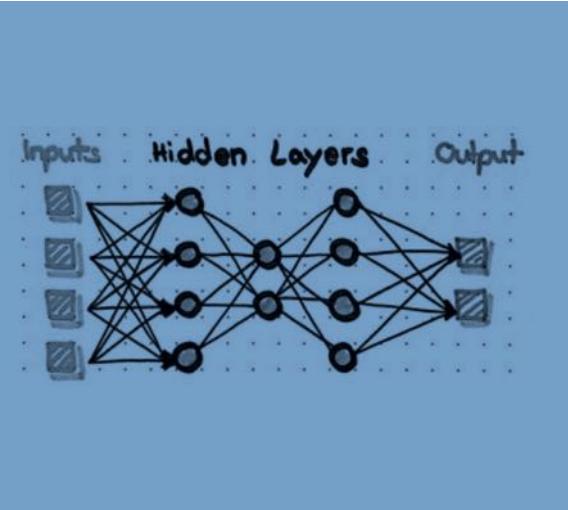


A deep learning library

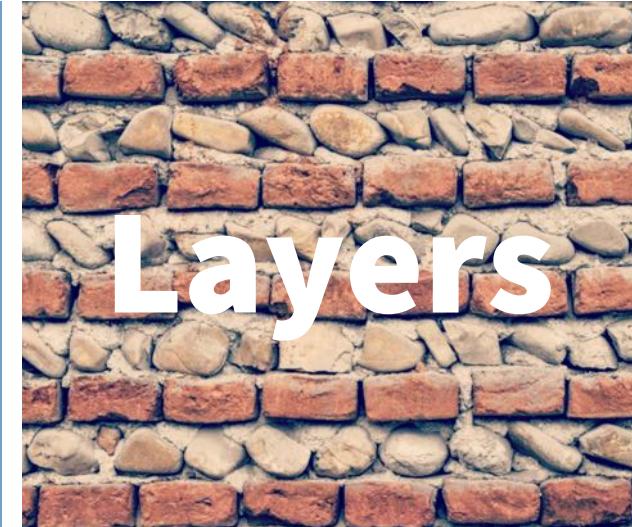
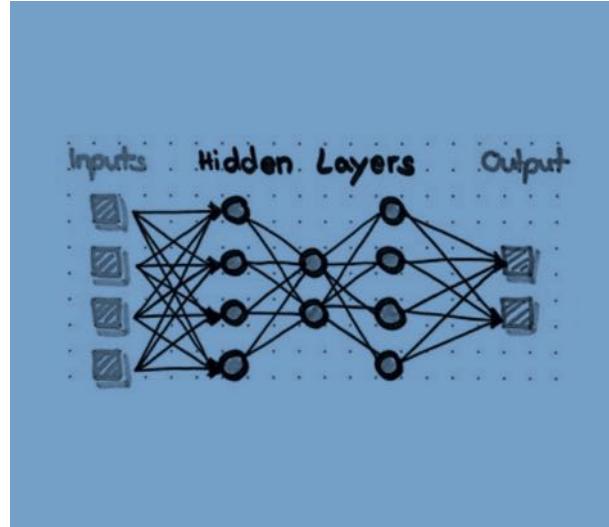
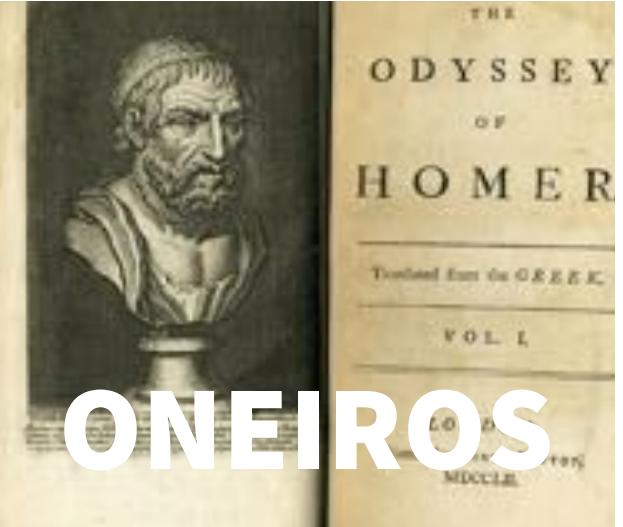
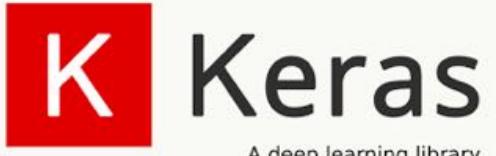


"Oneiroi are beyond our unravelling -
who can be sure what tale they tell?
Not all that men look for comes to pass.
Two gates there are that give passage to fleeting Oneiroi;
one is made of horn, one of ivory.
The Oneiroi that pass through sawn ivory are deceitful,
bearing a message that will not be fulfilled;
those that come out through polished horn have truth behind them,
to be accomplished for men who see them."

Homer, *Odyssey* 19. 562 ff (Shewring translation).

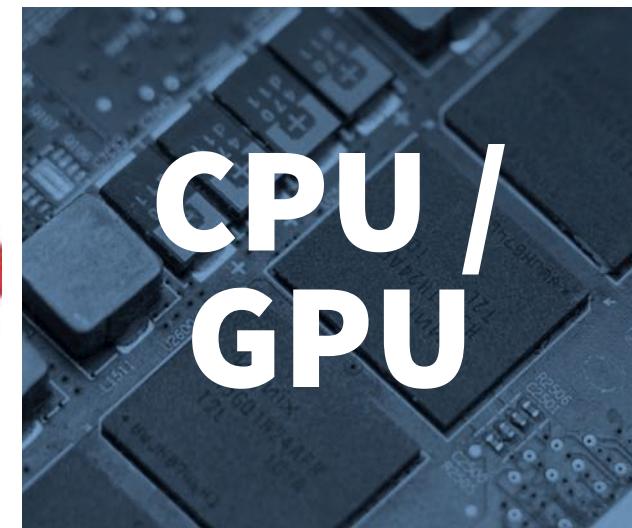


Benefit from the community

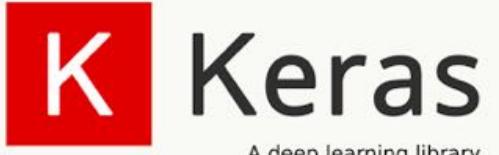


"Oneiroi are beyond our unravelling -
who can be sure what tale they tell?
Not all that men look for comes to pass.
Two gates there are that give passage to fleeting Oneiroi;
one is made of horn, one of ivory.
The Oneiroi that pass through sown ivory are deceitful,
bearing a message that will not be fulfilled;
those that come out through polished horn have truth behind them,
to be accomplished for men who see them."

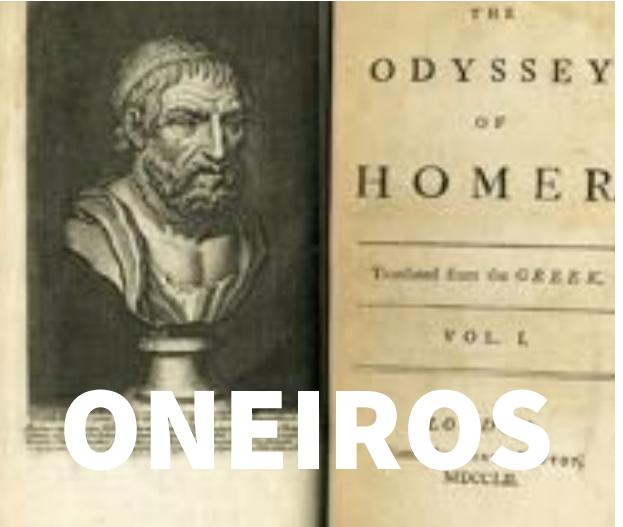
Homer, Odyssey 19. 562 ff (Shewring translation).



Models can train on CPU & GPU



A deep learning library

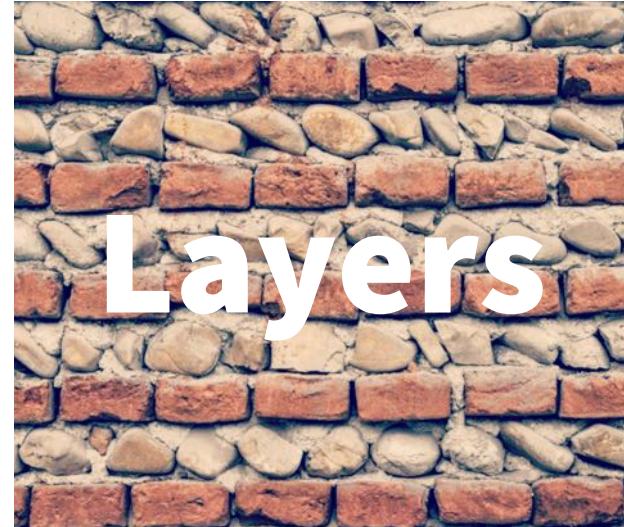
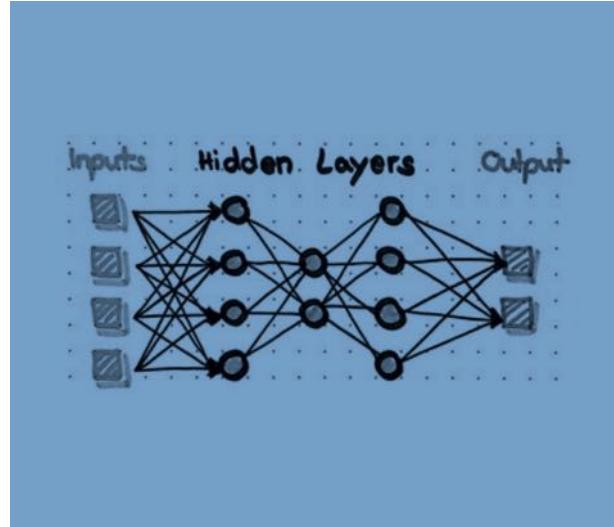


ONEIROS

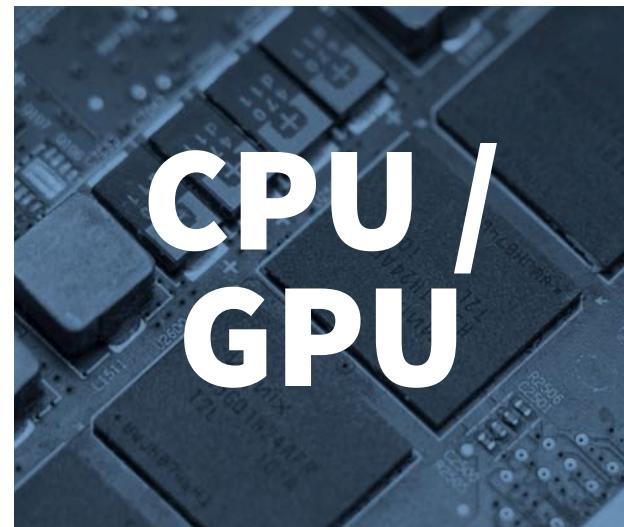


"Oneiroi are beyond our unravelling -
who can be sure what tale they tell?
Not all that men look for comes to pass.
Two gates there are that give passage to fleeting Oneiroi;
one is made of horn, one of ivory.
The Oneiroi that pass through sown ivory are deceitful,
bearing a message that will not be fulfilled;
those that come out through polished horn have truth behind them,
to be accomplished for men who see them."

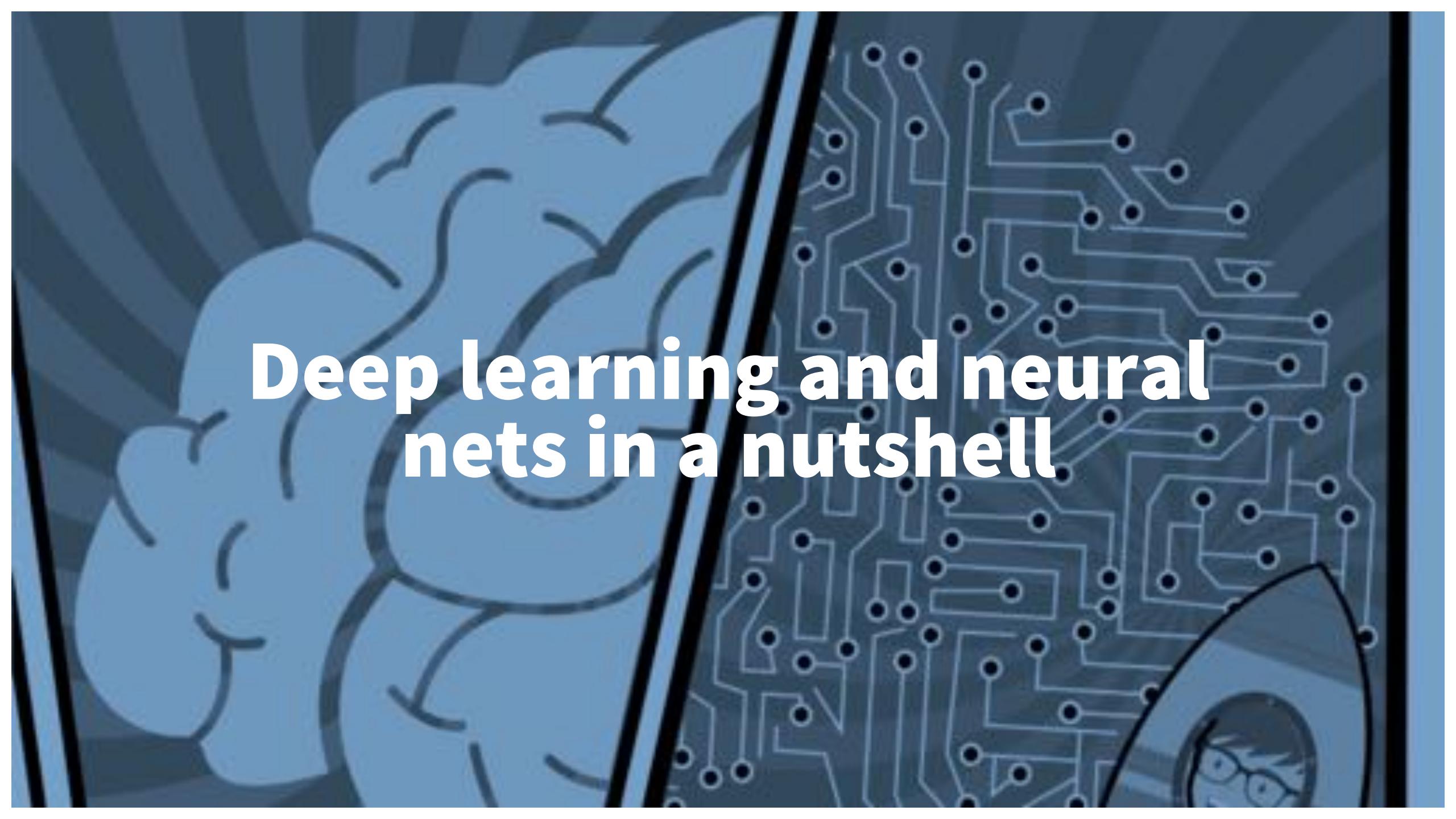
Homer, Odyssey 19. 562 ff (Shewring translation).



Layers

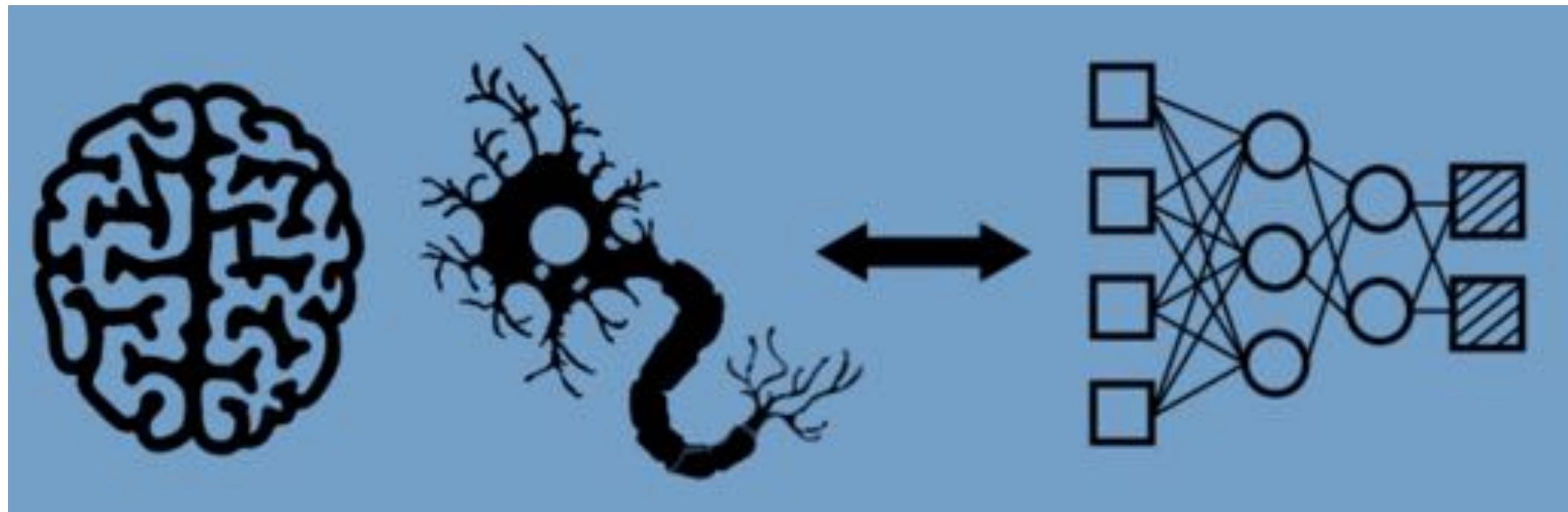


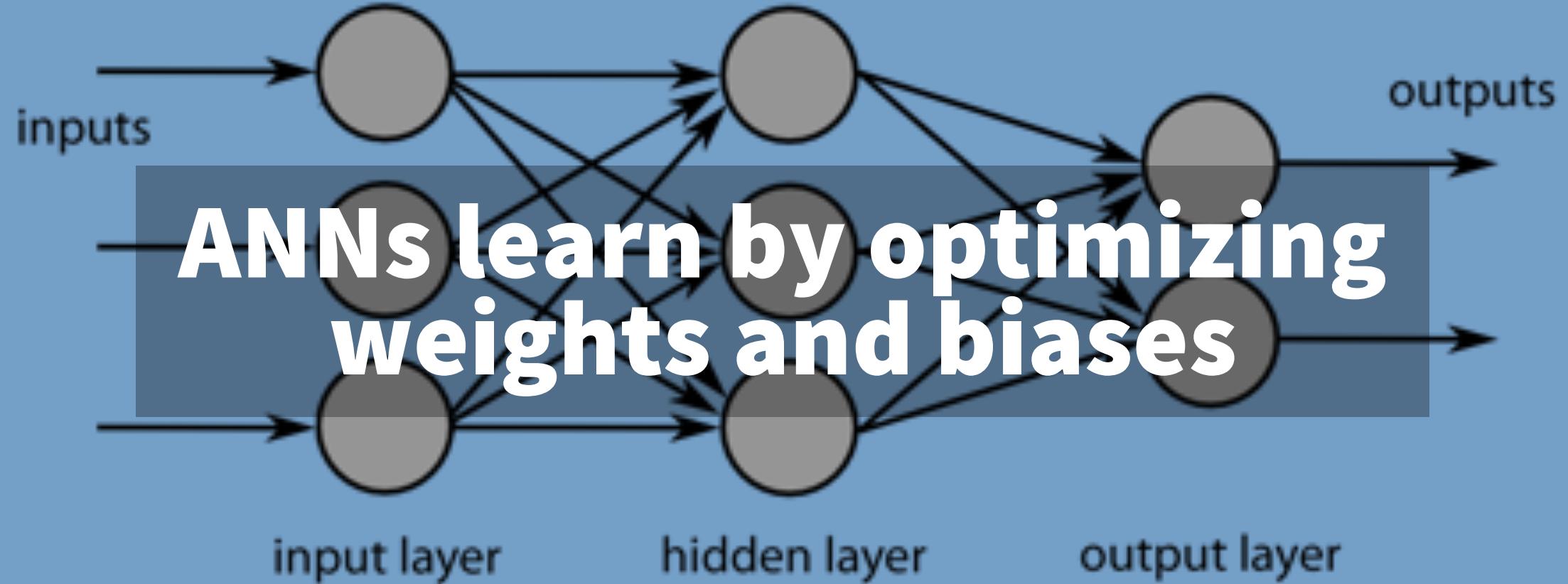
CPU / GPU

The background of the slide features a stylized, blue-tinted illustration of a human brain on the left, showing its gyri and sulci. On the right, there is a detailed illustration of a computer circuit board with various silver-colored metal tracks and black circular pads. The overall theme is the intersection of biology and technology.

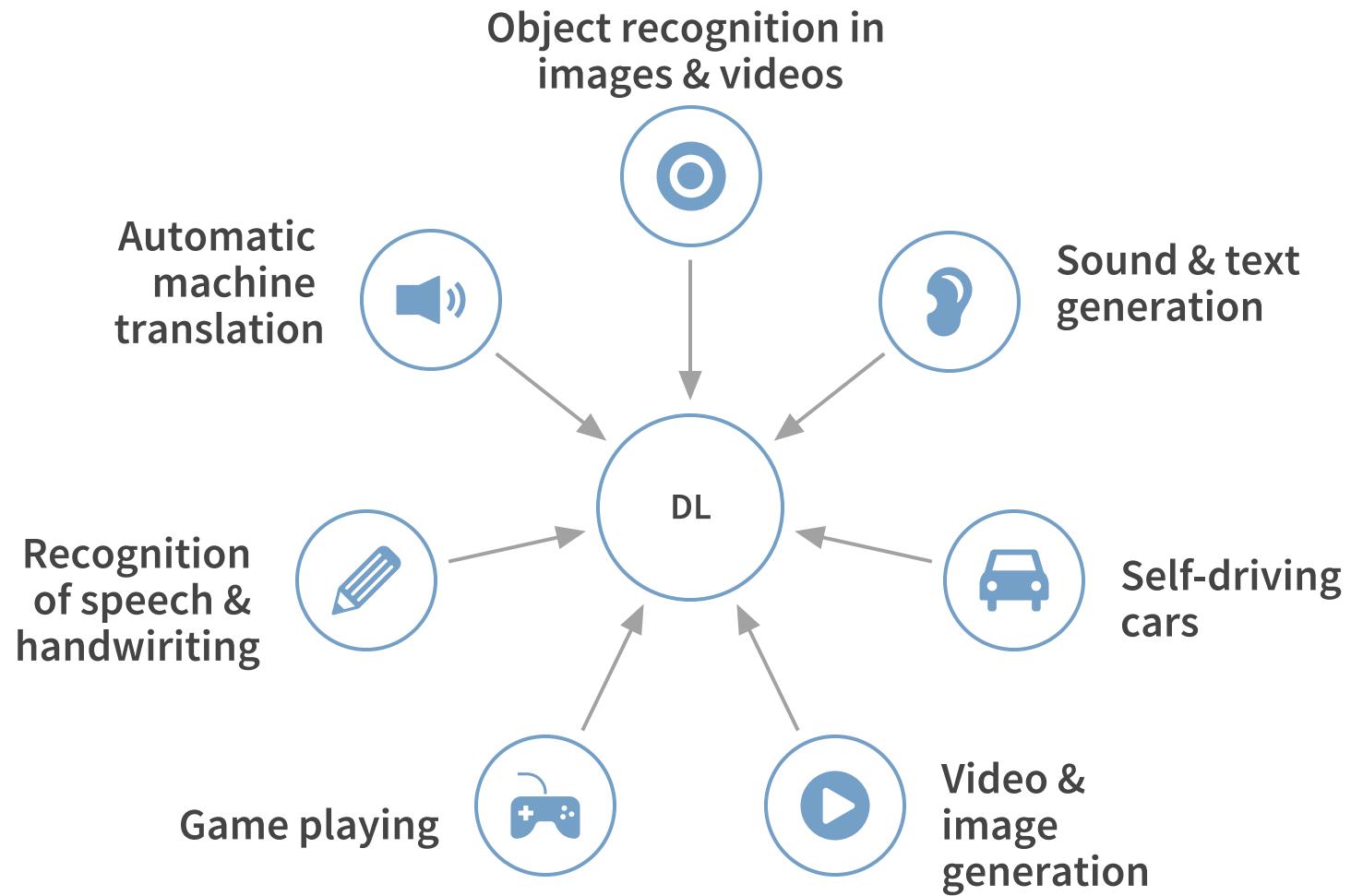
Deep learning and neural nets in a nutshell

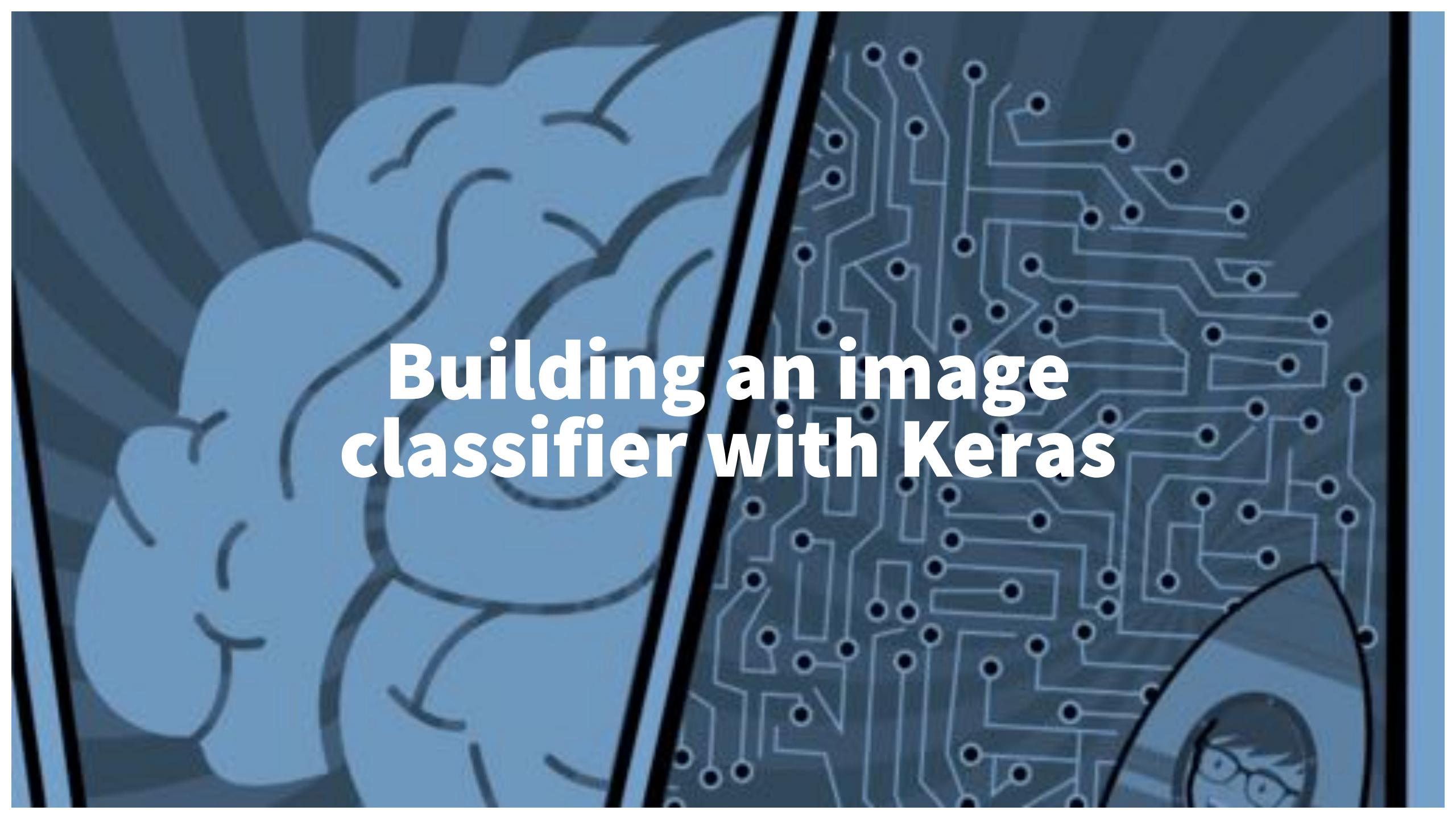
Neural nets are modeled after the human brain.





Deep learning solves complex tasks.





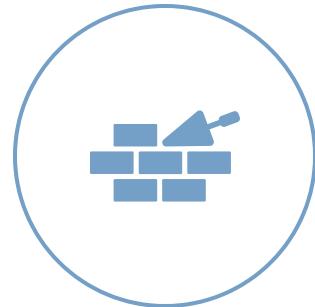
Building an image classifier with Keras



How does
Keras work?

Build neural
nets like LEGO
systems

Two types of APIs exist for model building



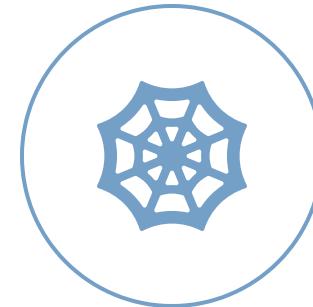
Sequential models

simple

suitable for most cases

linear order of layers

only one direction from input to output



Functional API

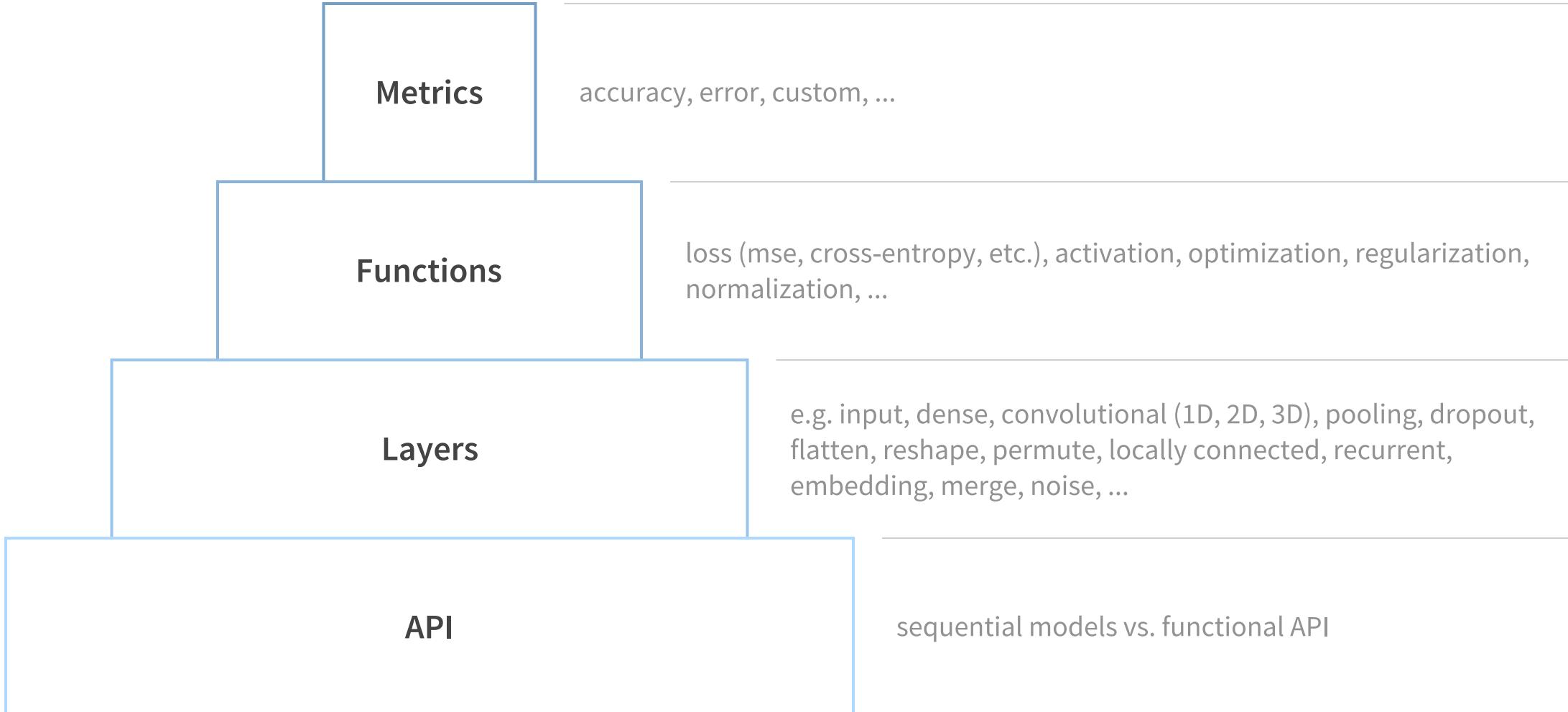
more complex

suitable for complex models

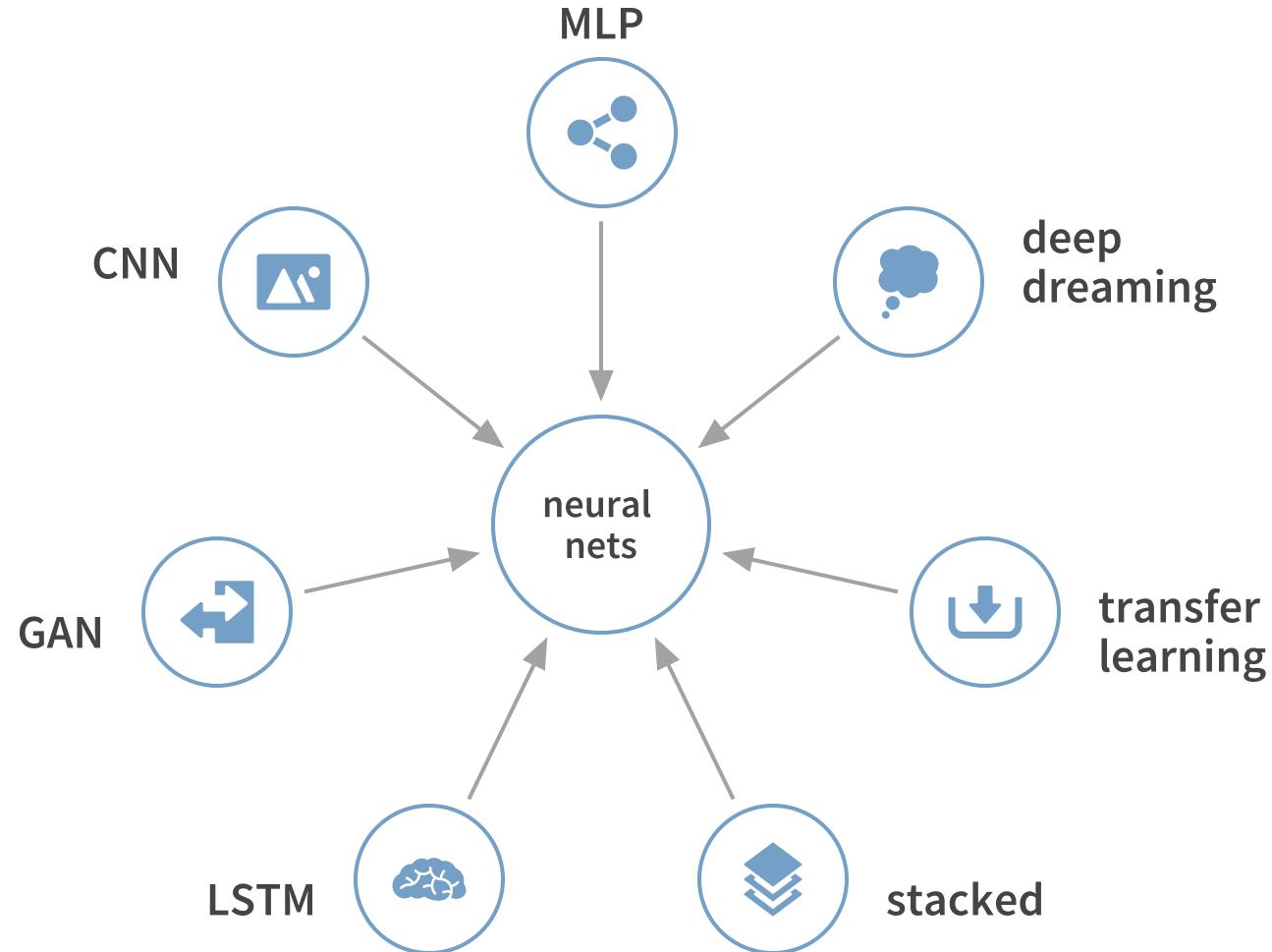
can have multiple in- or outputs

layers can be non-sequential, e.g. LSTM

Modularity of layers



Keras lets you build all kinds of neural nets



There are many options for working with Keras



Train models

- Jupyter notebook
- Python
- R

There are many options for working with Keras



Backend

- TensorFlow
- Caffe
- Theano
- ...

There are many options for working with Keras



Deploy

- TensorFlow object & serving
- JavaScript
- Scala
- ...

There are many options for working with Keras



Scale

- distributed learning with Spark and elephas
- ...

There are many options for working with Keras



Cloud

- AWS
- Azure

Let's train an image classifier

Code: https://shirinsplayground.netlify.com/2018/06/keras_fruits/

Data: <https://www.kaggle.com/moltean/fruits/data>



- **Classify fruits**

Kiwi, Banana, Apricot, Avocado, Cocos, Clementine, Mandarine, Orange, Limes, Lemon, Peach, Plum, Raspberry, Strawberry, Pineapple

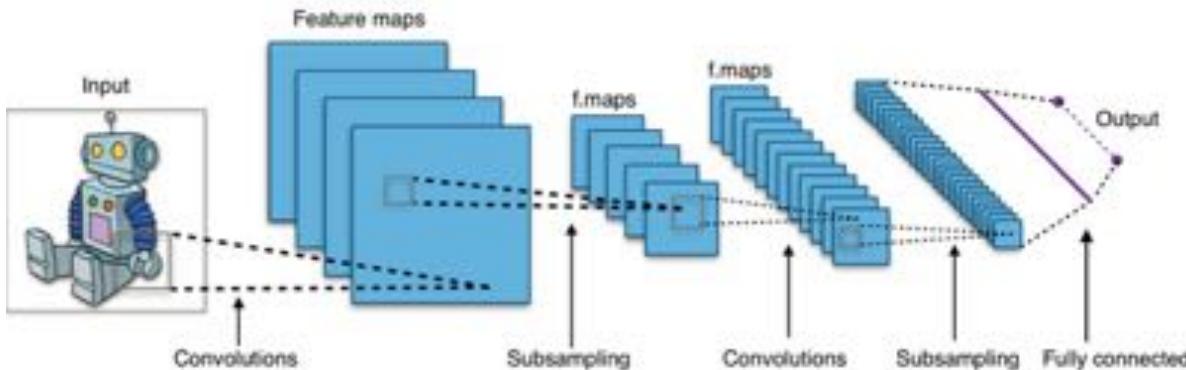
- **ca 490 images per class**

3 channels

100 x 100 px reduced to 20 x 20 px
scaled (divided by 255)

Let's train an image classifier

with a Convolutional Neural Net



- **Convolutional Layer**

1 - many

2D or 3D

sliding window calculates convolutions => filter kernels
edge detection => ... => complex features

- **Pooling Layer**

follows a (set of) convolutional layers

condensing information, e.g. max pooling

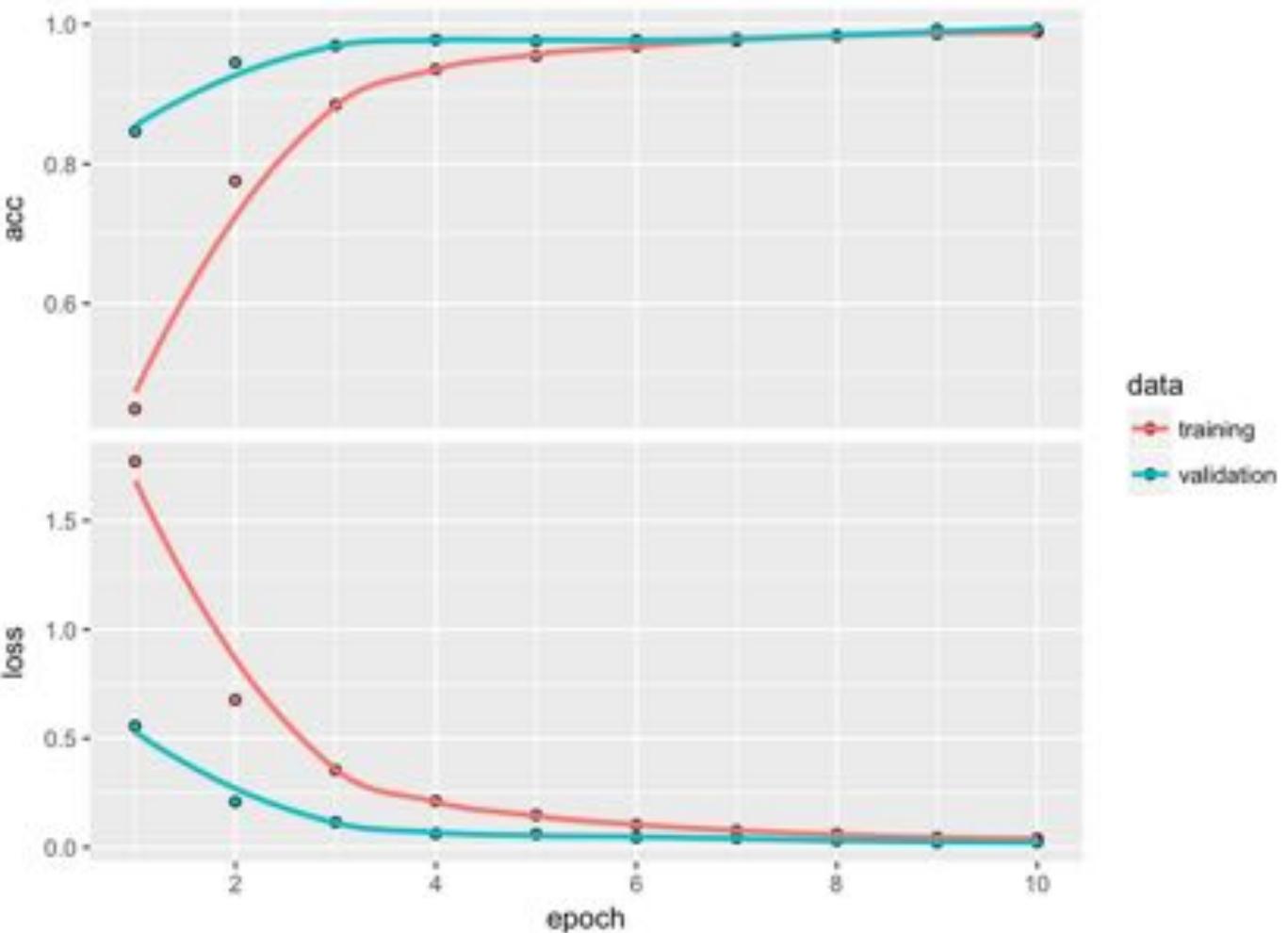
```

# initialise model
model <- keras_model_sequential()

# add layers
model %>%
  layer_conv_2d(filter = 32, kernel_size = c(3,3), padding = "same", input_shape = c(img_w,
layer_activation("relu")) %>%
  
  # Second hidden layer
  layer_conv_2d(filter = 16, kernel_size = c(3,3), padding = "same") %>%
  layer_activation_leaky_relu(0.5) %>%
  layer_batch_normalization() %>%
  
  # Use max pooling
  layer_max_pooling_2d(pool_size = c(2,2)) %>%
  layer_dropout(0.25) %>%
  
  # Flatten max filtered output into feature vector
  # and feed into dense layer
  layer_flatten() %>%
  layer_dense(100) %>%
  layer_activation("relu") %>%
  layer_dropout(0.5) %>%
  
  # Outputs from dense layer are projected onto output layer
  layer_dense(output_n) %>%
  layer_activation("softmax")

# compile
model %>% compile(
  loss = "categorical_crossentropy",
  optimizer = optimizer_rmsprop(lr = 0.0001, decay = 1e-6),
  metrics = "accuracy"
)

```



TensorBoard

SCALARS

GRAPHS

File to screen

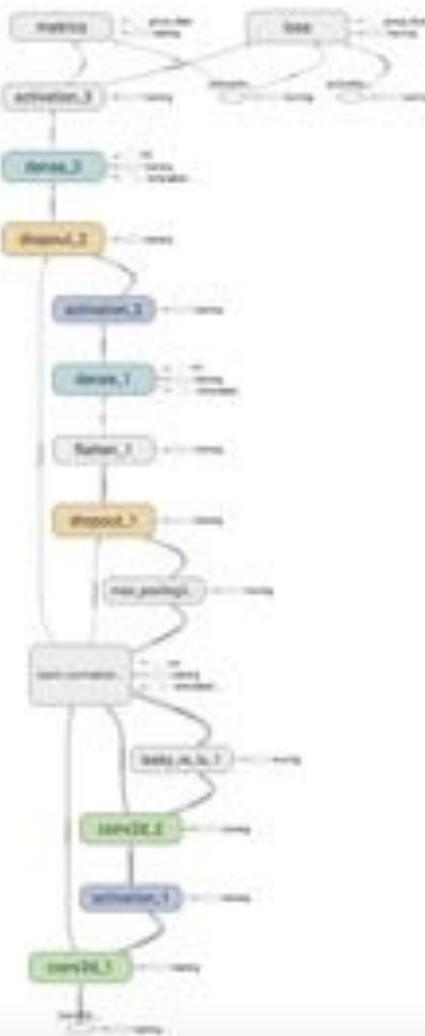
Download PNG

Run
logs/test.1
(1)Session
runs (0)Upload Trace inputsColor Structure Device XLA Cluster Compute time Memory TPU Compatibility

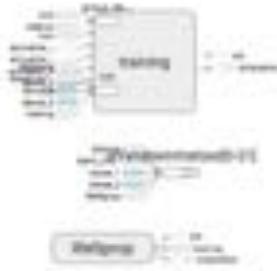
series same substructure

unique substructure

Main Graph



Auxiliary Nodes

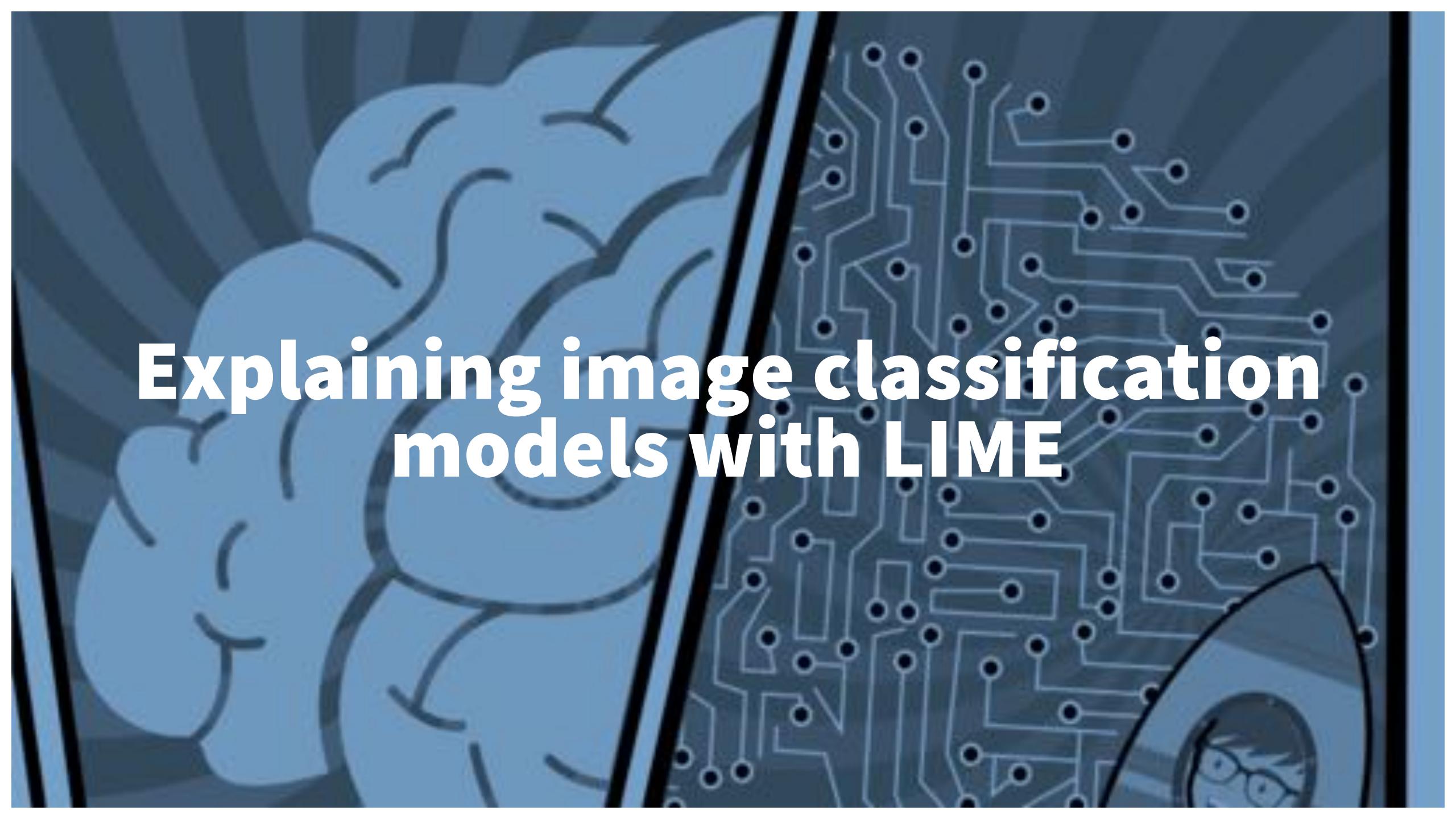
 Close legend.Graph Nonquantized OpsNode Unconnected series Connected series Control Summary Dataflow graph Control dependency edge Reference edge



But the question remains:

Why?

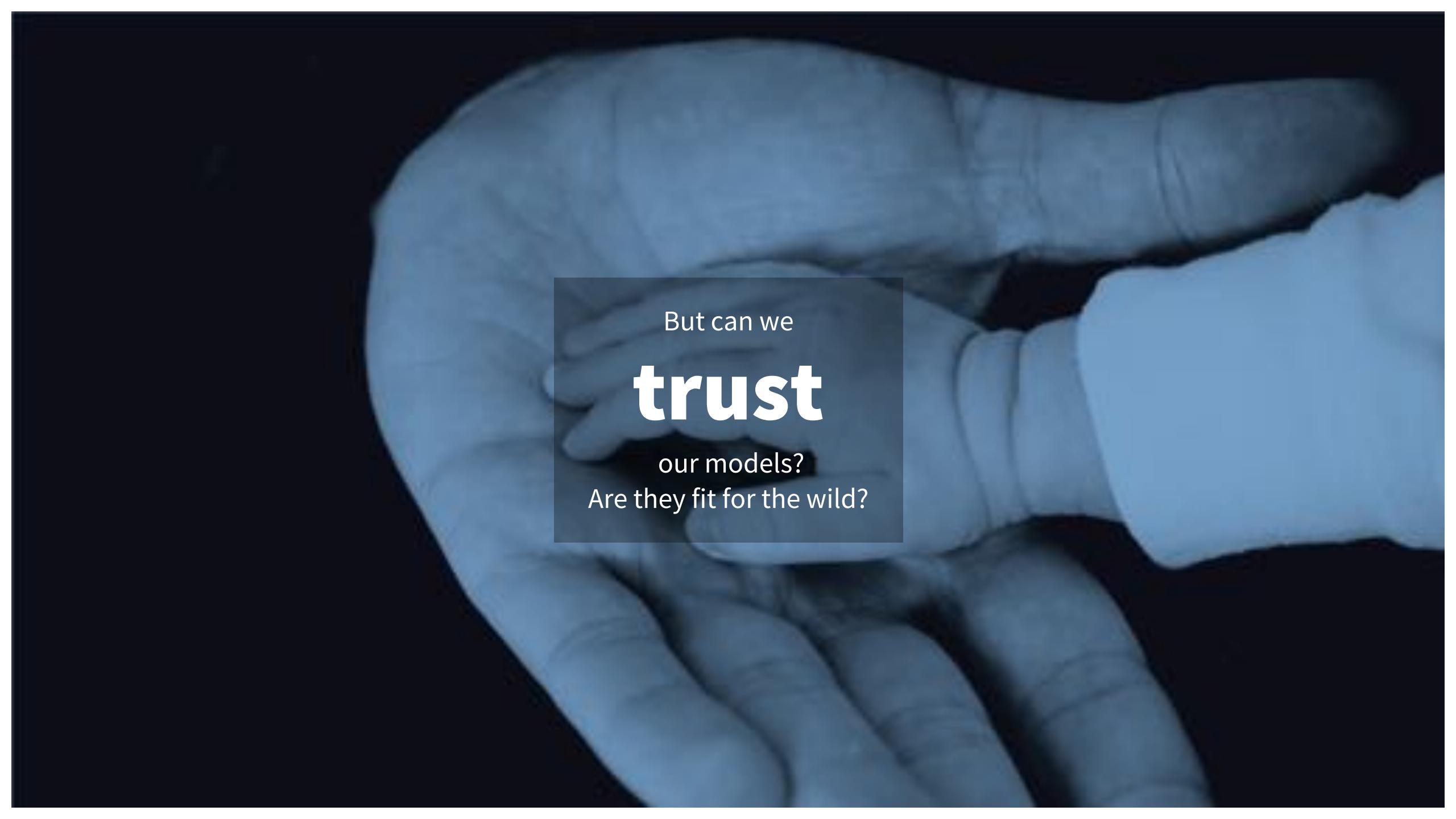
Why does my model make the predictions it makes?



Explaining image classification models with LIME

A typical ML workflow

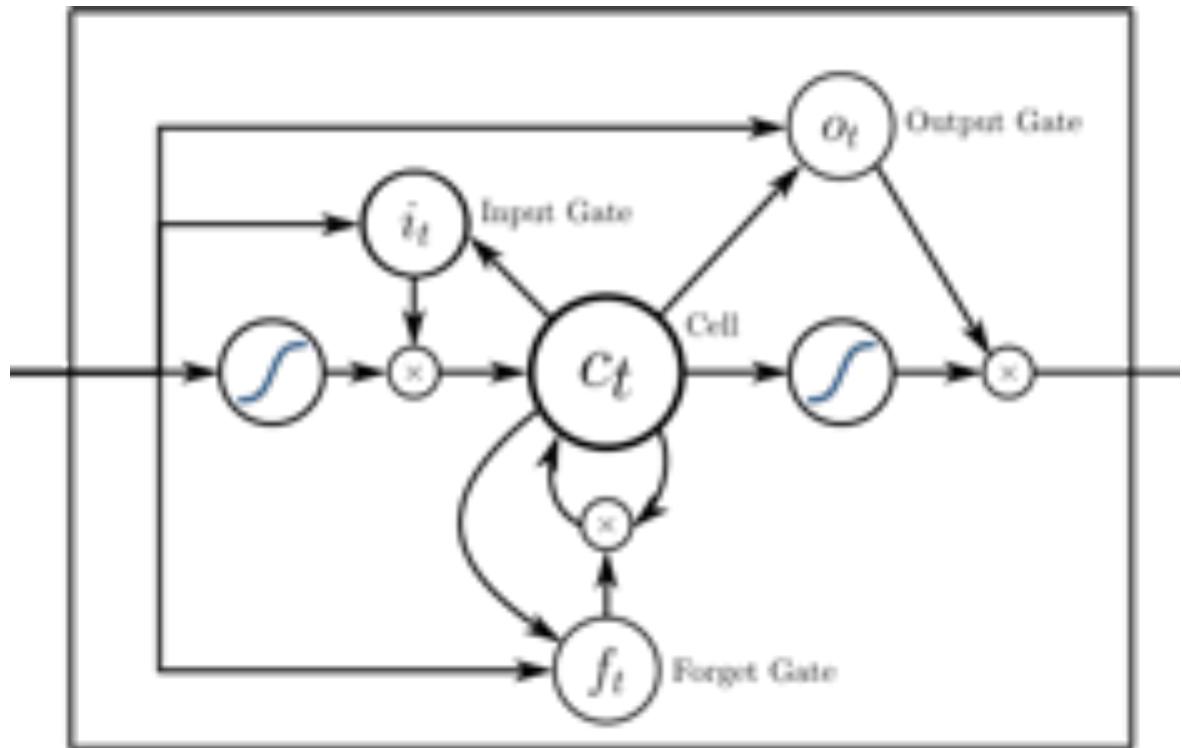
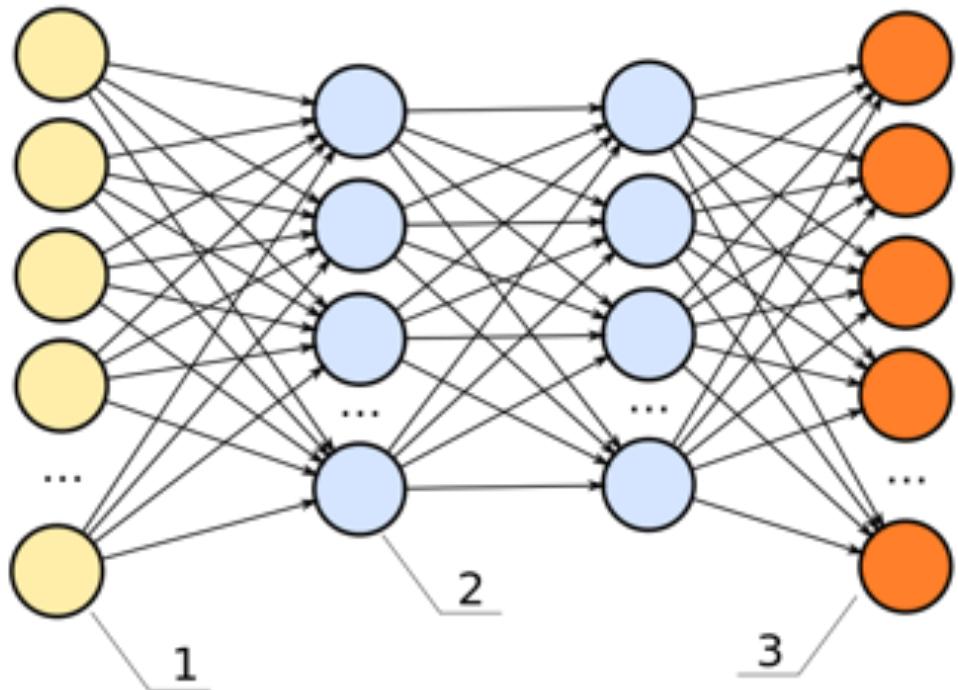




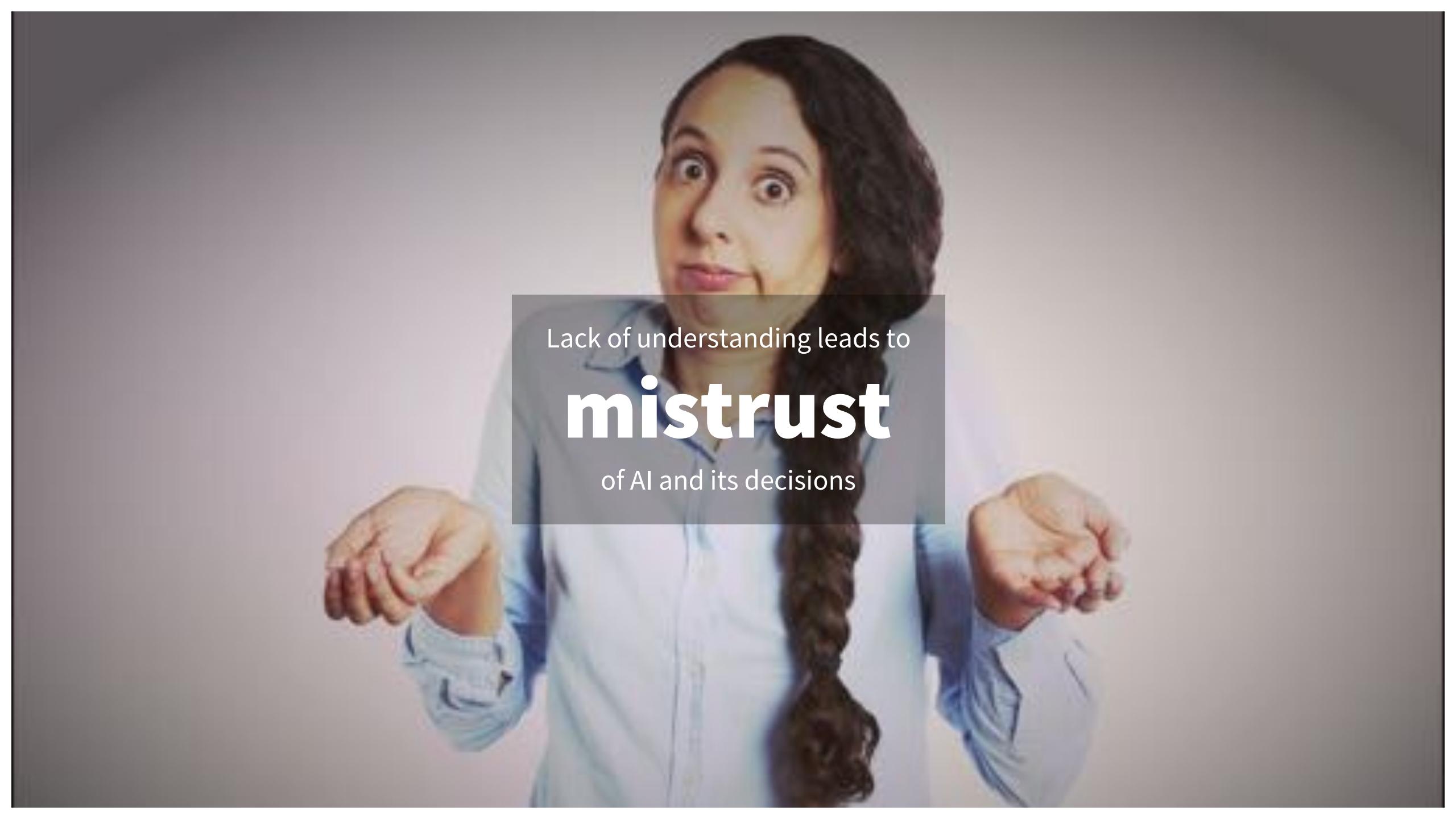
But can we
trust
our models?
Are they fit for the wild?

Trade-off between interpretability & complexity

Complexity tends to make models perform better but harder to understand.



Source: Wikipedia

A photograph of a woman with dark hair, wearing a light blue button-down shirt. She has a surprised or shocked expression, with her mouth slightly open and eyes wide. She is holding two closed fists up towards the camera at shoulder height. A semi-transparent dark gray rectangular box is overlaid on the center of the image, containing white text.

Lack of understanding leads to
mistrust
of AI and its decisions

General Data Protection Regulation (GDPR)



- “[...] the controller shall provide [...] the following information: the existence of automated decision-making and [...] meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.”

Art. 13-15 & 22 Regulation (EU) 2016/679

<https://dsgvo-gesetz.de/>

Why should we improve our understanding of ML models?

... if technically it isn't necessary ...



Improving our models

Generalisability

“Sanity Check”

Prevent wrong conclusions &
potentially adversarial attacks



Trust and transparency

Can I trust my model's decisions?

Why does my model make the predictions
it makes?

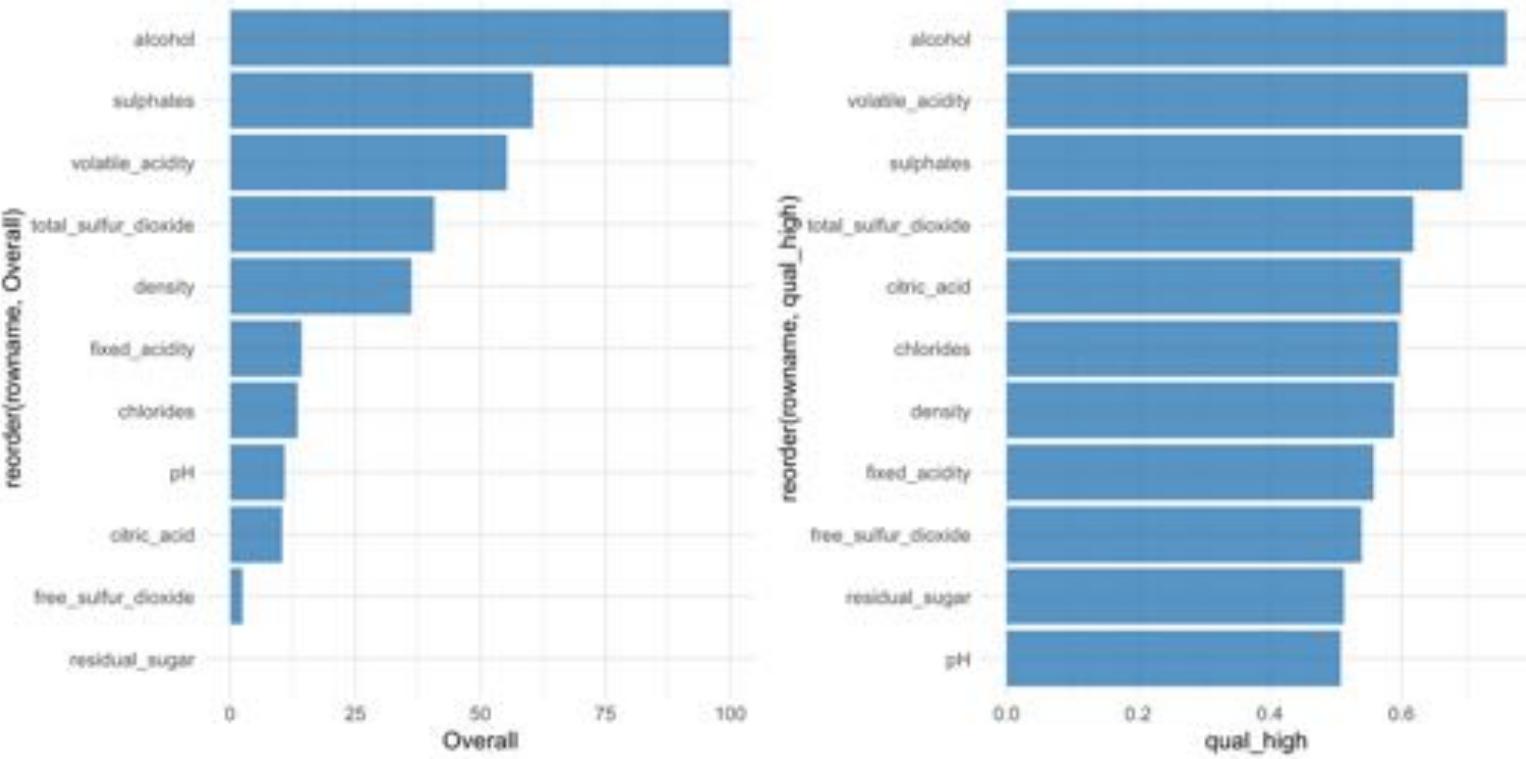


Prevent Bias

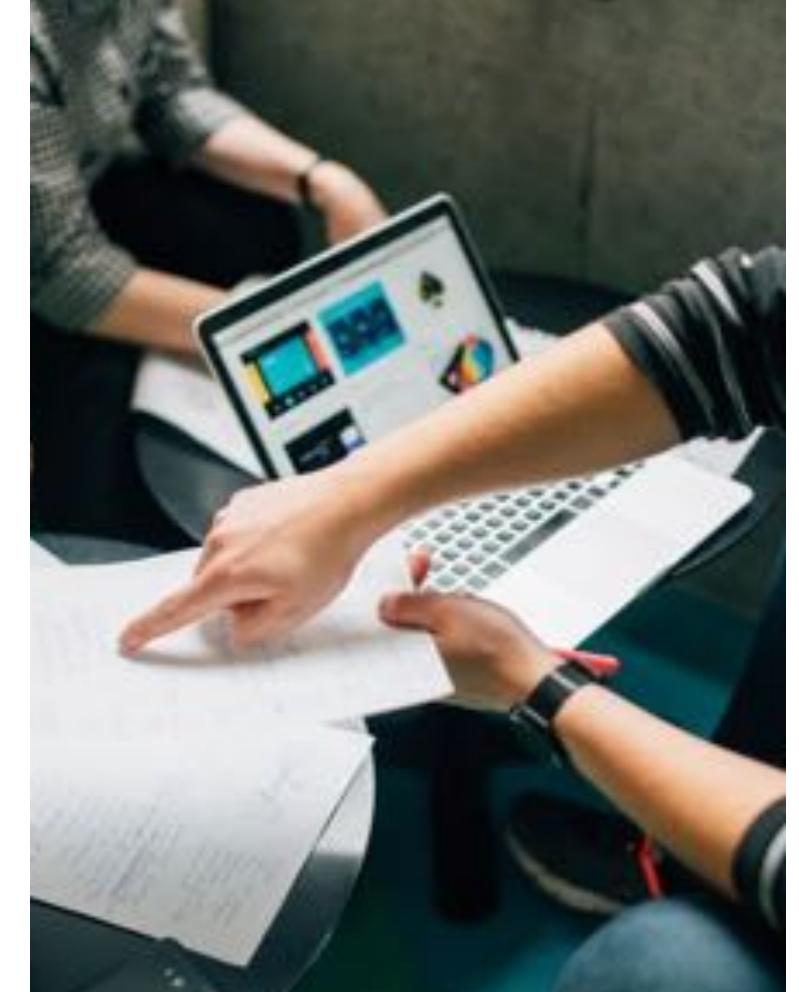
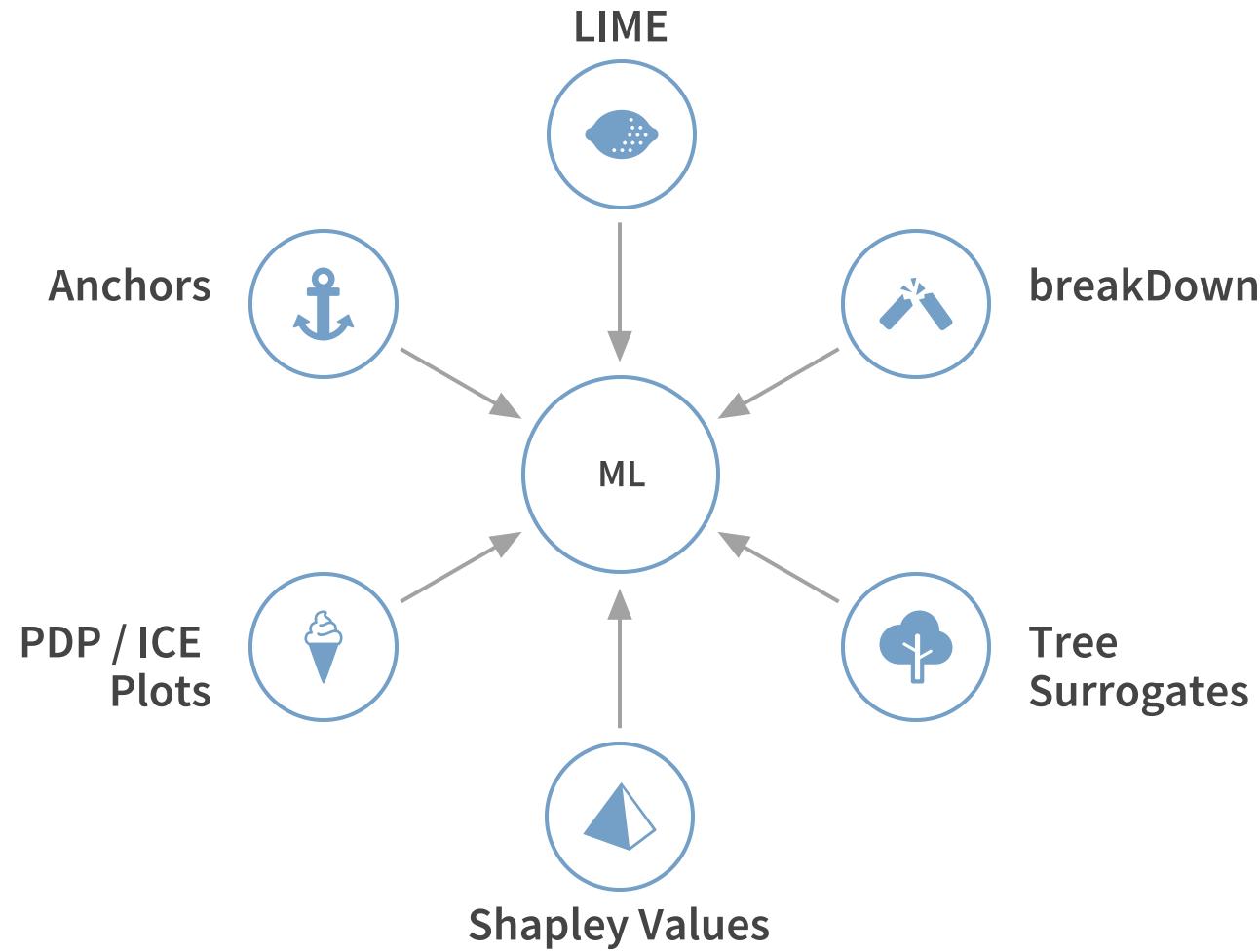
Fairness

Identify and prevent bias

Interpretability is more than feature importance!



Approaches to help explain and interpret ML models



Local Models: LIME

Local Interpretable Model-agnostic Explanations



- “Why Should I Trust You?”: Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin.

CoRR 2016

- Written in Python

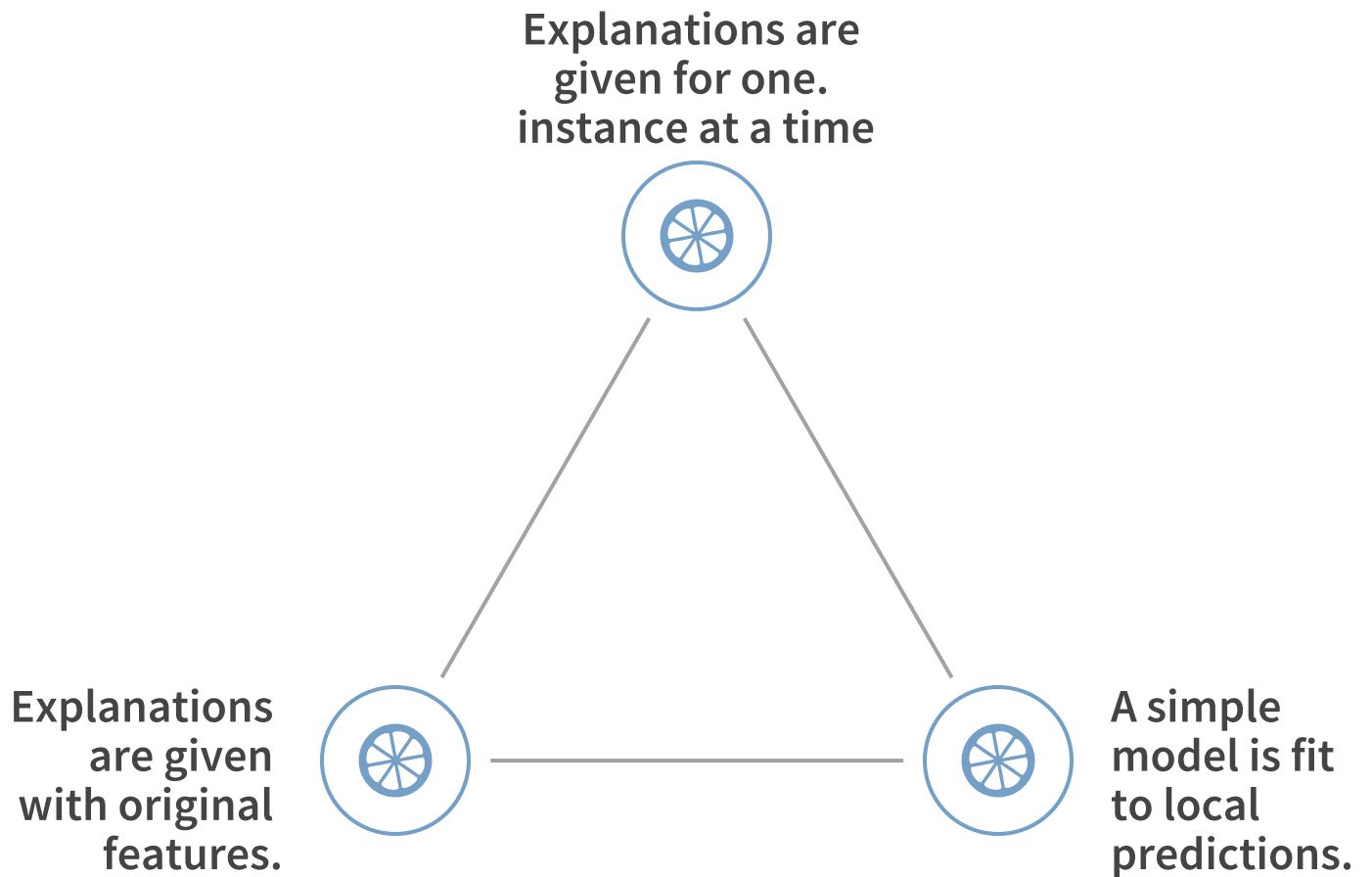
<https://github.com/marcotcr/lime>

- Adapted for R

<https://github.com/thomasp85/lime>

How does LIME work?

3 principles





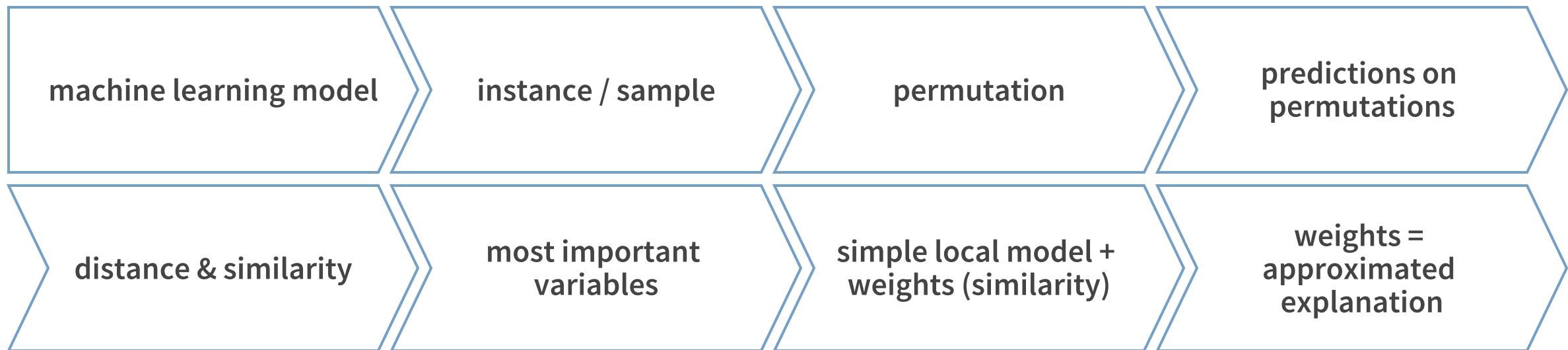
"Why should I trust you?" Explaining the predictions of any classifier

Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin
University of Washington

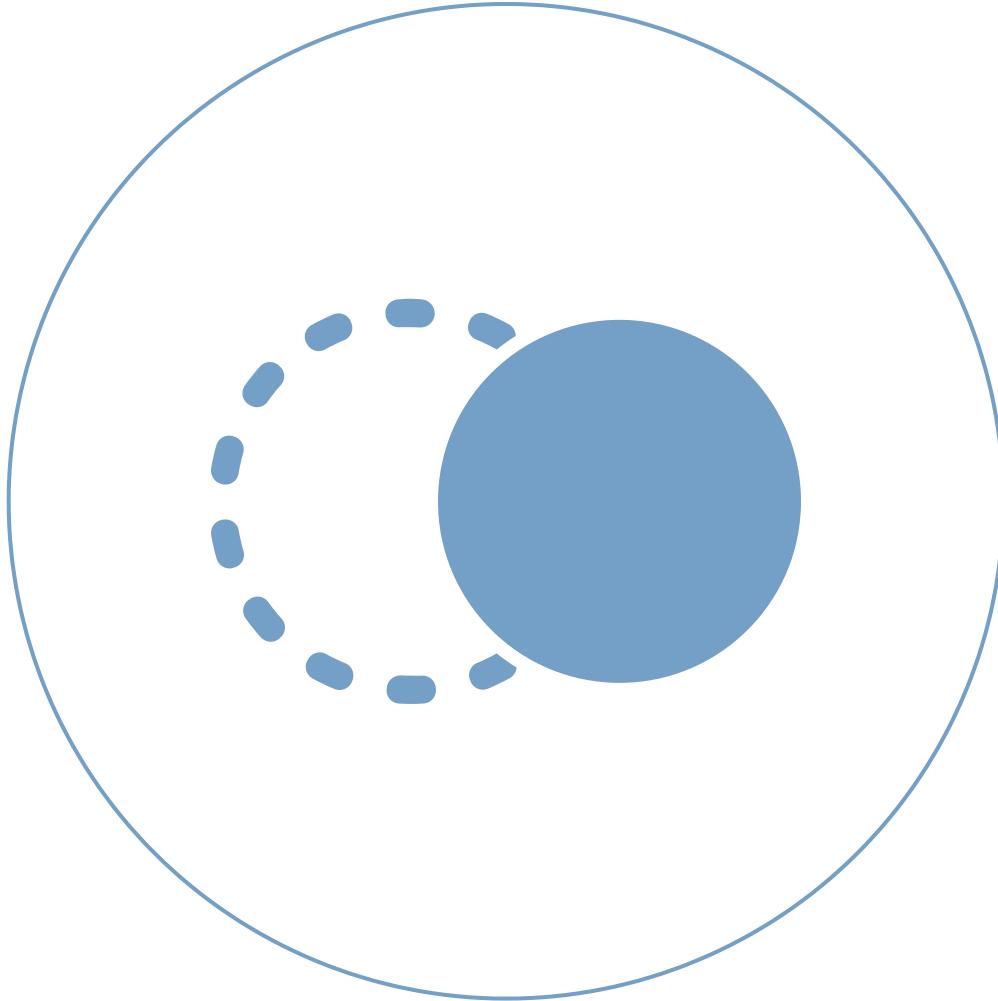
<https://youtu.be/hUnRCxnydCc>

How can LIME give explanations of complex models?

With the following workflow:



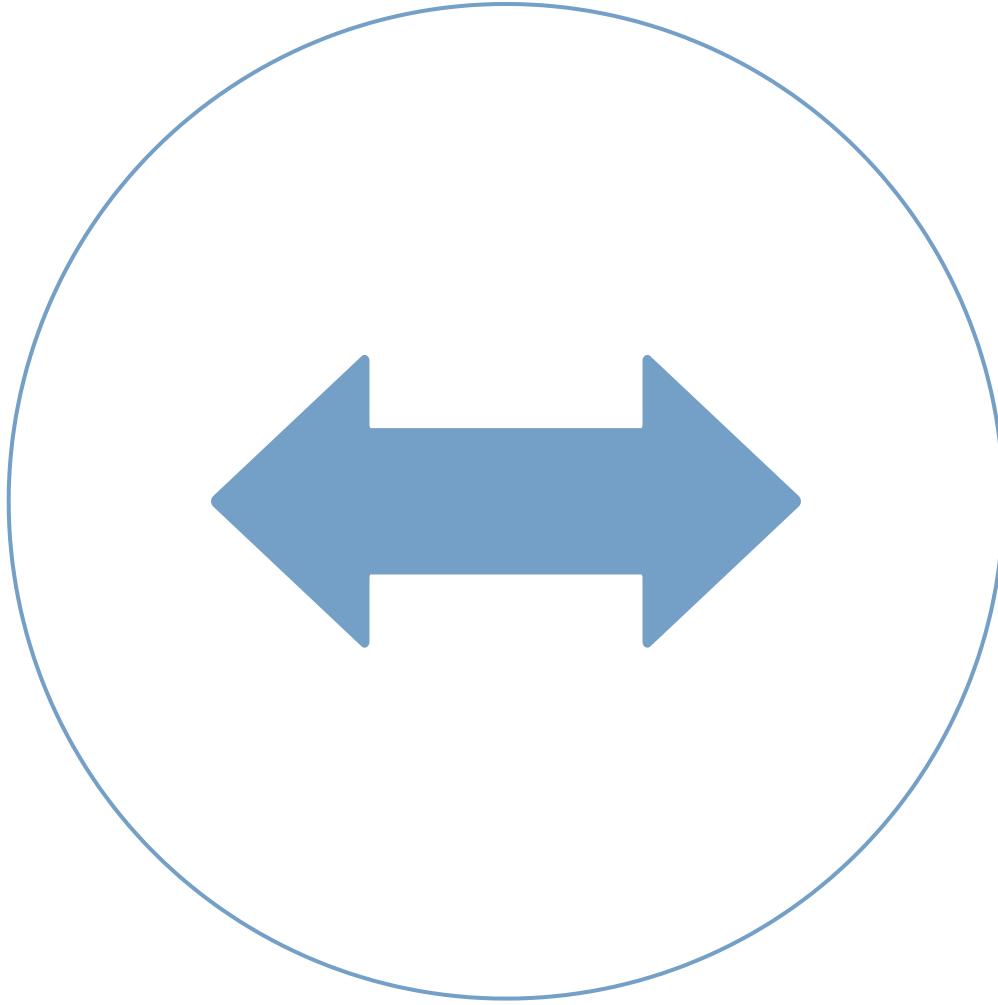
Steps of the LIME algorithm and workflow



Permutation

- Sampling from distribution curve or kernel density function
- Or by randomly leaving out words
- Predictions on every permutation with original (complex) machine learning model

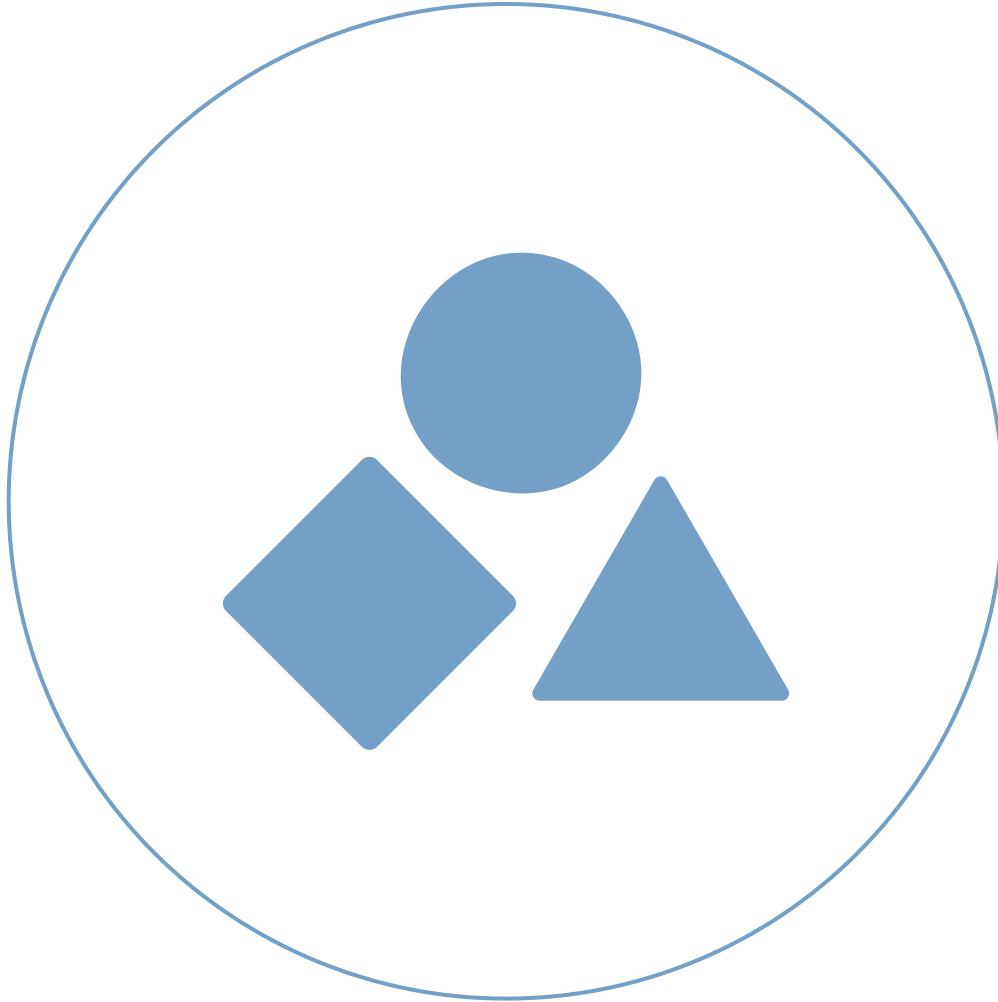
Steps of the LIME algorithm and workflow



Distance & Similarity

- Distance measure for each permutation with original instance
- Converted to similarity metric
- E.g. cosine similarity (text), gower (default for tabular data), euclidean distance or any other

Steps of the LIME algorithm and workflow



Most important variables

- Choose number of features
- Forward selection & ridge regression fit
- Highest weights in ridge regression fit
- Lasso regularization
- Decision tree splits

Steps of the LIME algorithm and workflow

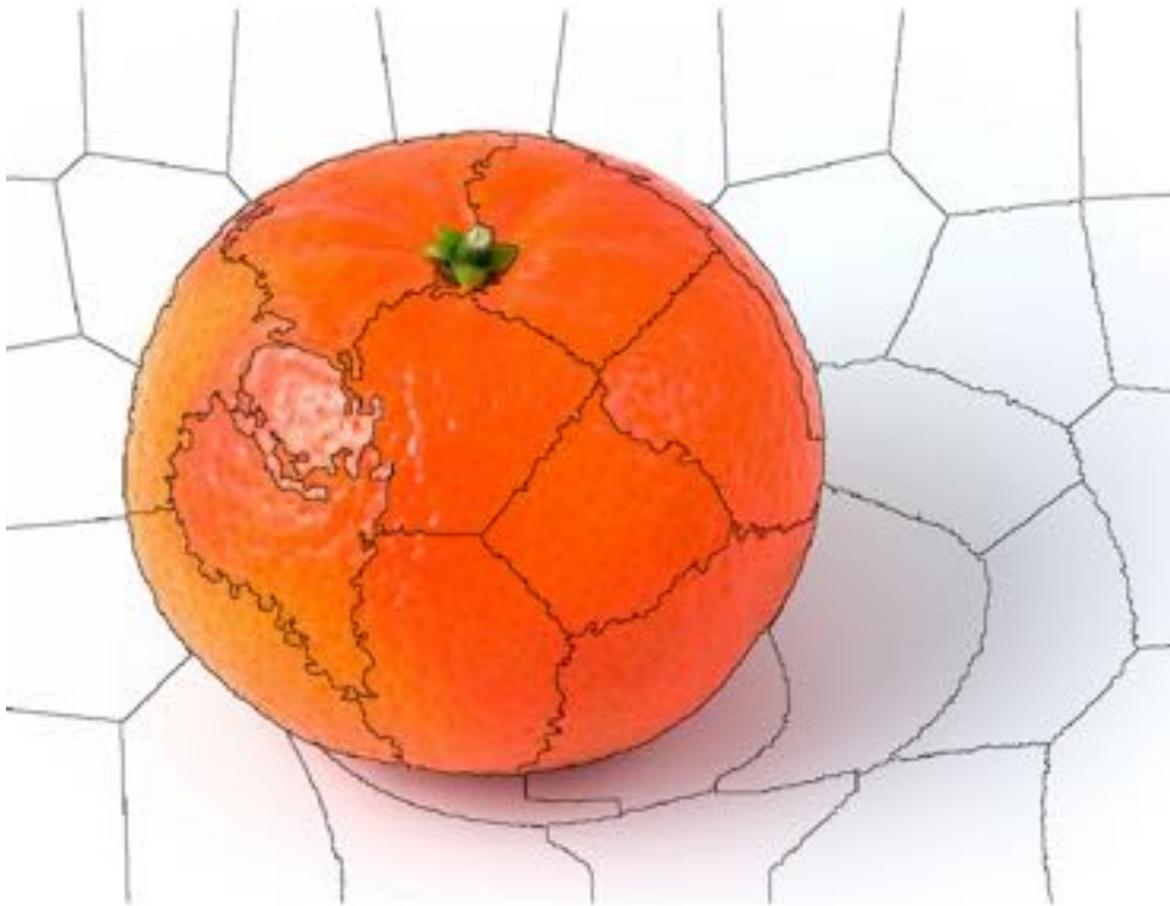


Simple local model

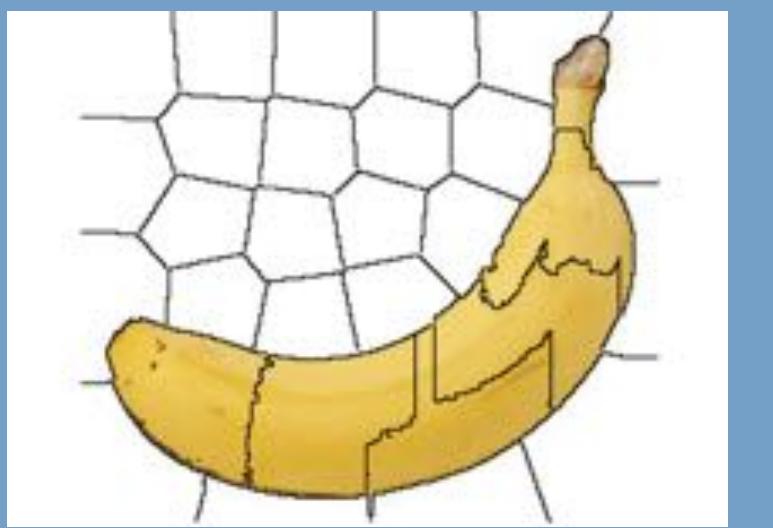
- Fit on output from machine learning model
- Output used as probability of each possible class
- Ridge regression (default in lime package)
- But principally, any model could be used

Applying LIME to our image classification model

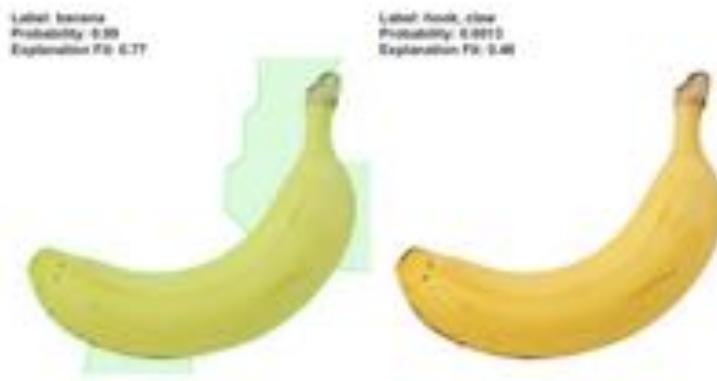
Code: https://shirinsplayground.netlify.com/2018/06/keras_fruits_lime/



- **Explaining predictions of**
 - our Keras model
 - a pretrained Imagenet model (VGG16)
- **On two new images**
downloaded from the internet



Label: banana
Probability: 0.99
Explanation F1: 0.677



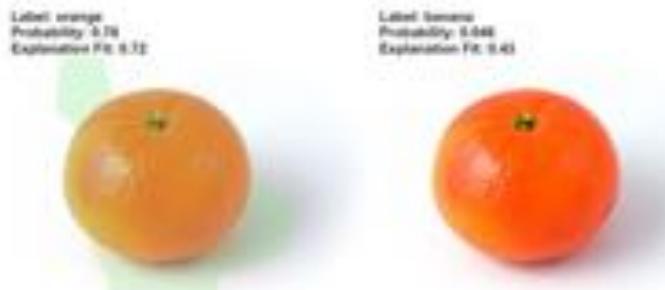
Label: knob, clear
Probability: 0.9913
Explanation F1: 0.446



Label: banana
Probability: 1
Explanation F1: 0.622



Label: orange
Probability: 0.78
Explanation F1: 0.72



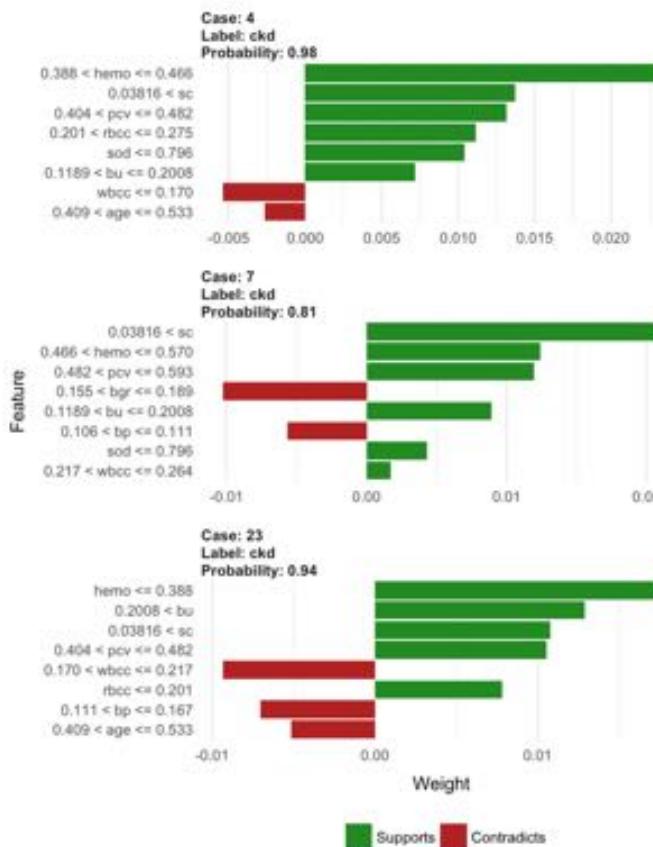
Label: banana
Probability: 0.046
Explanation F1: 0.43



Label: Clementine
Probability: 0.88
Explanation F1: 0.917

Explanations of tabular models

Chronic Kidney Disease (ckd) model trained with caret



Explanations of text models

Women's clothing reviews model trained with xgboost

Absolutely wonderful - silky and sexy and comfortable

Label predicted: 1 (97.25%)

Explainer fit: 0.94

Some major design flaws I had such high hopes for this dress and really wanted it to work for me. I initially ordered the petite small (my usual size) but I found this to be outrageously small, so small in fact that I could not zip it up! I reordered it in petite medium, which was just ok. Overall, the top half was comfortable and fit nicely, but the bottom half had a very tight under layer and several somewhat cheap (net) over layers. IMO, a major design flaw was the net over layer sewn directly into the zipper - it c

Label predicted: 0 (94.56%)

Explainer fit: 0.34

Flattering shirt This shirt is very flattering to all due to the adjustable front tie. It is the perfect length to wear with leggings and it is sleeveless so it pairs well with any cardigan. Love this shirt!!!

Label predicted: 1 (98.61%)

Explainer fit: 0.62

Pretty party dress with some issues. This is a nice choice for holiday gatherings. I like that the length grazes the knee so it is conservative enough for office related gatherings. The size small fit me well - I am usually a size 2/4 with a small bust. In my opinion it runs small and those with larger busts will definitely have to size up (but then perhaps the waist will be too big). The problem with this dress is the quality. The fabrics are terrible. The delicate netting type fabric on the top layer of skirt got stuck in the zip

Label predicted: 0 (98.61%)

Explainer fit: 0.44

Problems and Pitfalls with LIME



- Uses linearly weighted combinations of features
- Depending on the goodness of fit linearity does not apply to new instances.
- Coverage is not clear

Anchors

If-then rules locally anchor predictions with sufficiently high probability.

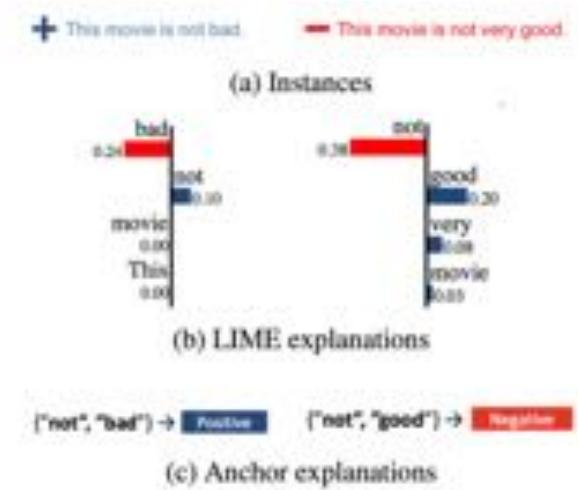
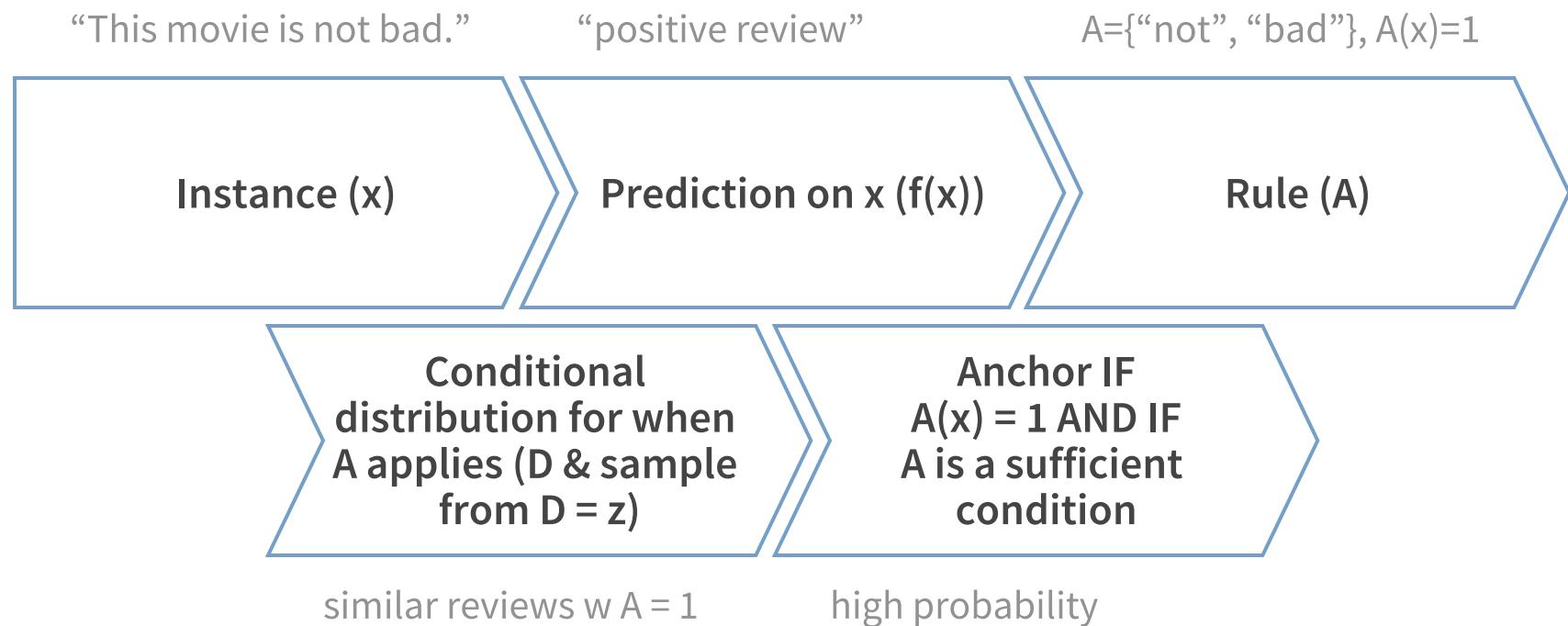


Figure 1: Sentiment predictions, LSTM

<https://homes.cs.washington.edu/~marcotcr/aaai18.pdf> & <https://github.com/marcotcr/anchor>

Advantages: Easy to understand & clear coverage!

Anchor explanations

Figures 3 & 5 and Tables 2 & 3 from Ribeiro, Singh and Guestrin 2018



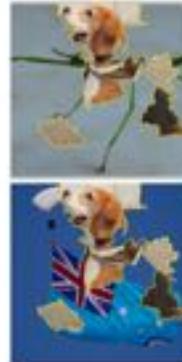
(a) Original image



(b) Anchor for “beagle”



(c) Images where Inception predicts $P(\text{beagle}) > 90\%$



	English	Portuguese
	This is the question we must address	Esta é a questão que temos que enfrentar
	This is the problem we must address	Este é o problema que temos que enfrentar
	This is what we must address	É isso que temos de enfrentar

Table 2: Anchors (in bold) of a machine translation system for the Portuguese word for “This” (in pink).

	If	Predict
adult	No capital gain or loss, never married	$\leq 50K$
	Country is US, married, work hours > 45	$> 50K$
rev	No priors, no prison violations and crime not against property	Not rearrested
	Male, black, 1 to 5 priors, not married, and crime not against property	Re-arrested
lending	FICO score ≤ 649	Bad Loan
	$649 \leq \text{FICO score} \leq 699$ and $\$5,400 \leq \text{loan amount} \leq \$10,000$	Good Loan

Table 3: Generated anchors for Tabular datasets

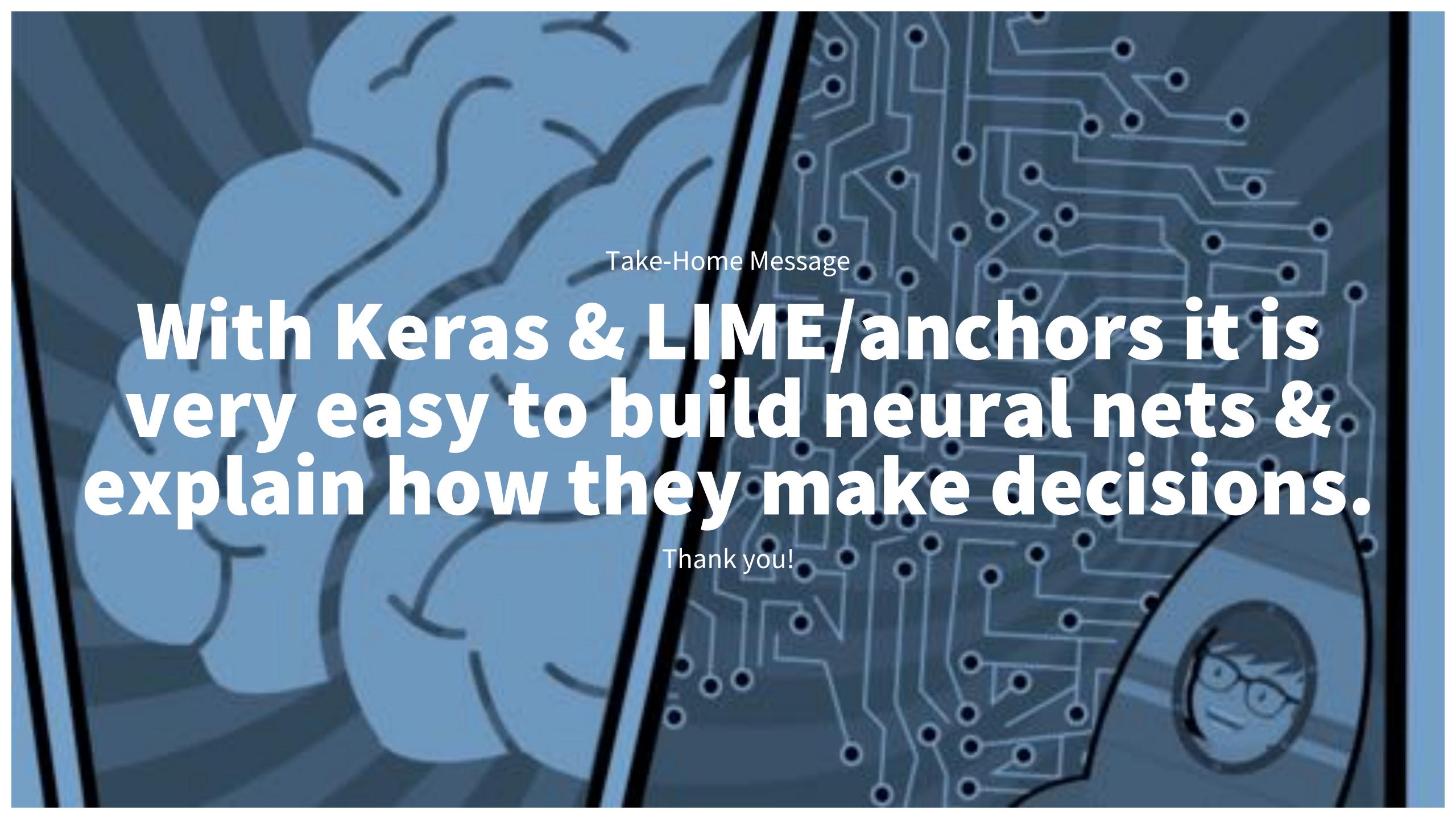
**IF Country = United-States AND Capital Loss = Low
AND Race = White AND Relationship = Husband
AND Married AND $28 < \text{Age} \leq 37$
AND Sex = Male AND High School grad
AND Occupation = Blue-Collar
THEN PREDICT Salary > \$50K**



RIGHT

Take-Home Message

**WE need to make sure that our
models are not going to cause
harm to people or society!**

The background of the slide features a blue-toned illustration of a human brain on the left, rendered in a soft, glowing blue. On the right, there is a detailed blue and white circuit board with various electronic components like resistors and capacitors. The overall theme is the intersection of neuroscience and technology.

Take-Home Message

**With Keras & LIME/anchors it is
very easy to build neural nets &
explain how they make decisions.**

Thank you!

codecentric.AI Bootcamp

<https://bootcamp.codecentric.ai/>



Für wen?

Lernzettel

Module

Disserten

Firmen Deep-Dive

Login

Register

ARTIFICIAL INTELLIGENCE BOOTCAMP

Um Grundlagen der künstlichen Intelligenz und des Machine Learning zu beherrschern, braucht man keinen Doktor in Mathematik. Mit unserem codecentric.AI Bootcamp geben wir dir einen komprimierten Schnelldurchlauf durch ML, Deep Learning, Computer Vision und Co. - hands-on und kostengünstig.

KOSTENLOS TEILNEHMEN





Thank you!
And stay
connected

Am Mittelhafen 14, Münster



ShirinGlander



@ shirin.glander@codecentric.de

www.shirin-glander.de

Additional information

https://youtu.be/RcUdUZf8_SU

<https://youtu.be/CY3t11vuuOM>

<https://github.com/marcotcr/lime>

<https://github.com/thomasp85/lime>

<https://github.com/marcotcr/anchor>