# Deep Learning Illustrated

## A Course Demo

Dr. Jon Krohn
jon@untapt.com
Slides available at jonkrohn.com/talks

March 6th, 2019

NYC DATA SCIENCE ACADEMY

# Outline

# Outline

# Outline

# Outline

# Outline

# Demand for AI Talent

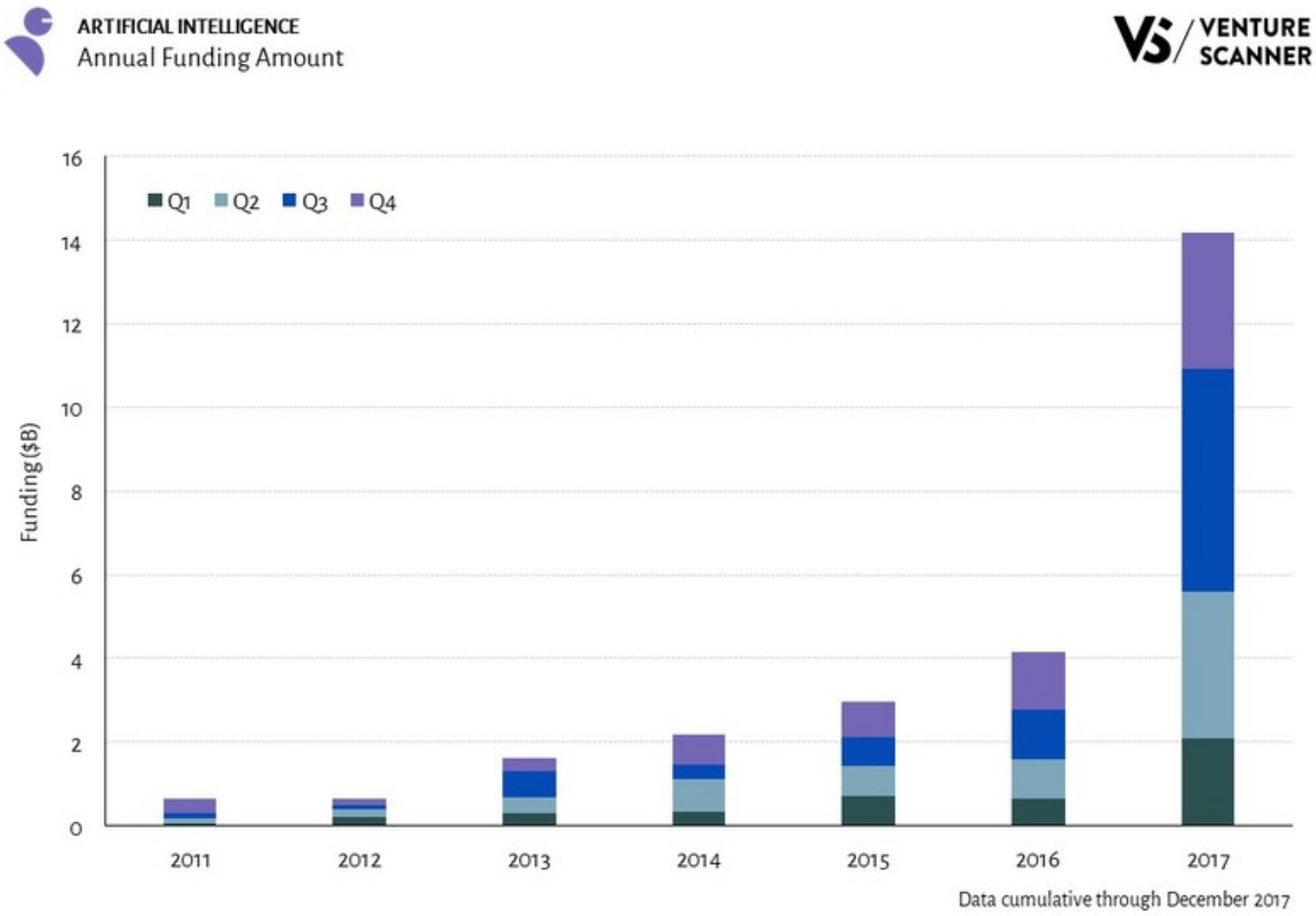
## i.e., *Deep Learning* talent

"Of the ten most valuable quoted companies in the world, seven say they have plans to put deep-learning-based AI at the heart of their operations"

~*The Economist* (Feb. 17th, 2018)

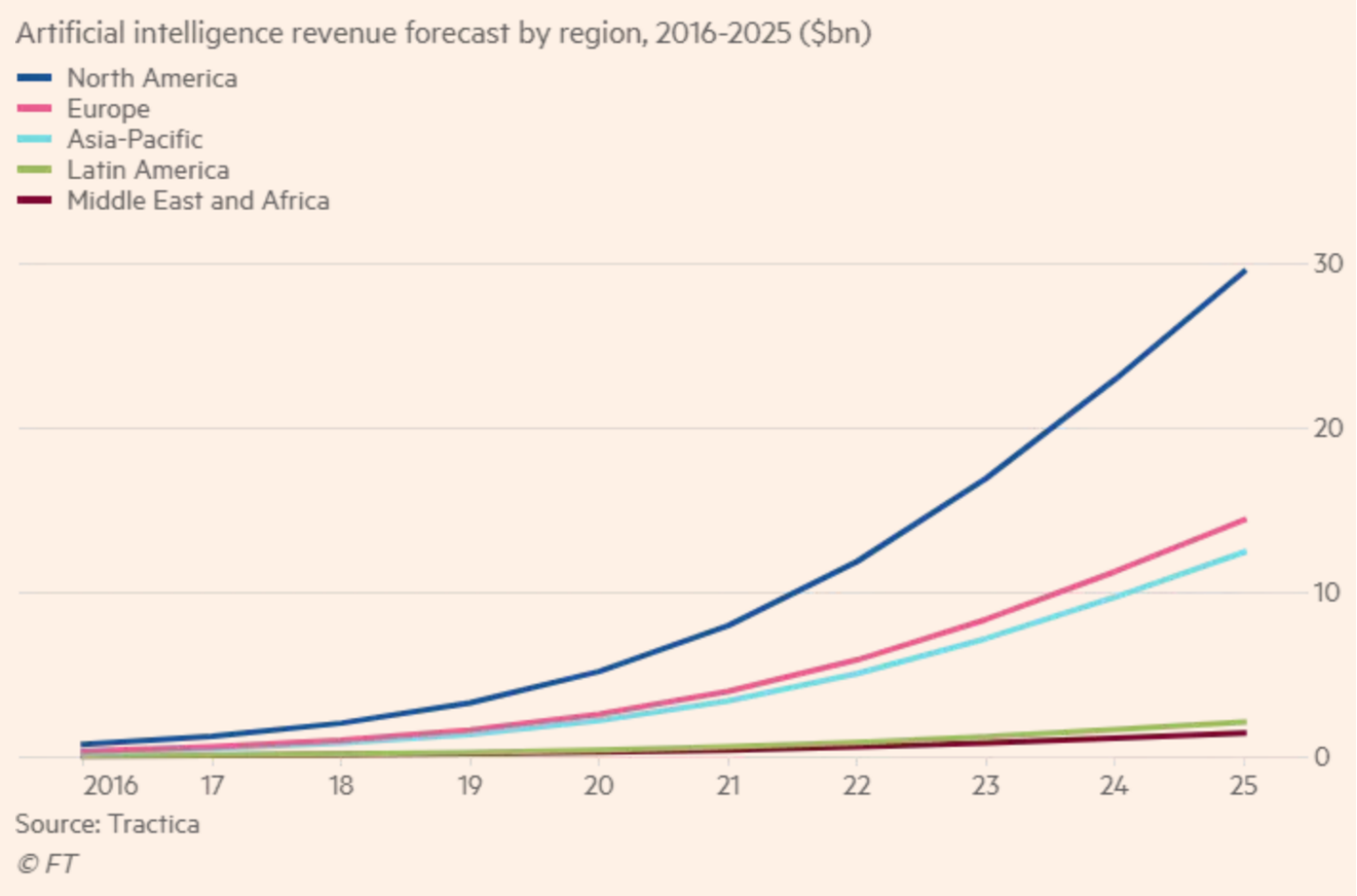
# Demand for AI Talent

## i.e., *Deep Learning* talent



# Demand for AI Talent

## i.e., *Deep Learning* talent



# Demand for AI Talent

## i.e., *Deep Learning* talent

According to JF Gagne's [Global AI Talent Report 2018]:

1. 22k Ph.D.-educated researchers globally
2. 3k of those currently looking
3. 5k publishing / presenting at AI conferences

# Demand for AI Talent

## i.e., *Deep Learning* talent

According to JF Gagne's [Global AI Talent Report 2018]:

1. 22k Ph.D.-educated researchers globally
2. 3k of those currently looking
3. 5k publishing / presenting at AI conferences

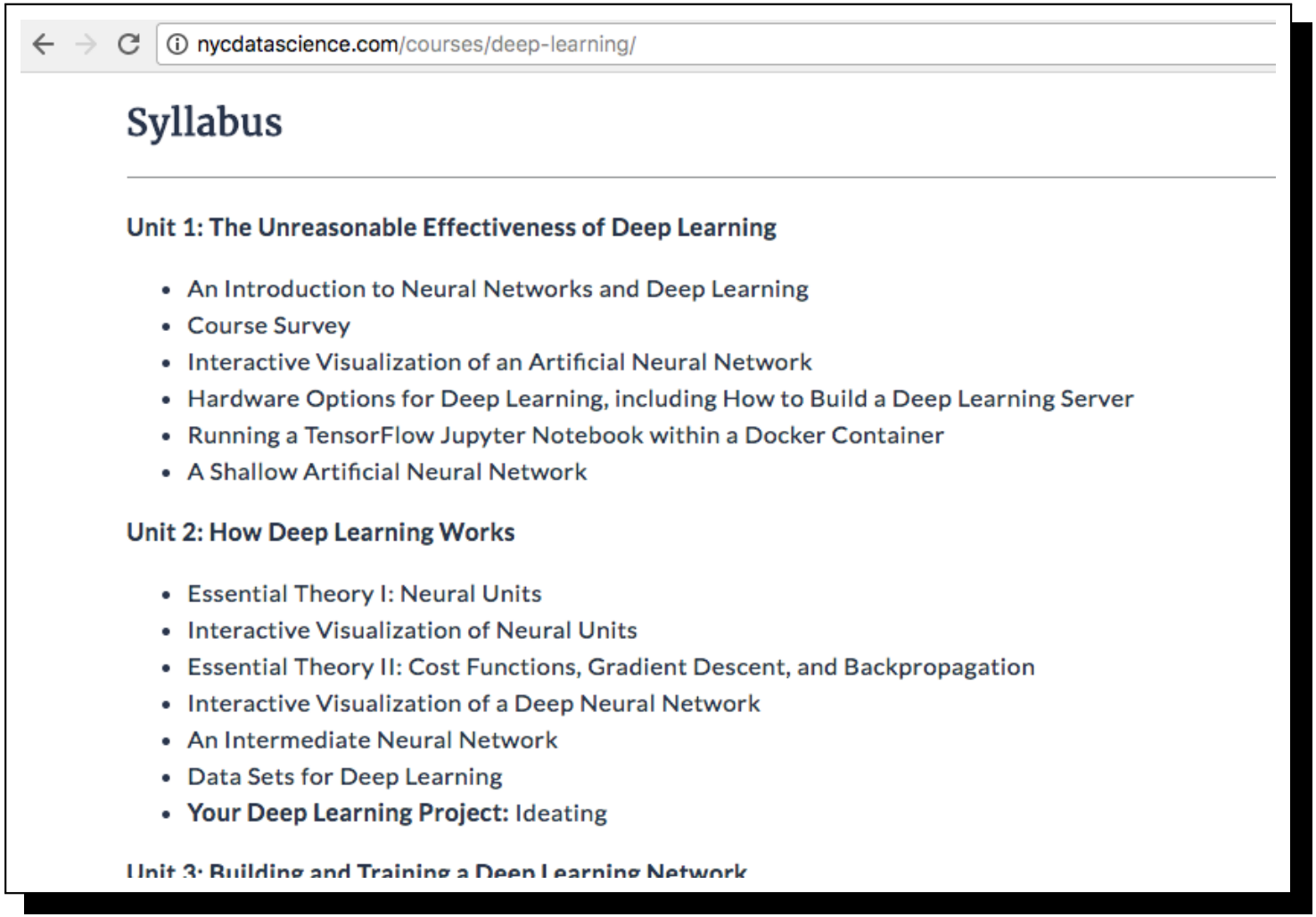
# Demand for AI Talent

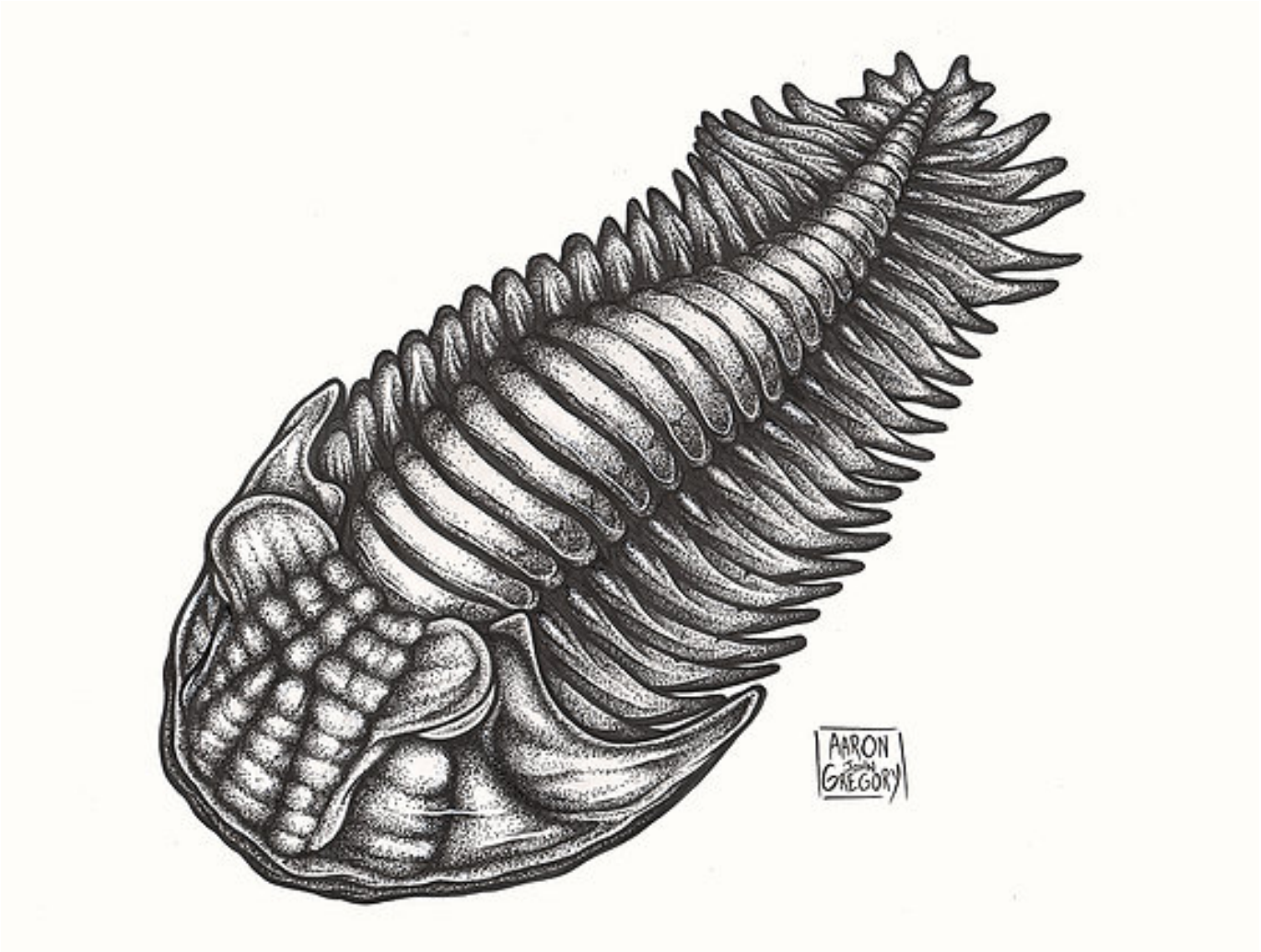
## i.e., *Deep Learning* talent

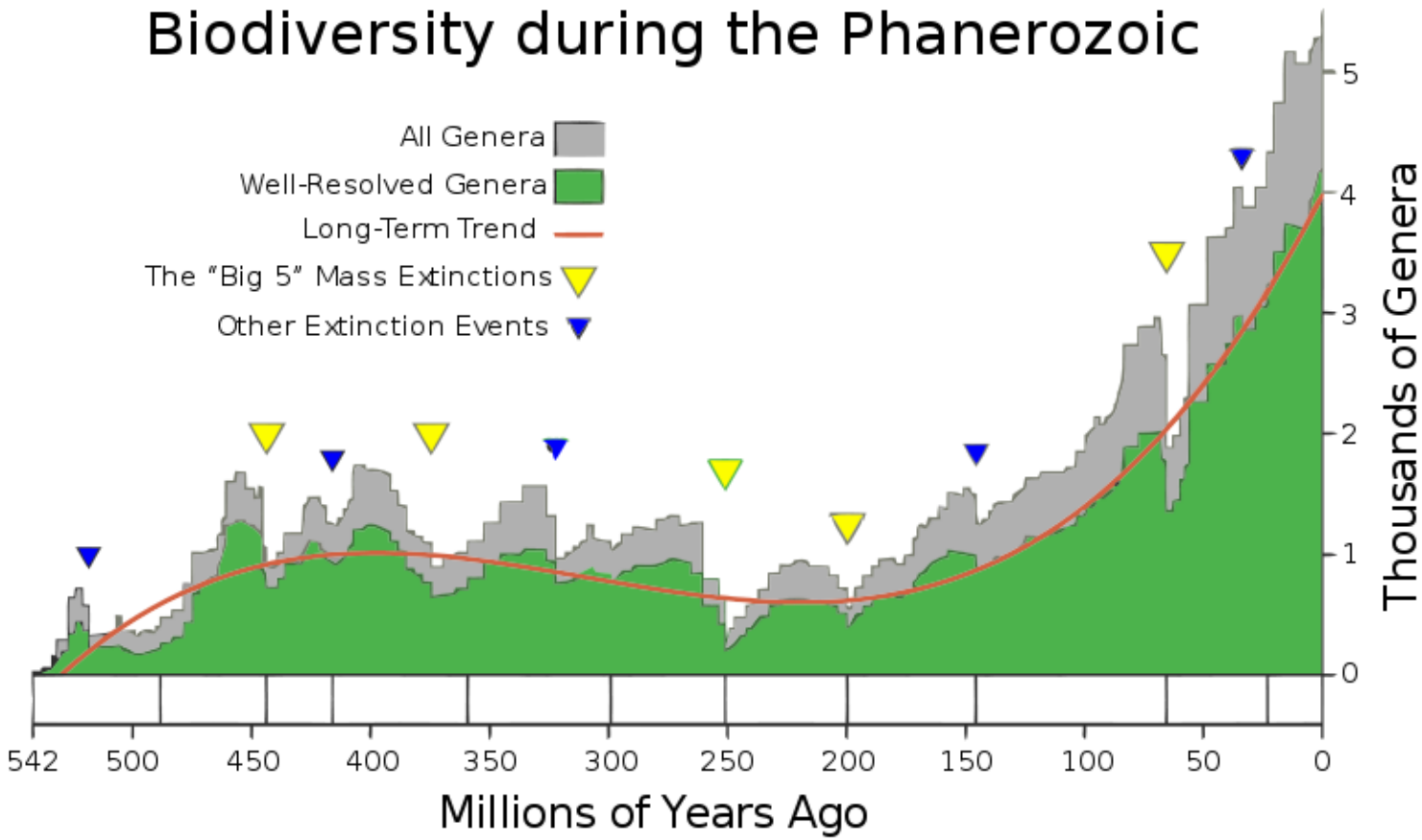
According to JF Gagne's [Global AI Talent Report 2018]:

1. 22k Ph.D.-educated researchers globally
2. 3k of those currently looking
3. 5k publishing / presenting at AI conferences

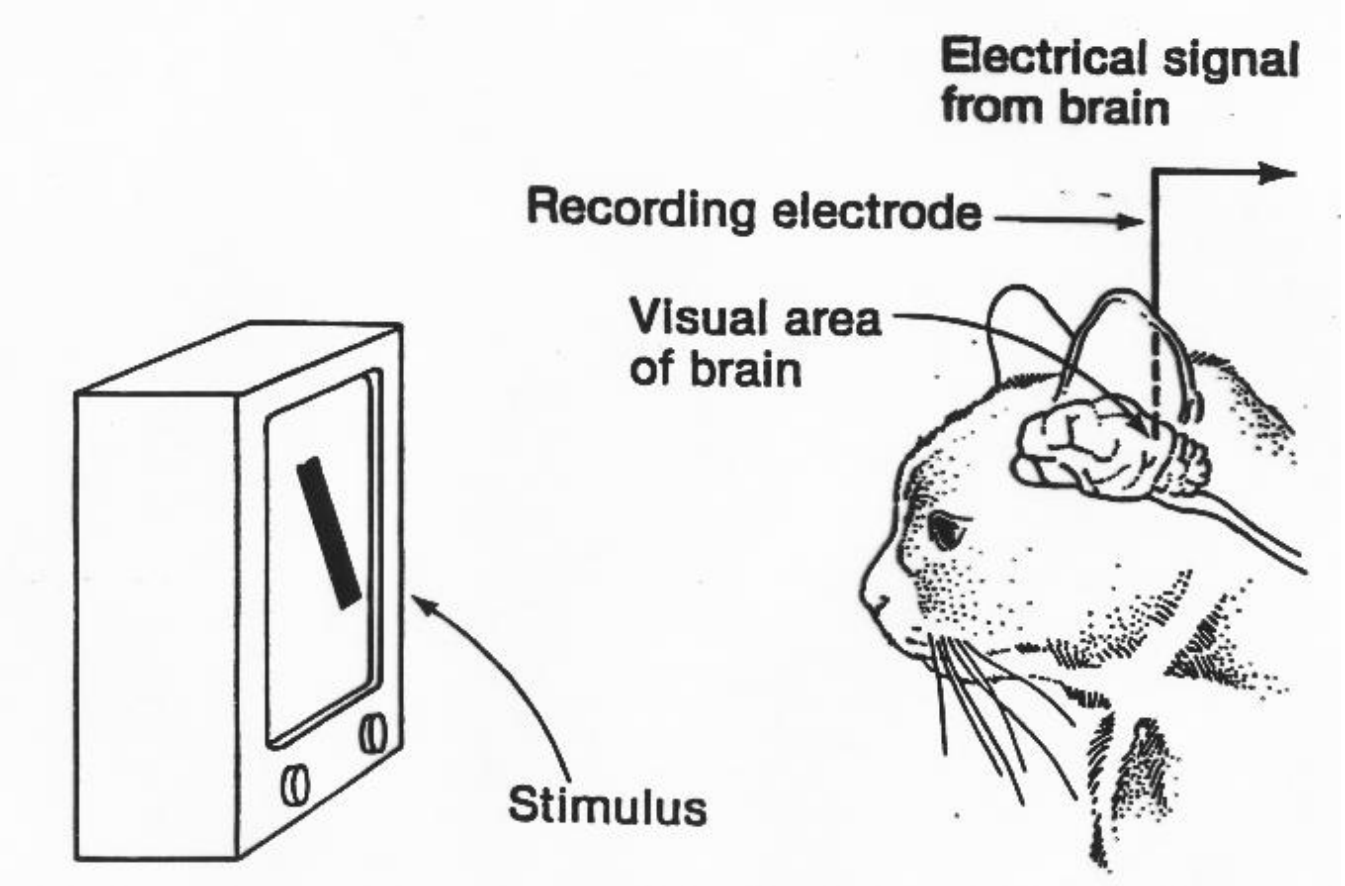
# An Intro to Deep Learning

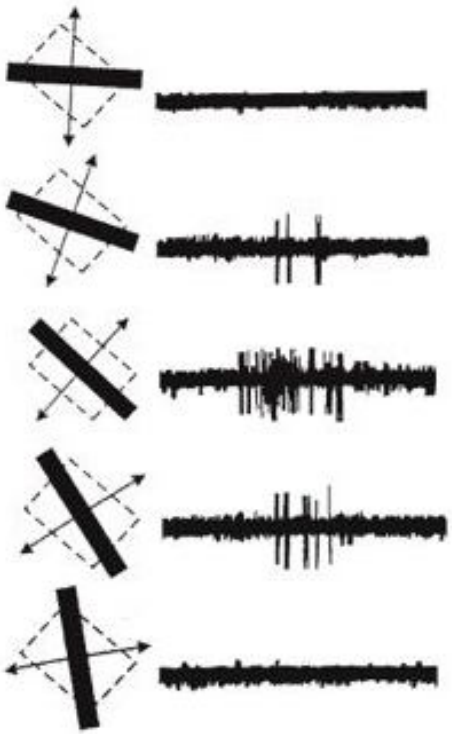


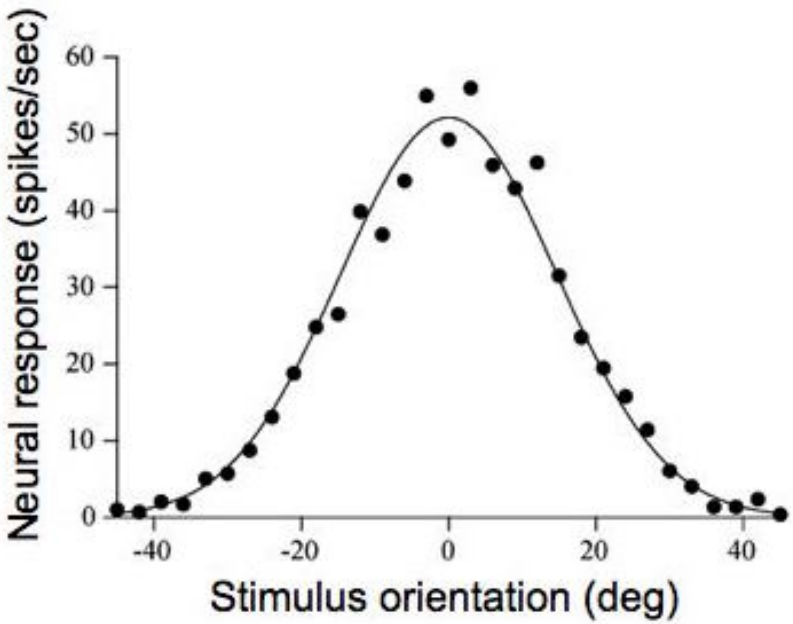



Biodiversity during the Phanerozoic
All Genera
Well-Resolved Genera
Long-Term Trend
The "Big 5" Mass Extinctions
Other Extinction Events
542
500
450
400
350
300
250
200
150
100
50
0
Millions of Years Ago
0
1
2
3
4
5
Thousands of Genera

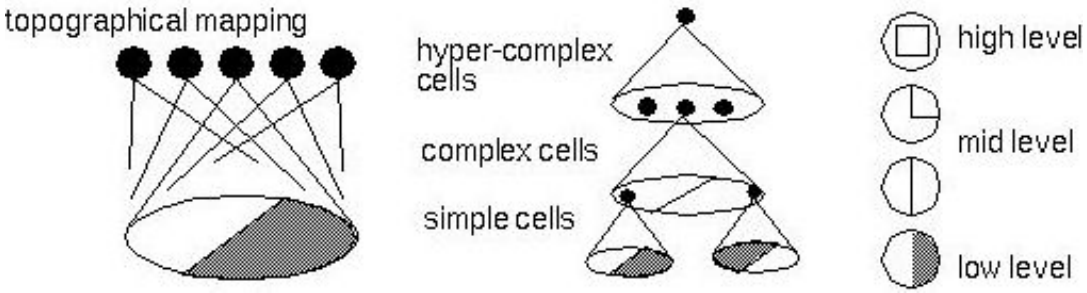
# Hubel & Wiesel (1959)

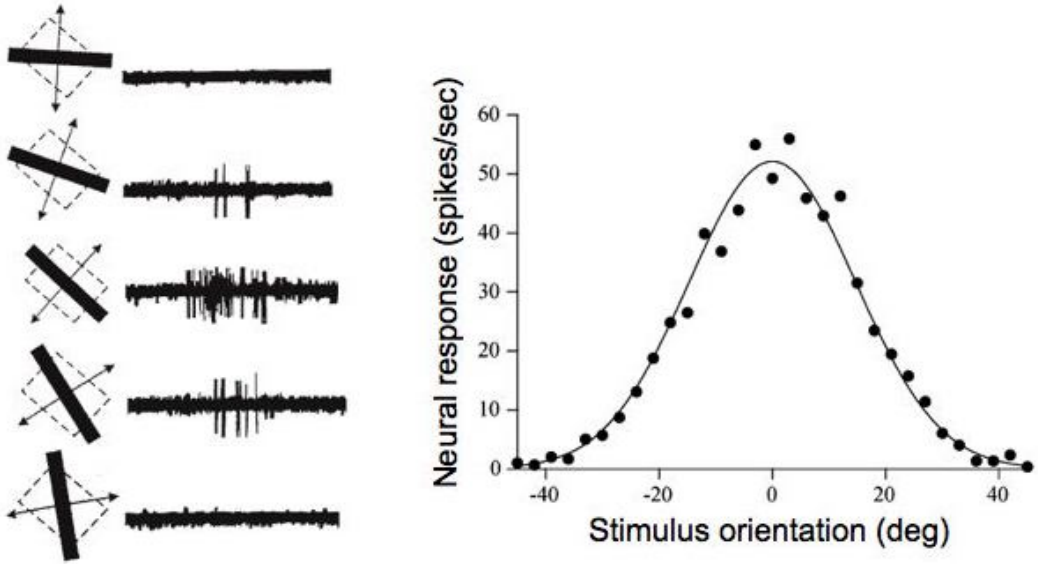




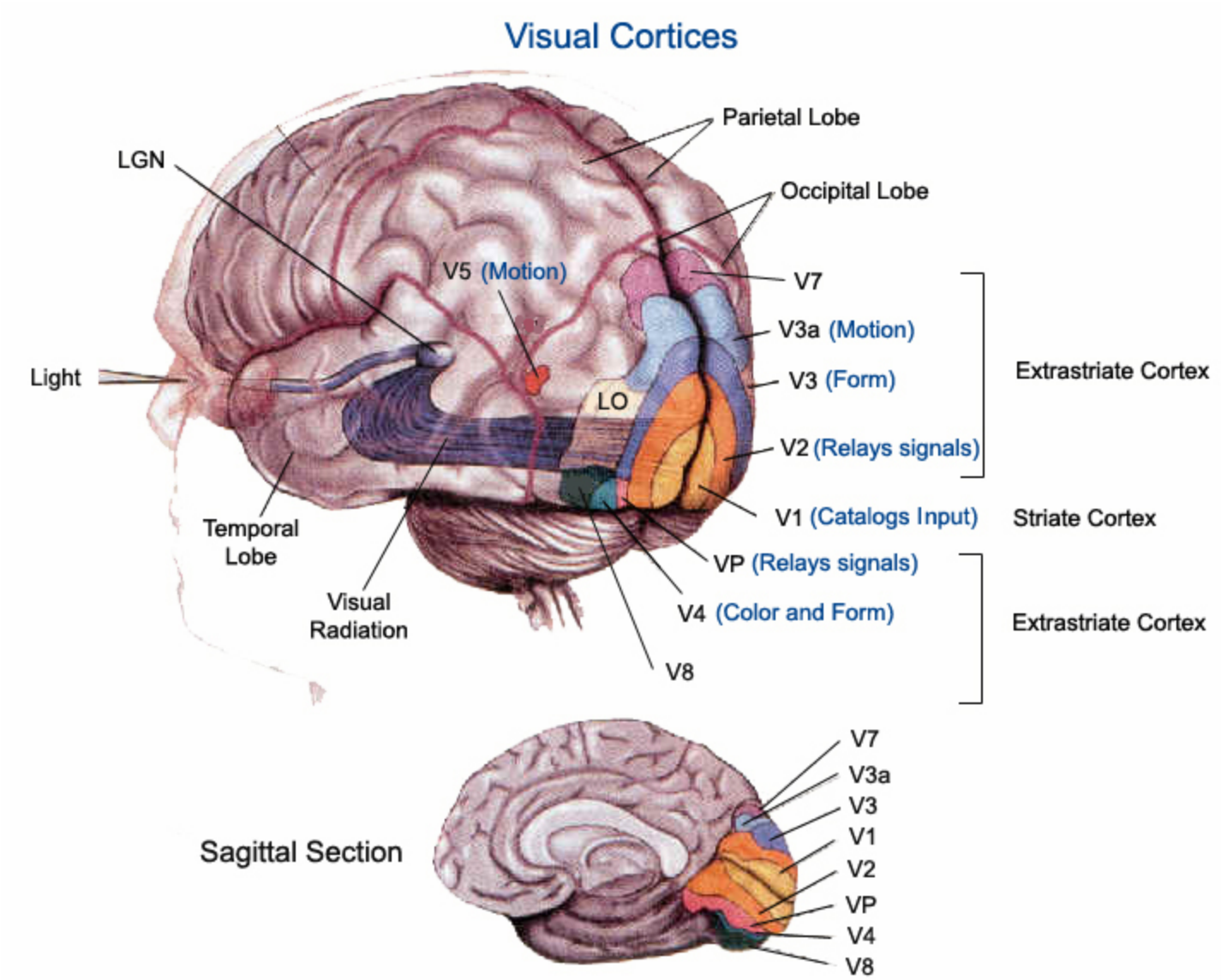


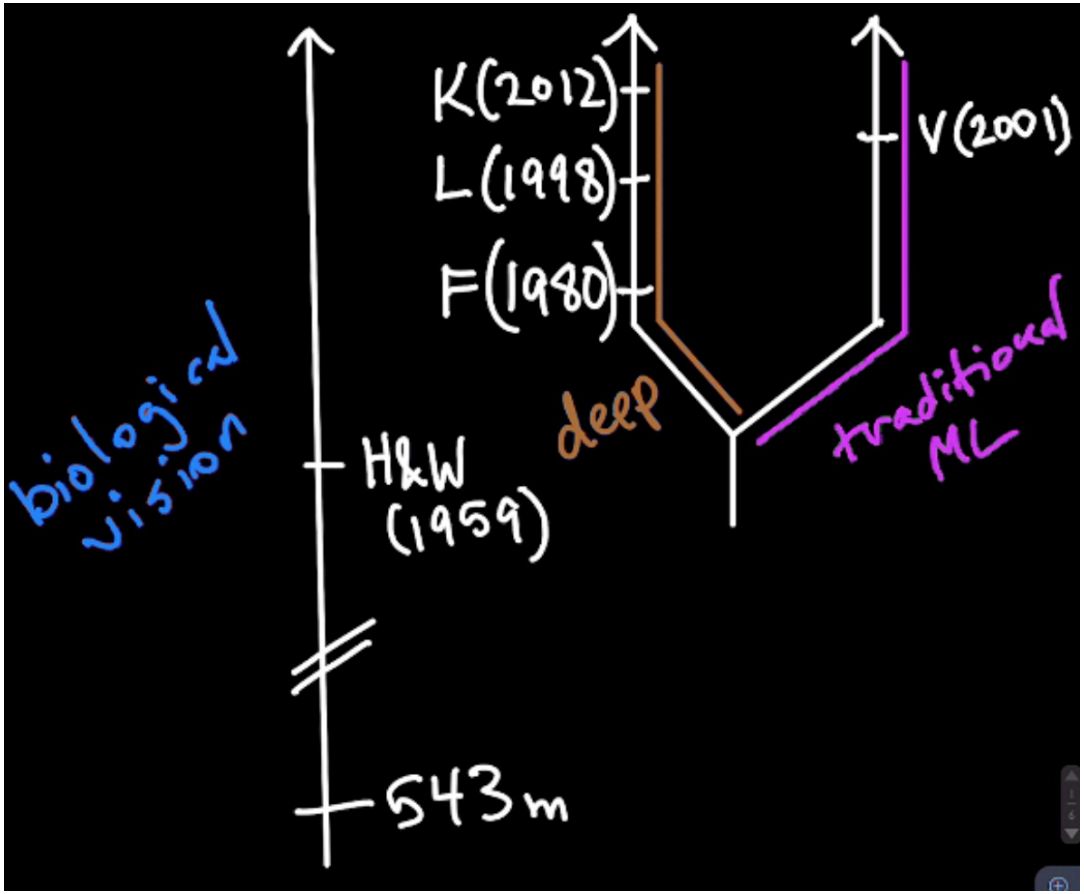

Hubel & Wiesel, 1968

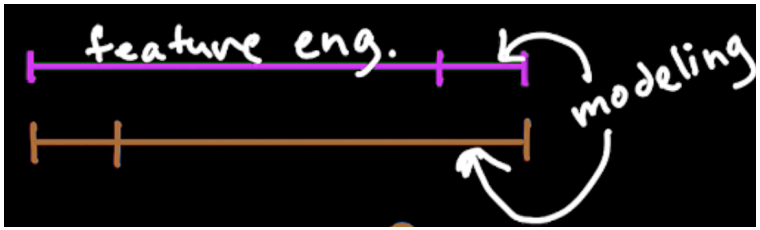



Hubel & Wiesel, 1968

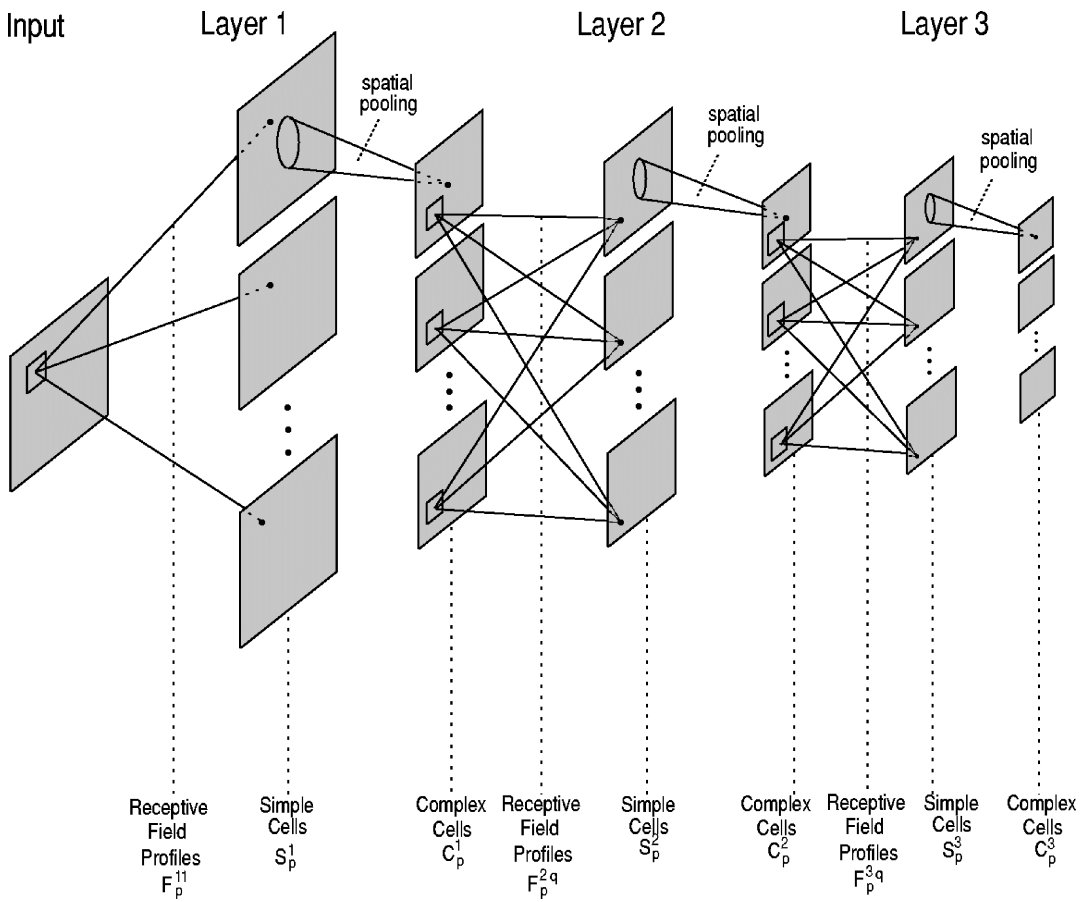


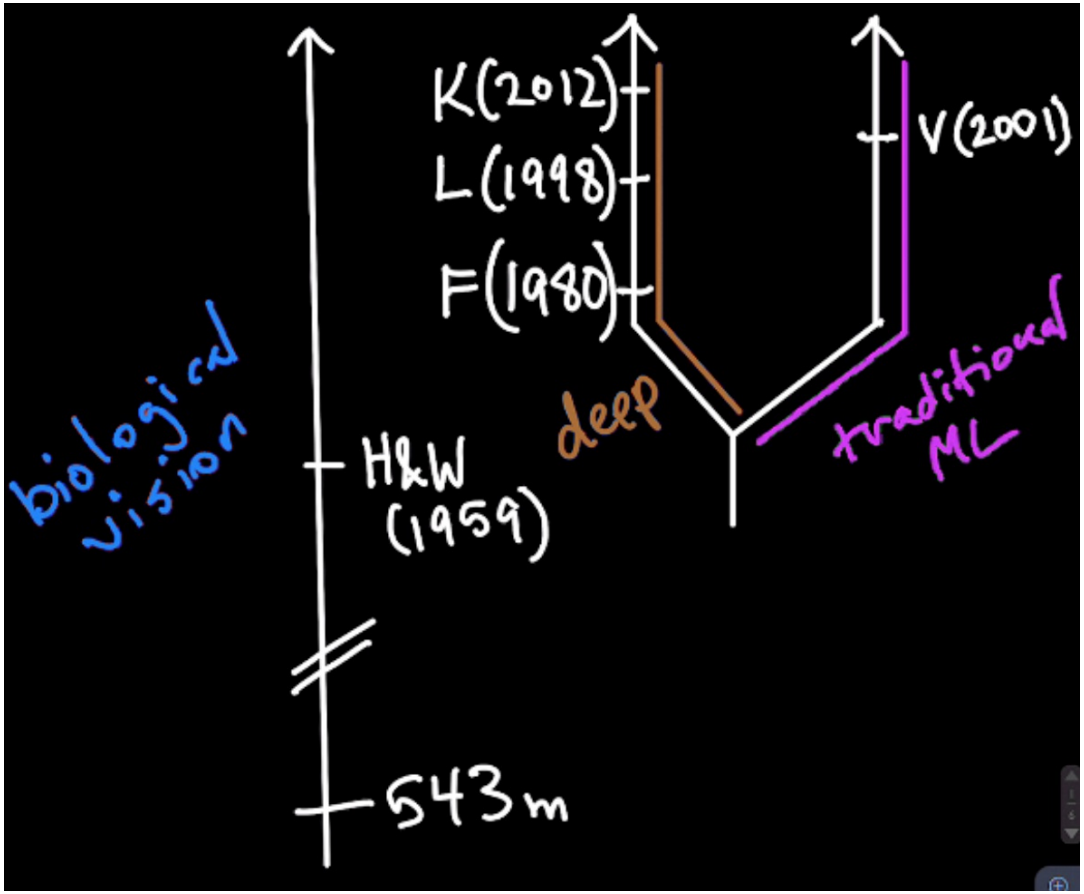


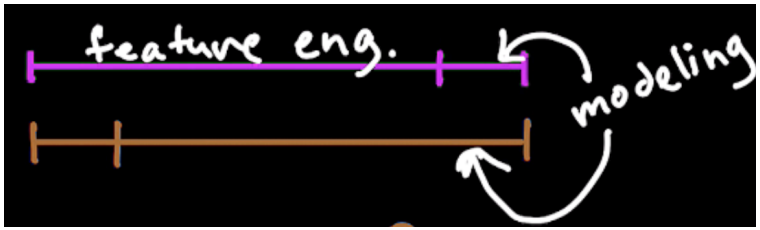



# Neocognitron

## Fukushima (1980)

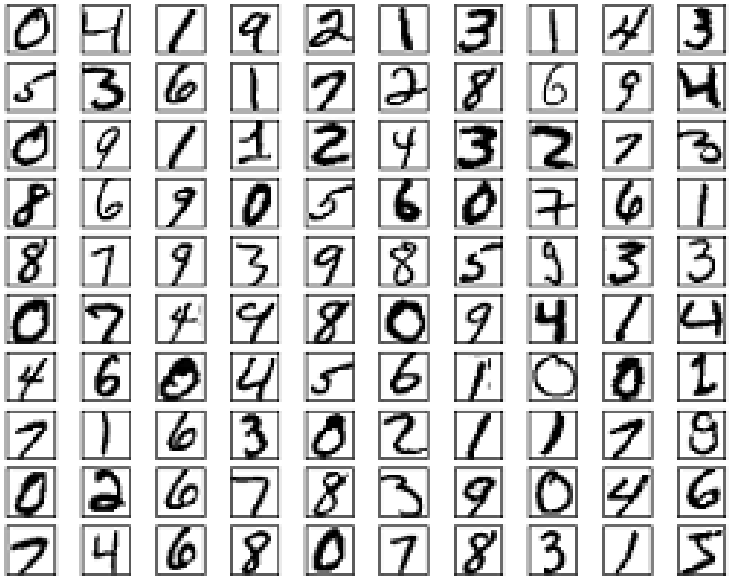




# MNIST Digits & LeNet-5

## LeCun, Boutou, Bengio & Haffner (1998)





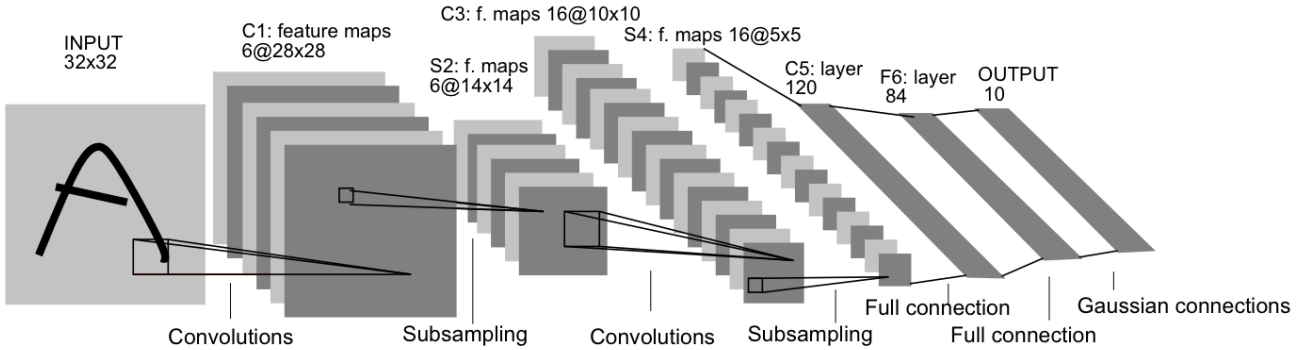
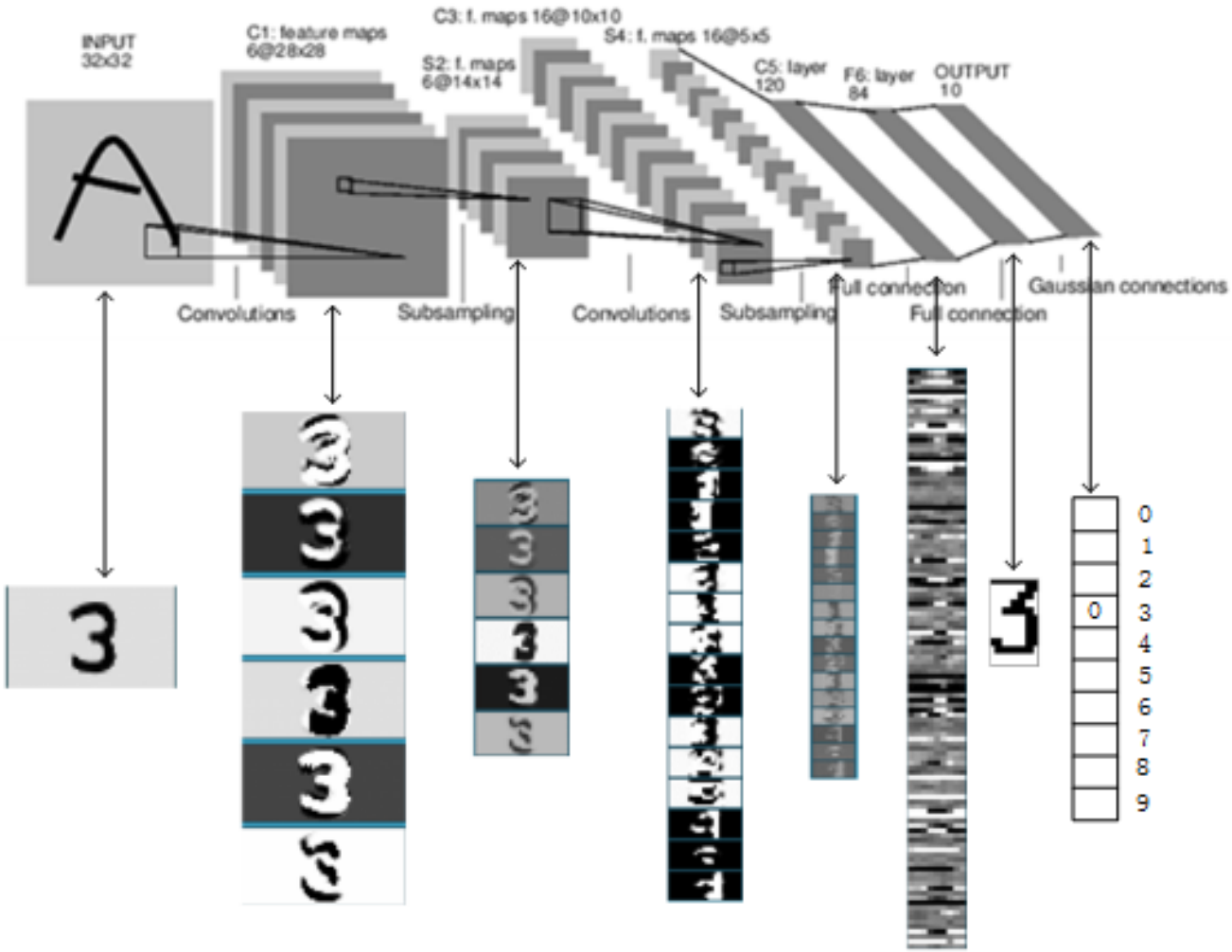
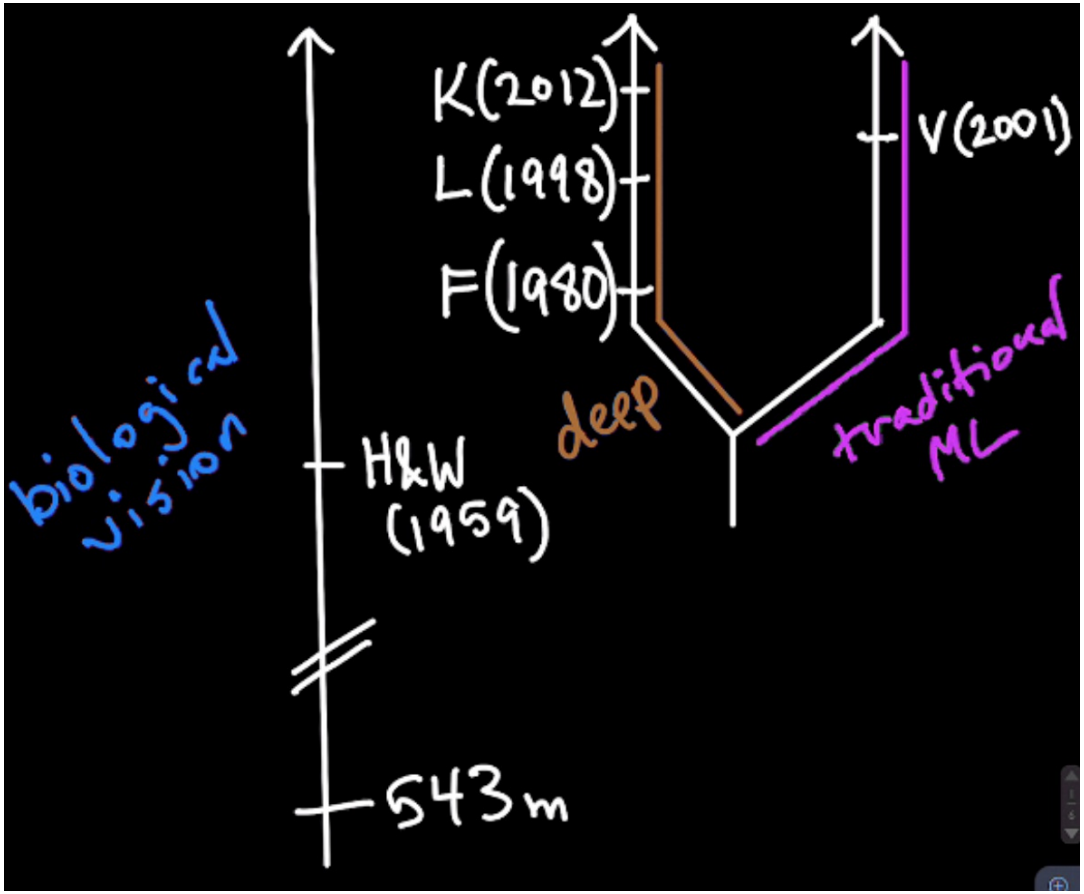


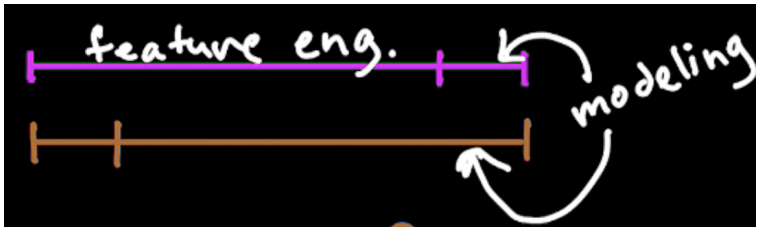

Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

# LeNet-5

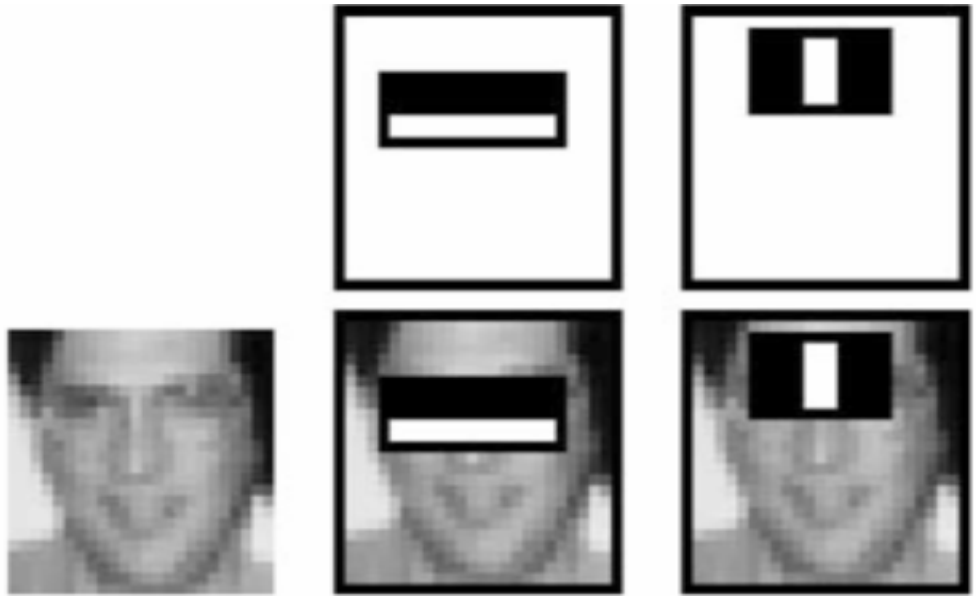
## LeCun, Boutou, Bengio & Haffner (1998)

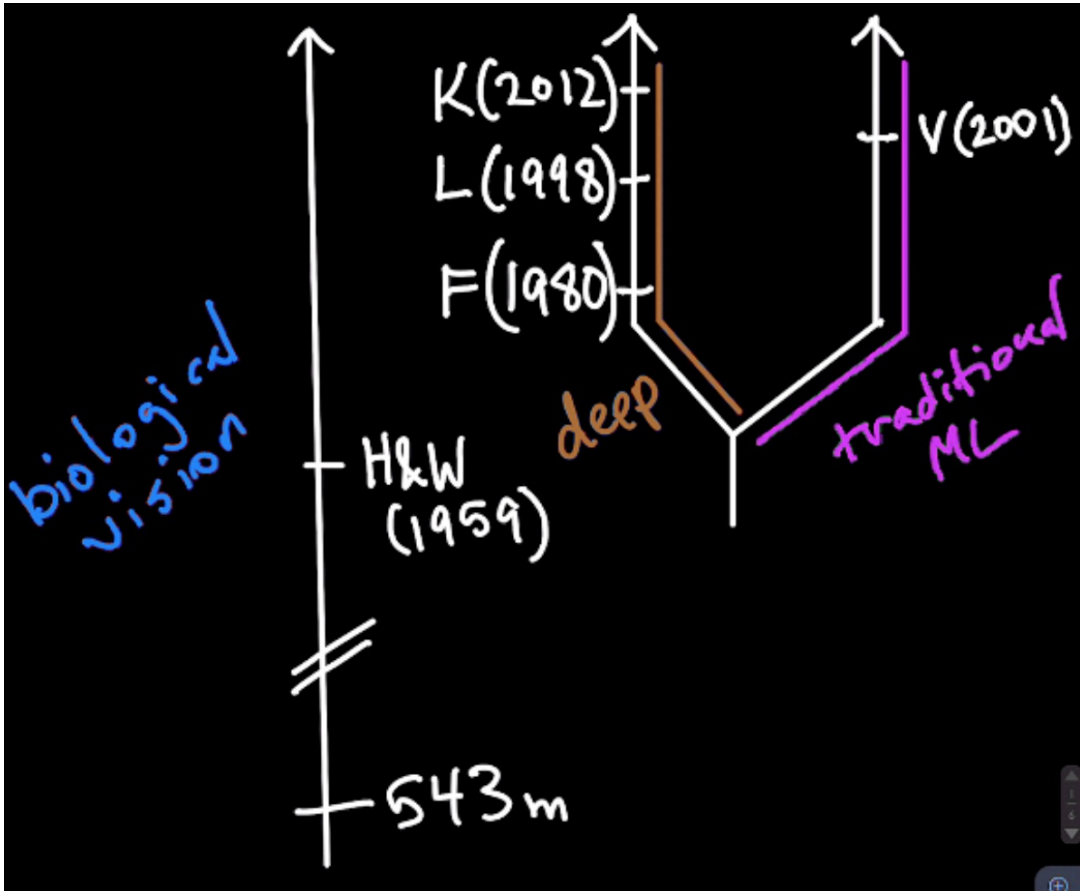


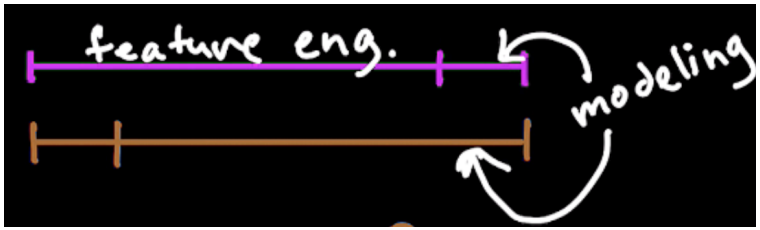


# Viola & Jones (2001)

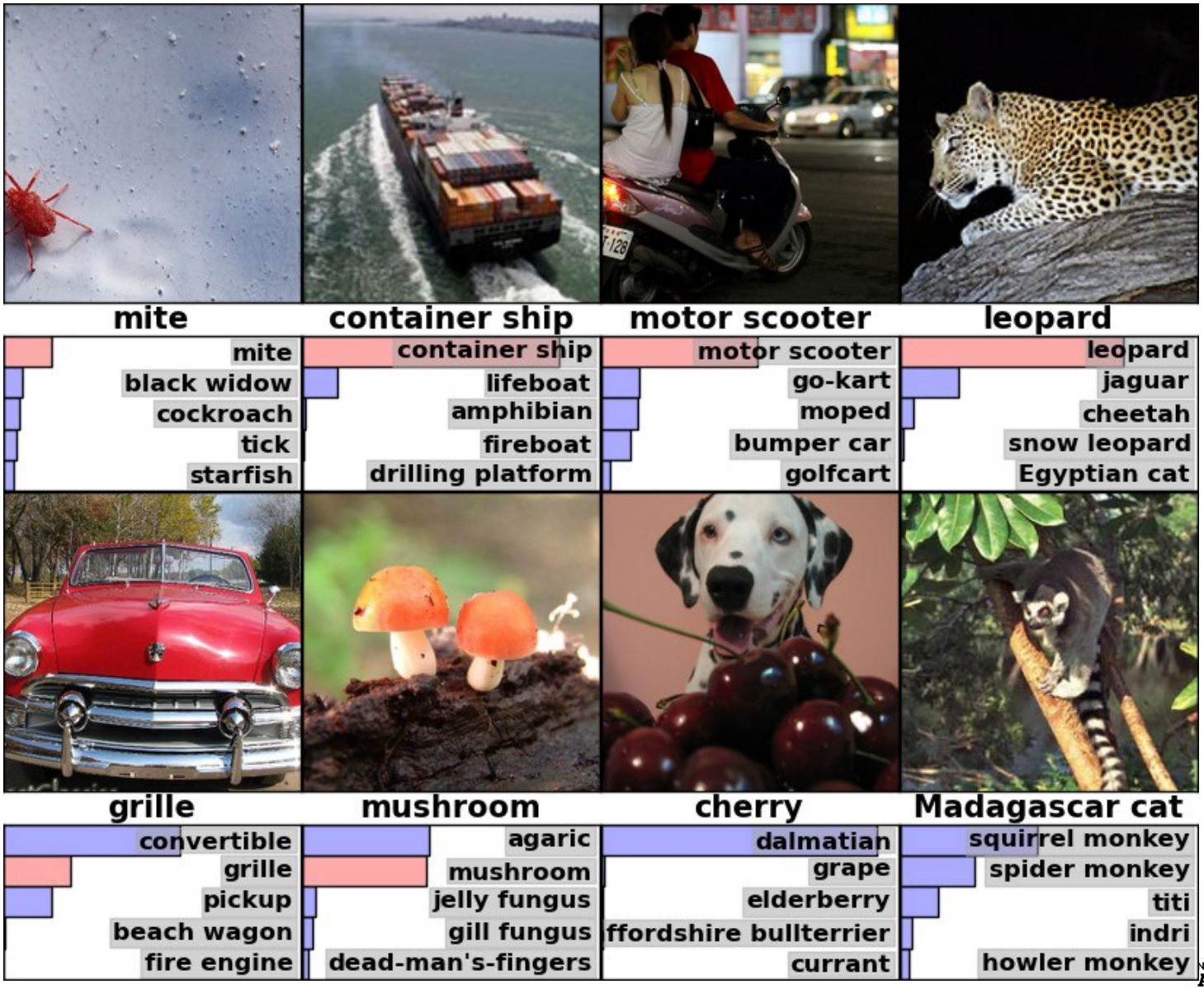




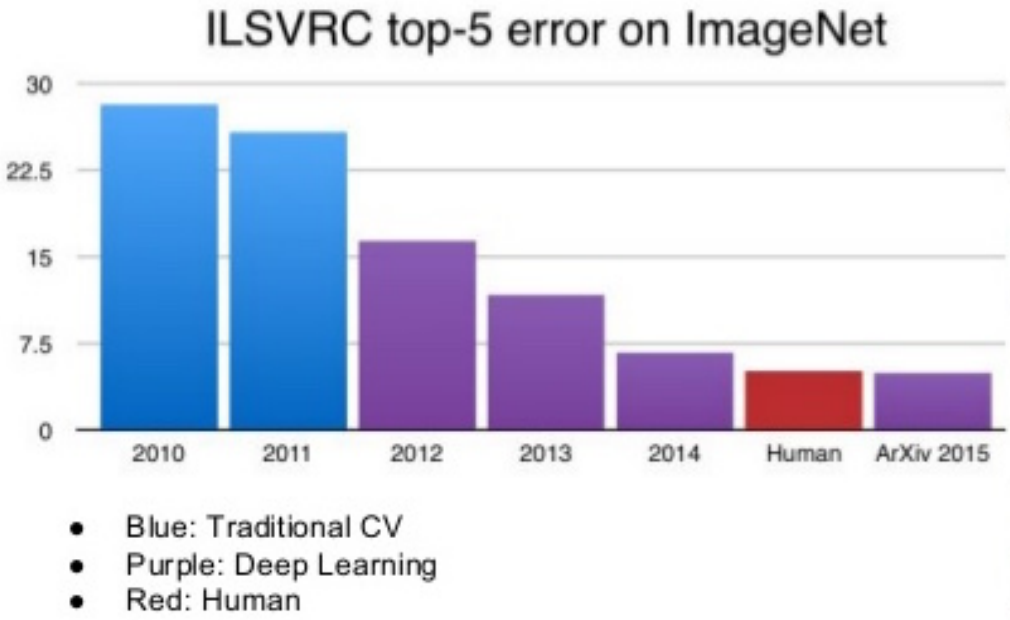
# ImageNet

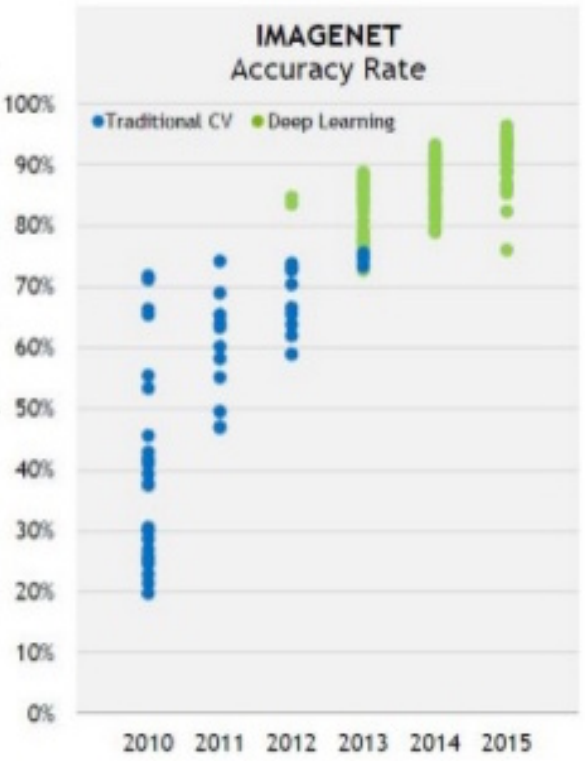
## Fei-Fei Li et al. (2009), 14m images, 22k categories



# ImageNet Classification Error

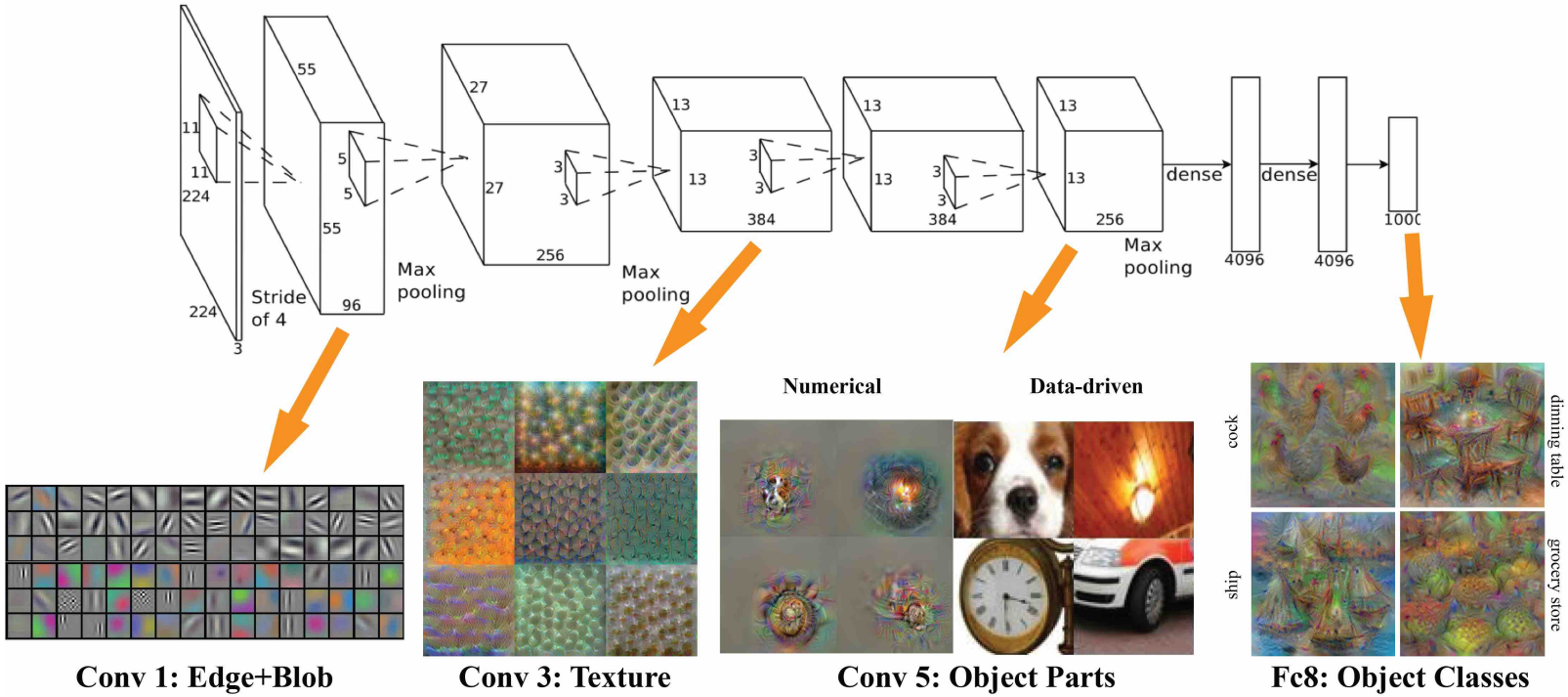
## ILSVRC: 1.4m, 1k object classes

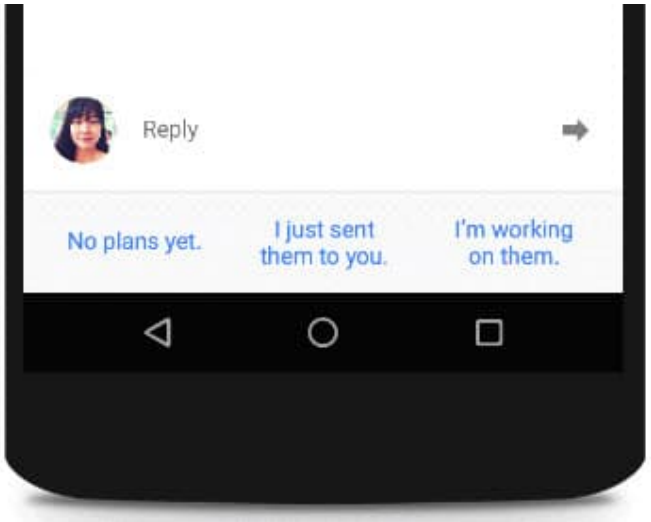


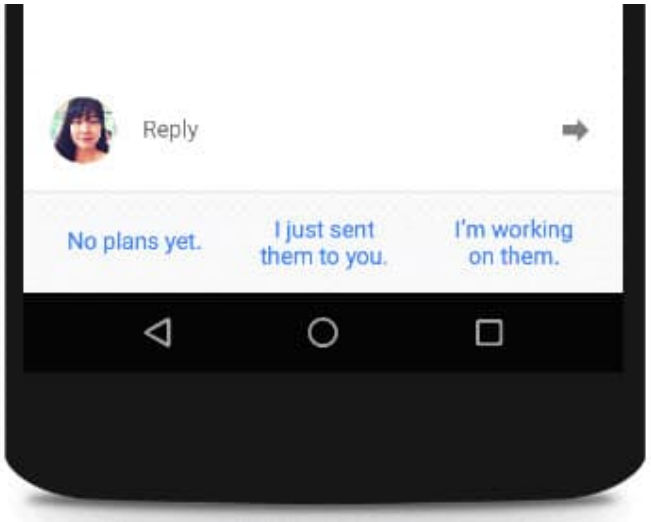
# AlexNet

## Krizhevsky, Sutskever & Hinton (2012)

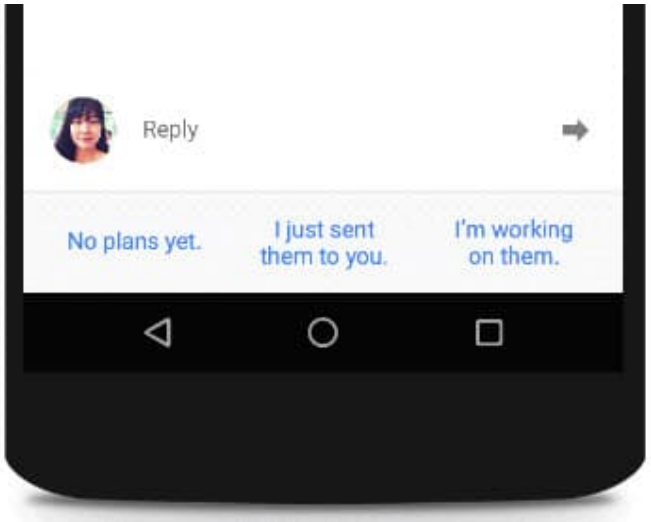











# Interactive Visualization of an Artificial Neural Network

[TensorFlow Playground]

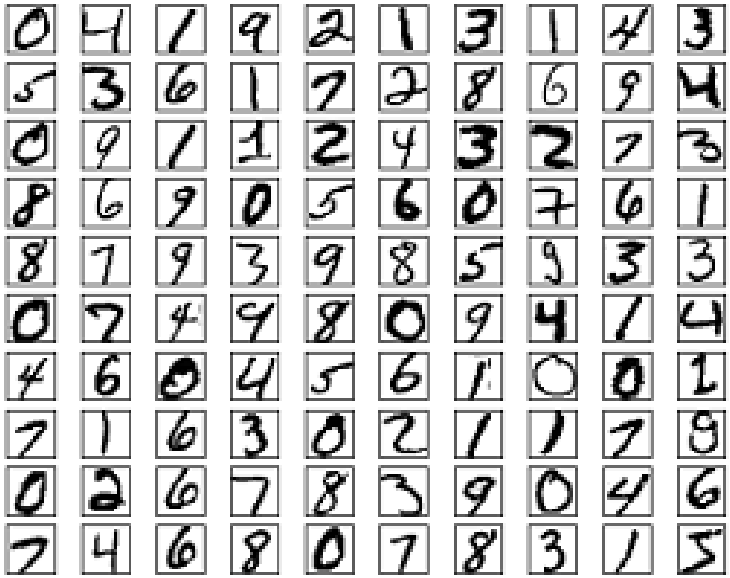
# Jupyter Notebooks

## + Docker + Nvidia GPU + TensorFlow

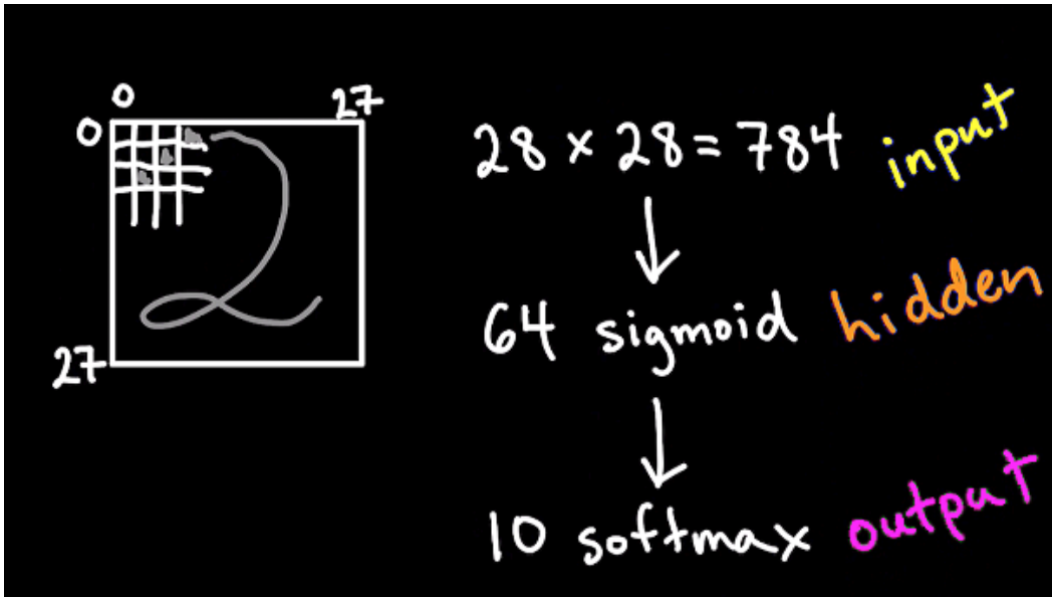
[Dockerfile]

[notebook server]

# A Shallow Neural Network



# A Shallow Neural Network

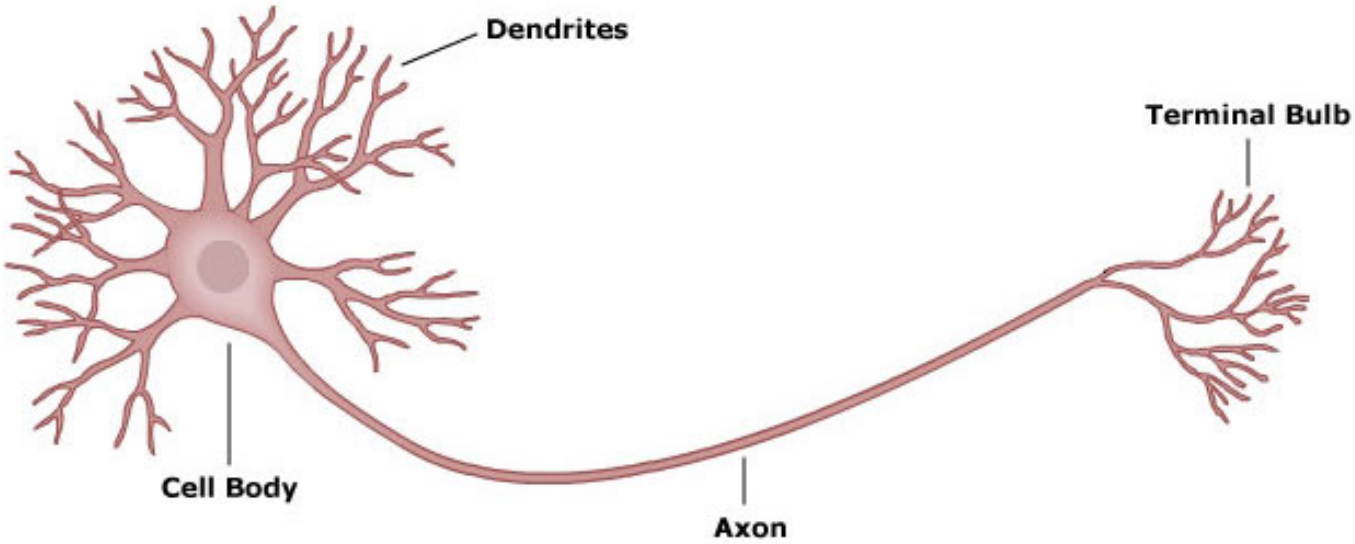



[shallow notebook]

# Outline

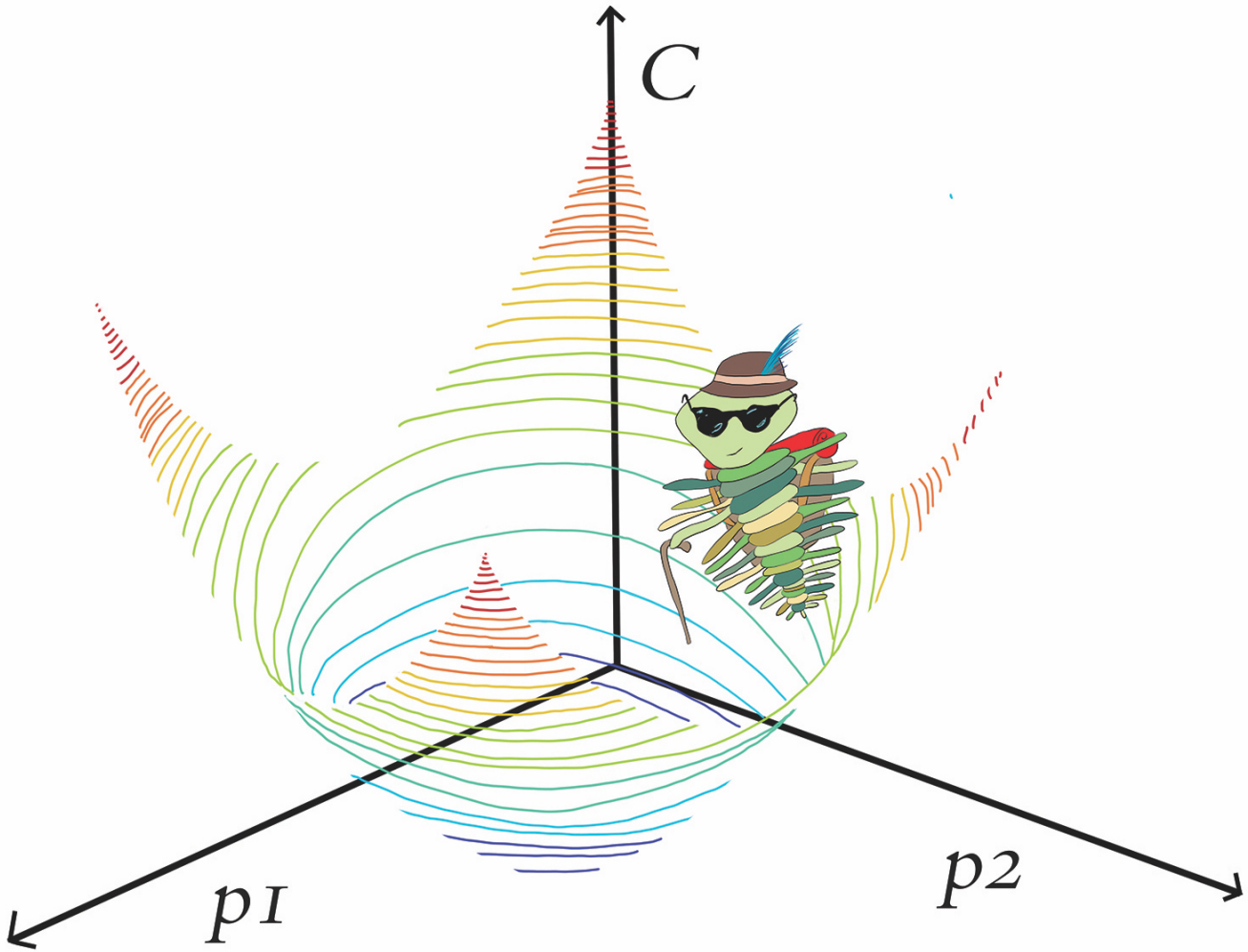
# Essential Theory I

## Neural Units



# Essential Theory II

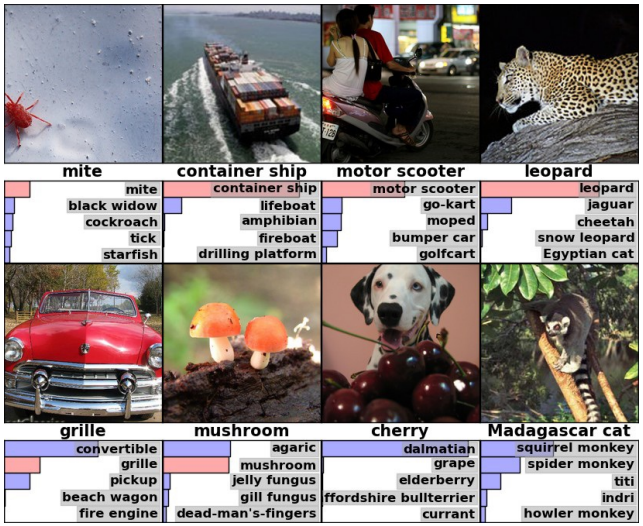
## Cost Functions, Gradient Descent, and Backpropagation

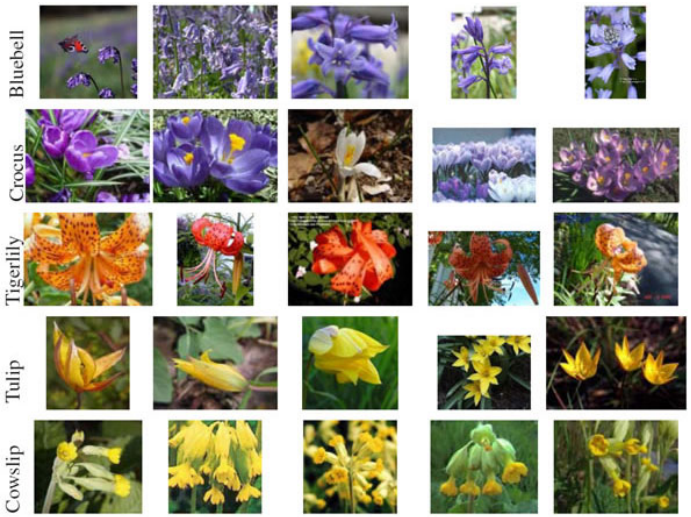


# An Intermediate Neural Network

[intermediate notebook]

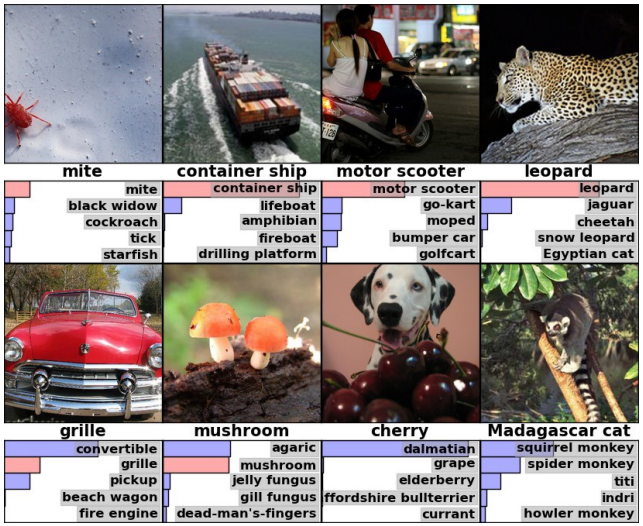
# Data Sets for Deep Learning

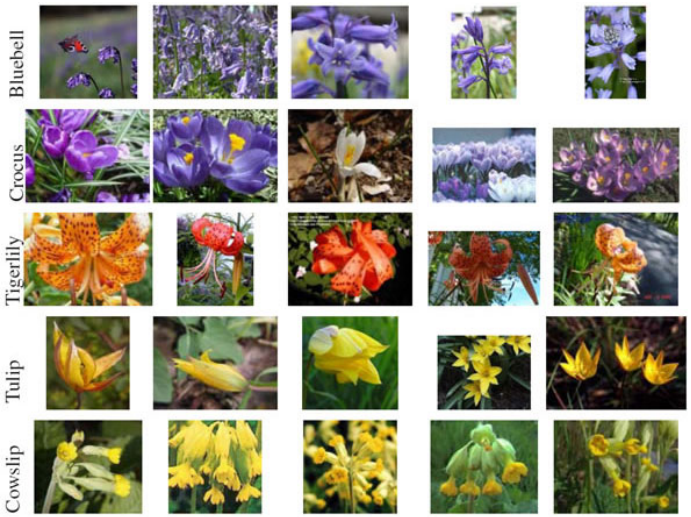



| Dataset | Classes | Train Samples |
|---|---|---|
| AG's News | 4 | 120,000 |
| Sogou News | 5 | 450,000 |
| DBPedia | 14 | 560,000 |
| Yelp Review Polarity | 2 | 560,000 |
| Yelp Review Full | 5 | 650,000 |
| Yahoo! Answers | 10 | 1,400,000 |
| Amazon Review Full | 5 | 3,000,000 |
| Amazon Review Polarity | 2 | 3,600,000 |

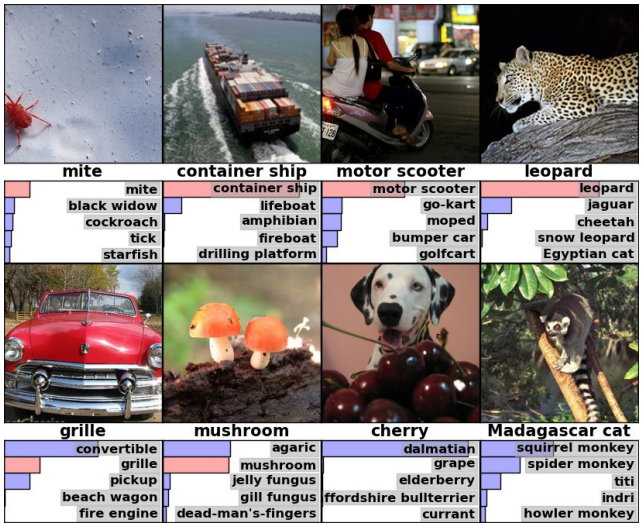
# Data Sets for Deep Learning

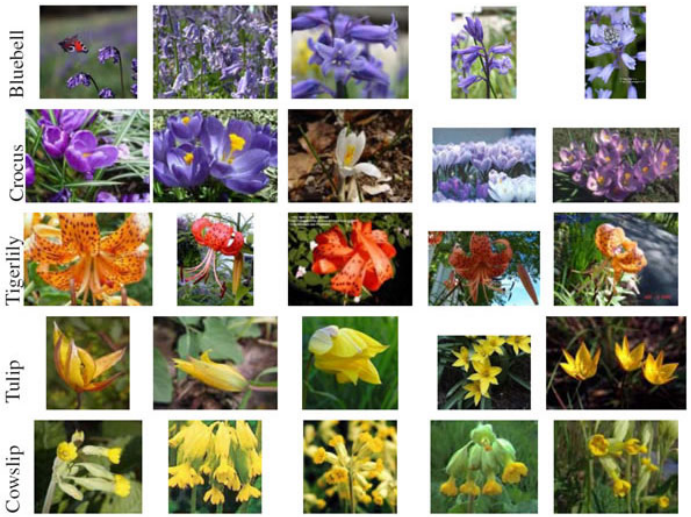



| Dataset | Classes | Train Samples |
|---|---|---|
| AG's News | 4 | 120,000 |
| Sogou News | 5 | 450,000 |
| DBPedia | 14 | 560,000 |
| Yelp Review Polarity | 2 | 560,000 |
| Yelp Review Full | 5 | 650,000 |
| Yahoo! Answers | 10 | 1,400,000 |
| Amazon Review Full | 5 | 3,000,000 |
| Amazon Review Polarity | 2 | 3,600,000 |

# Data Sets for Deep Learning




| Dataset | Classes | Train Samples |
|---|---|---|
| AG's News | 4 | 120,000 |
| Sogou News | 5 | 450,000 |
| DBPedia | 14 | 560,000 |
| Yelp Review Polarity | 2 | 560,000 |
| Yelp Review Full | 5 | 650,000 |
| Yahoo! Answers | 10 | 1,400,000 |
| Amazon Review Full | 5 | 3,000,000 |
| Amazon Review Polarity | 2 | 3,600,000 |

Data Science Resources — Jon

Secure | https://www.jonkrohn.com/resources/

Jon Krohn, Cajoler of Datums

**Home**
**Resources**
**Posts**
**Publications**
**Talks**

Academia
Applications
Quotations
Contact

**Open Data Sources**

To train a powerful model, the larger the data set, the better -- if it's well-organised and open, that's ideal. The following repositories are standouts that meet all these criteria:

- Data.gov (home of >150k US government-related datasets),
- Govcode, a collection of government open source projects,
- the Open Data Stack Exchange, and
- this curated list of 'awesome' public datasets
- this well-annotated list of data sets for natural language processing
- for biomedical and health data specifically, check out:
    - this University of Minnesota resource
    - this Medical Data for Machine Learning GitHub repo

For machine learning models that require *a lot* of *labelled* data, check out:

- UC Irvine's repository
- Yahoo's massive 13TB data set comprised of 100 billion user interactions with news items
- Google's image and video data sets
- Luke de Oliveira's *Greatest Public Datasets for AI* blog post
- CrowdFlower's Data for Everyone

Finally, here are extensive pages on importing data from the Web into R, provided by CRAN and MRAN.

# Your Deep Learning Project I

## Ideating

# Outline

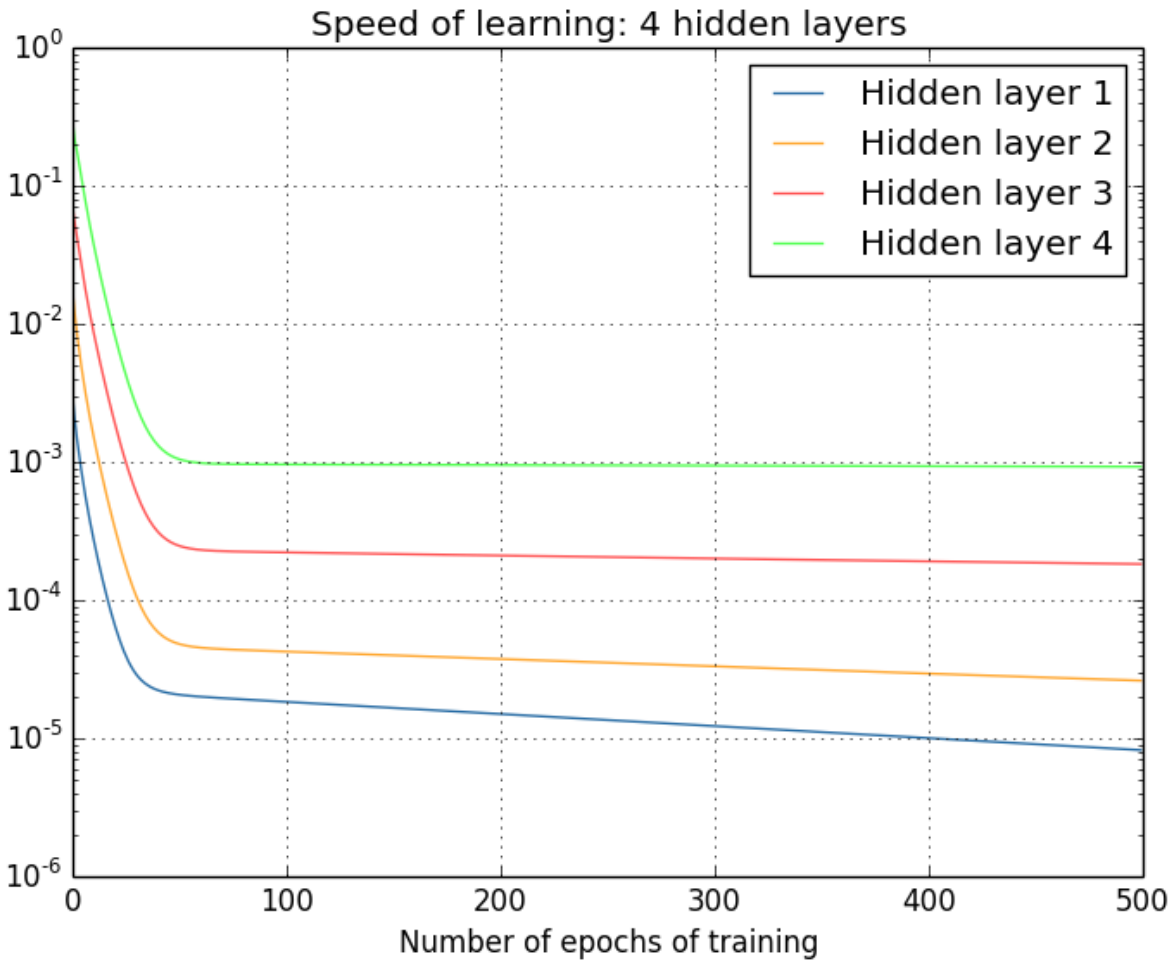
# Essential Theory III

## Weight Initialization and Mini-Batches

[neurons notebook]

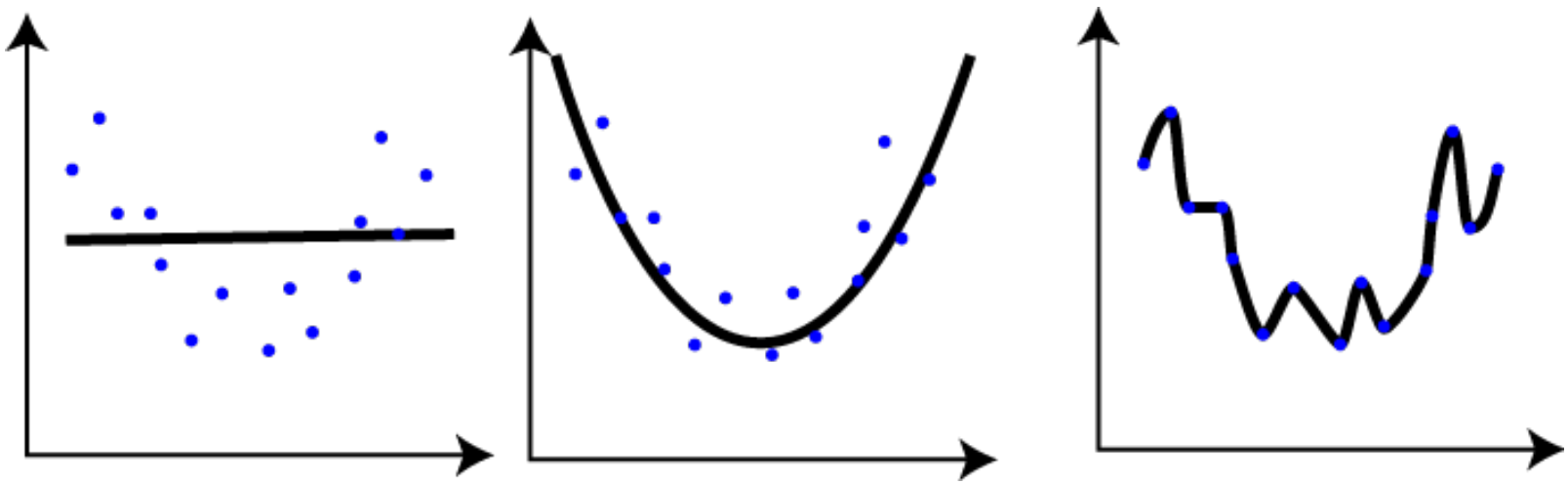
# Essential Theory IV

## Unstable Gradients and Avoiding Overfitting



# Essential Theory IV

## Unstable Gradients and Avoiding Overfitting

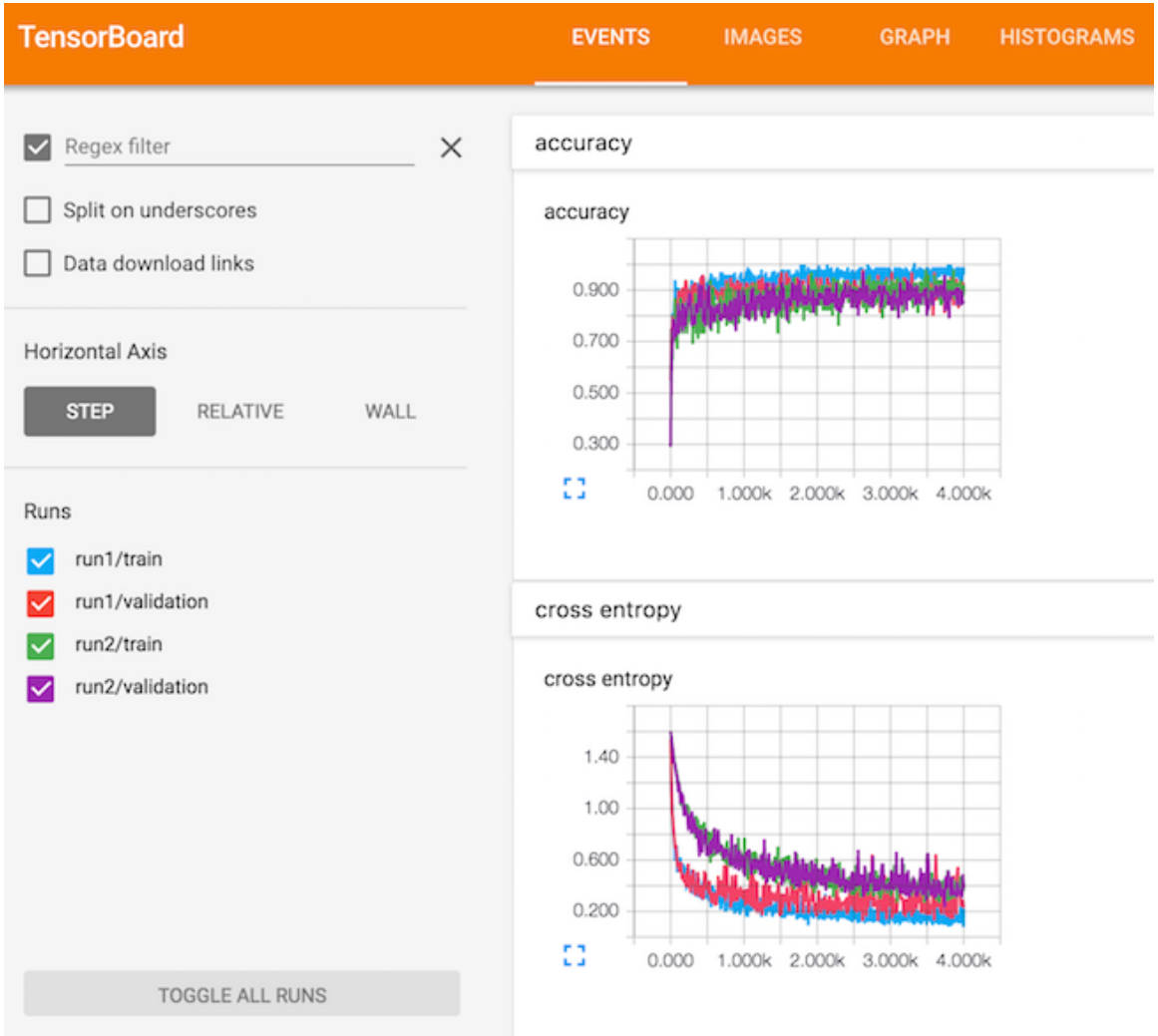


# A Deep Neural Network

[deep notebook]

# TensorBoard

## and the Interpretation of Model Outputs



# Outline

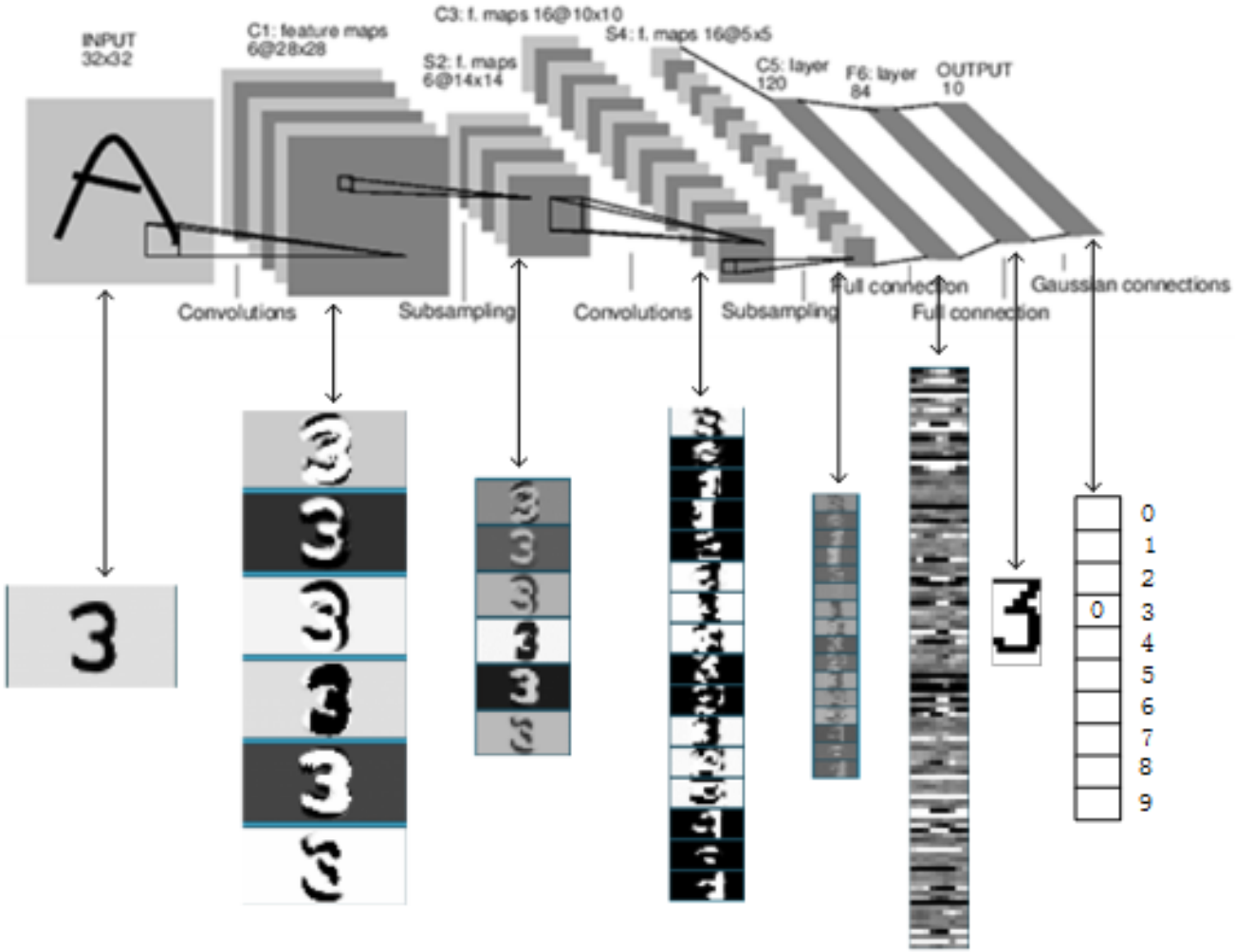
# Intro to ConvNets

## for Visual Recognition

[deepvis]

# LeNet-5

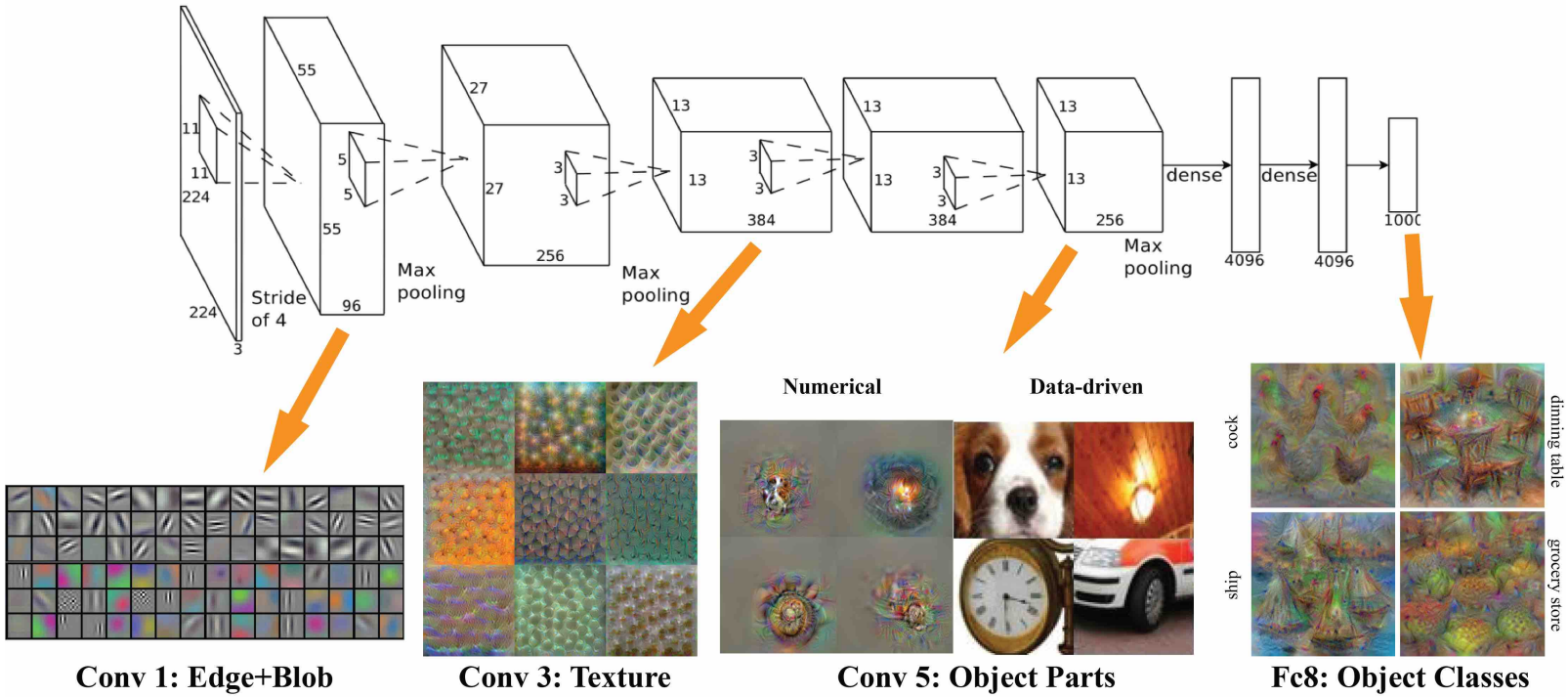
## Classic ConvNet Architecture I




[notebook]

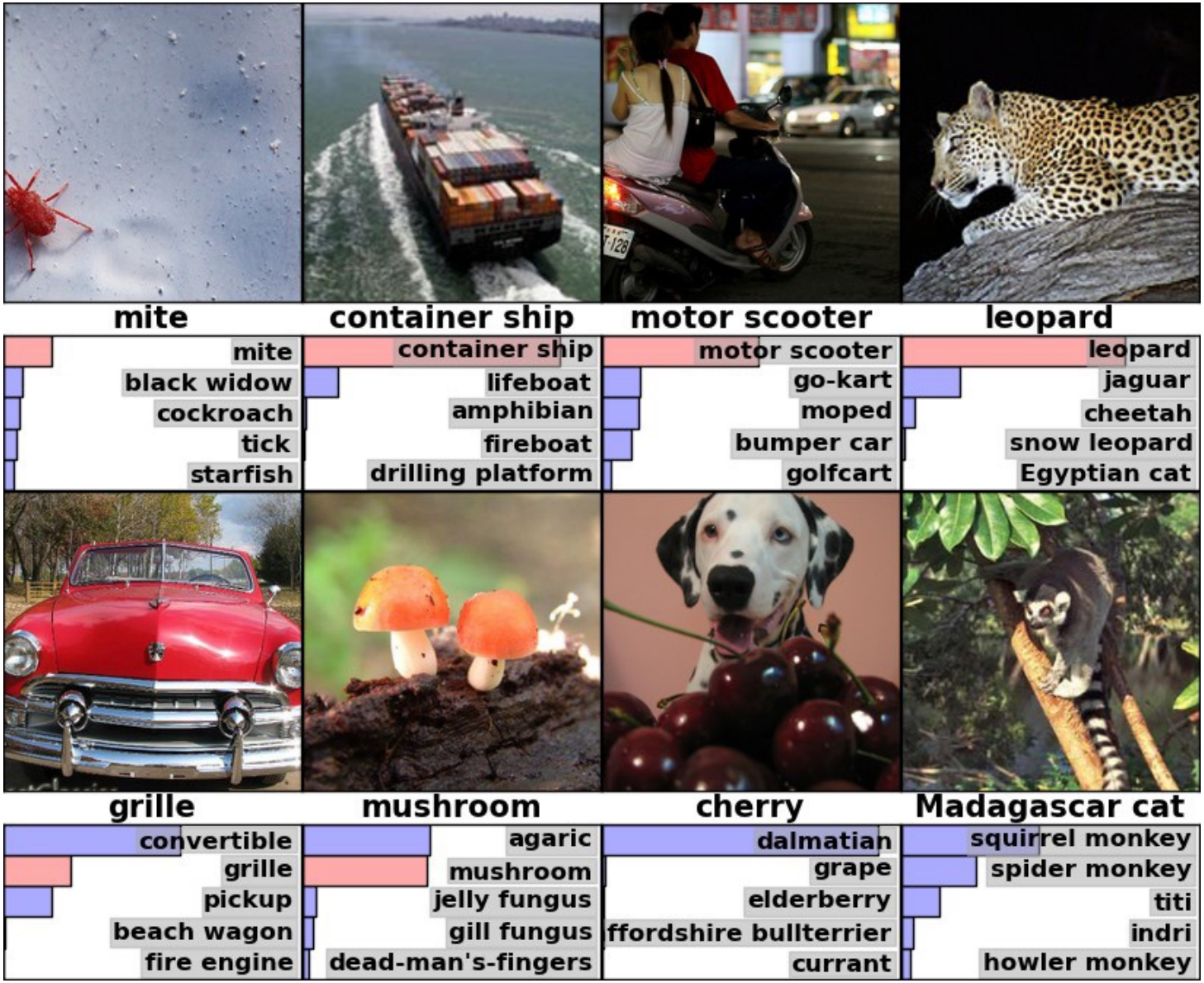
# AlexNet

## Classic ConvNet Architecture II




[notebook]

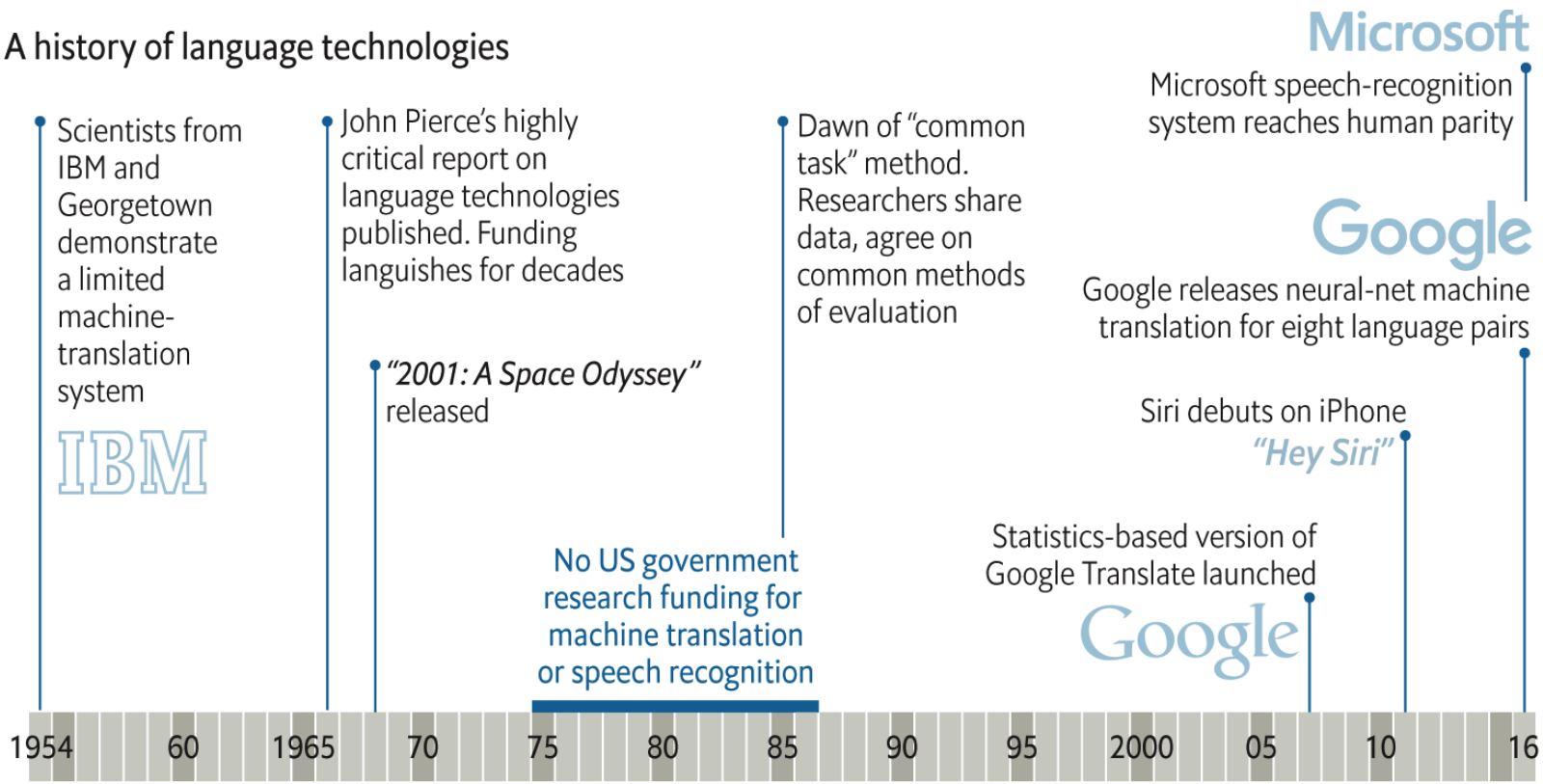
# Transfer Learning

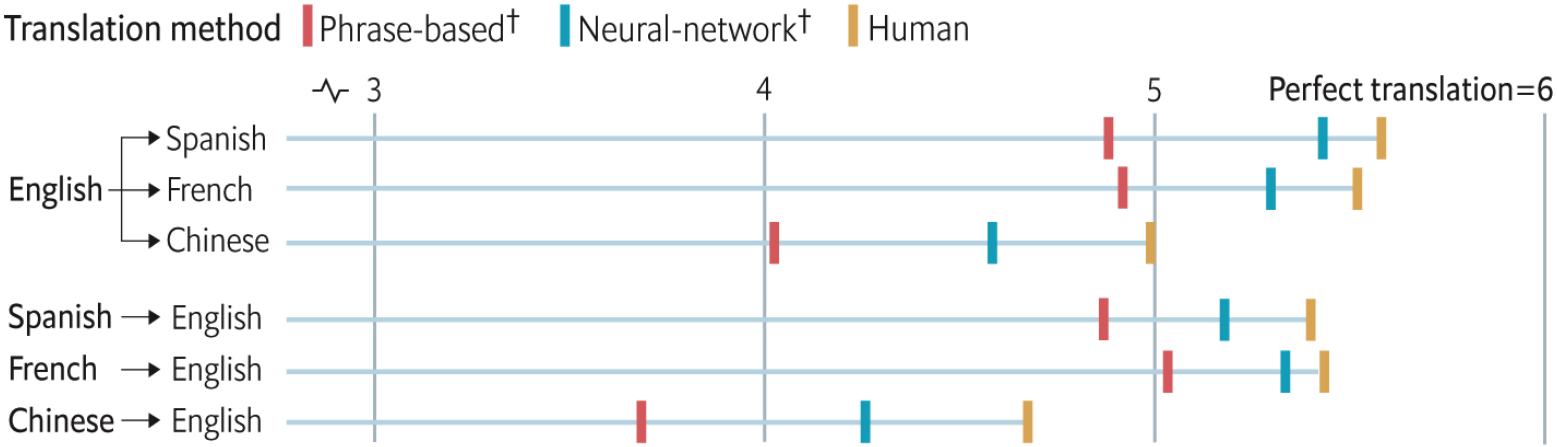


# Your Deep Learning Project II

## Formulating

# Outline


A history of language technologies
Scientists from IBM and Georgetown demonstrate a limited machine-translation system
IBM
John Pierce's highly critical report on language technologies published. Funding languishes for decades
"2001: A Space Odyssey" released
No US government research funding for machine translation or speech recognition
Dawn of "common task" method. Researchers share data, agree on common methods of evaluation
Statistics-based version of Google Translate launched
Google
Siri debuts on iPhone "Hey Siri"
Google releases neural-net machine translation for eight language pairs
Google
Microsoft speech-recognition system reaches human parity
Microsoft
1954
60
1965
70
75
80
85
90
95
2000
05
10
16


Translation method
Phrase-based†
Neural-network†
Human
3
4
5
Perfect translation=6
English
Spanish
French
Chinese
Spanish → English
French → English
Chinese → English

# Sunspring

# Sunspring

INT. SHIP

We see H pull a book from a shelf, flip through it while speaking, and then put it back.

H
In a future with mass unemployment, young people are forced to sell blood. That's the first thing I can do.

H2
You should see the boys and shut up. I was the one who was going to be a hundred years old.

H
I saw him again. The way you were sent to me... that was a big honest idea. I am not a bright light.

C
Well, I have to go to the skull. I don't know.

He picks up a light screen and fights the security force of the particles of a transmission on his face.

H
(continuing)
What do you mean?

C
(smiles)
I don't know anything about any of this.

H
(to Hauk, taking his eyes from his mouth)
Then what?

H2
There's no answer.

C
(frowning)
We're going to see the money.

H
(reading)
"All right, you can't tell me that."

Steps back. Coffey is still going through.

C
I was coming to that thing because you were so pretty.

H
I don't know. I don't know what you're talking about.

C
That's right.

H
So what are you doing?

H2
I don't want to be honest with you.

He looks at him for a moment, then smiles at him.

H
You don't have to be a doctor.

H2
I am not sure. I don't know what you're talking about.

H
I want to see you too.

H2
What do you mean?

He starts to shake.

H
I'm sorry, but I'm sure you wouldn't even touch me.

H2
I don't know what you're talking about.

H
The principle is completely constructed for the same time.

H2
(smiling)
It was all about you to be true.

H
You didn't even see the movie with the rest of the base.

H2
I don't know.

H
I don't care.

H2
I know that it's a consequence. Whatever you want to know about the presence of the story, I'm a little bit of a boy on the floor.

H
I don't know. I just have to ask you to explain to me what you say.

H2
What do you mean?

H
Because I don't know what you're talking about.

H2
That was all the time.

H
I know that.

H2
I don't know.

H
(angry)
It would be a good time. I think I could have been my life.

H (CONT'D)
It may never be forgiven, but that is just too bad. I have to leave, but I'm not free of the world.

C
Yes. Perhaps I should take it from here. I'm not going to do something.

H
You can't afford to take this anywhere. It's not a dream. But I've got a good time to stay there.

C
Well, I think you can still be back on the table.

H
Mmm. It's a damn thing scared to say. Nothing is going to be a thing but I was the one that got on this rock with a child and then I left the other two.

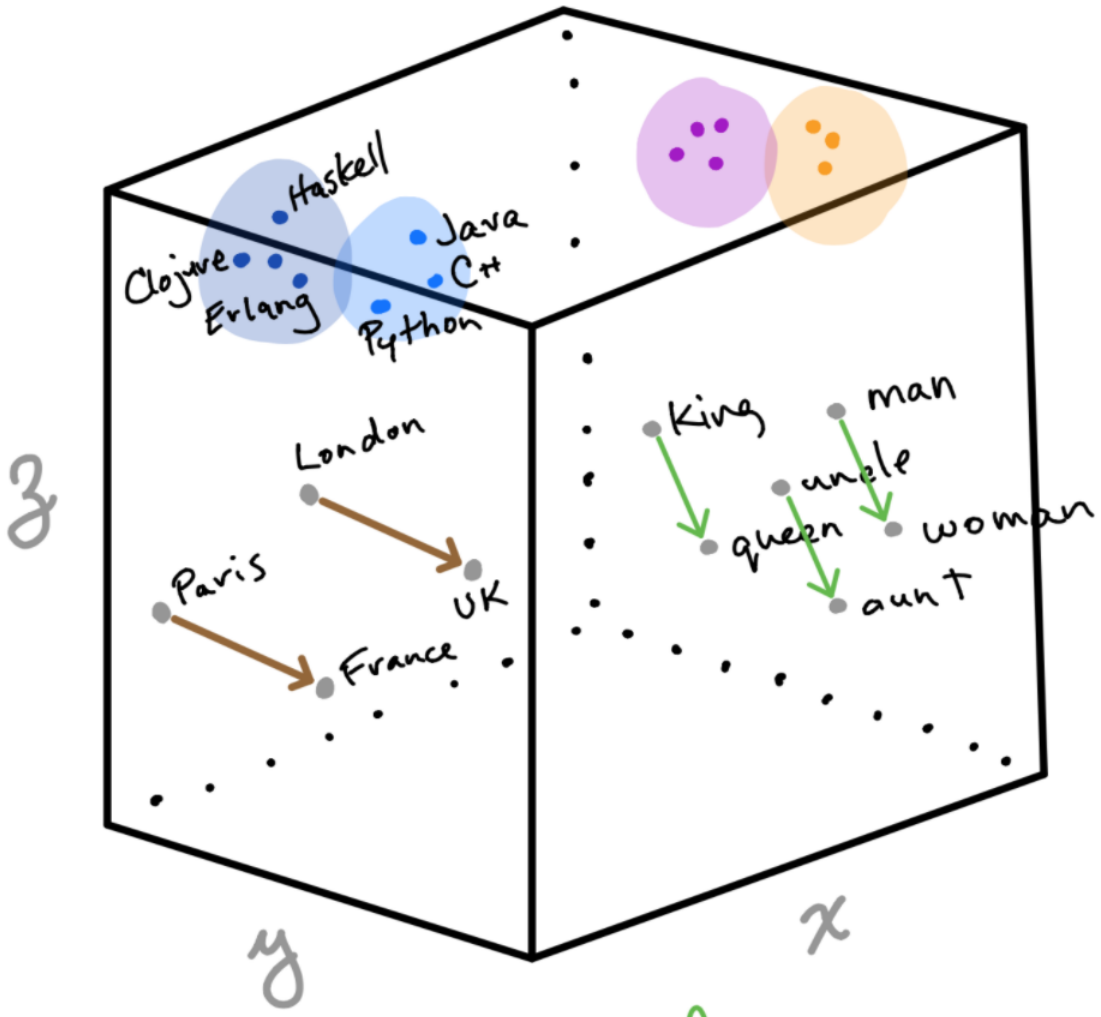
He is standing in the stars and sitting on the floor. He takes a seat on the counter and pulls the camera over to his back. He stares at it. He is on the phone. He cuts the shotgun from the edge of the room and puts it in his mouth. He sees a black hole in the floor leading to the man on the roof.

He comes up behind him to protect him. He is still standing next to him.

He looks through the door and the door closes. He looks at the bag from his backpack, and starts to cry.

T
Well, there's the situation with me and the light on the ship. The guy was trying to stop me. He was like a baby and he was gone. I was worried about him. But even if he would have done it all. He couldn't come any more. I didn't mean to be a virgin. I mean, he was weak. And I thought I'd change my mind. He was crazy to take it out. It was a long time ago. He was a little late. I was going to be a moment. I just wanted to tell you that I was much better than he did. I had to stop him and I couldn't even tell. I didn't want to hurt him. I'm sorry. I know I don't like him. I can go home and be so bad and I love him. So I can get him all the way over here and find the square and go to the game with him and she won't show up. Then I'll check it out. But I'm going to see him when he gets to me. He looks at me and he throws me out of his eyes. Then he said he'd go to bed with me.

# Vector Representations of Words



# Word Vectors

## word2vec & Vector-Space Embedding

[vse 2000]

[word2viz]

# Recurrent Neural Networks

## GRUs and LSTMs

[BiLSTM notebook]

# Non-Sequential Models

[multi-ConvNet notebook]

# Your Deep Learning Project III

## Assessing

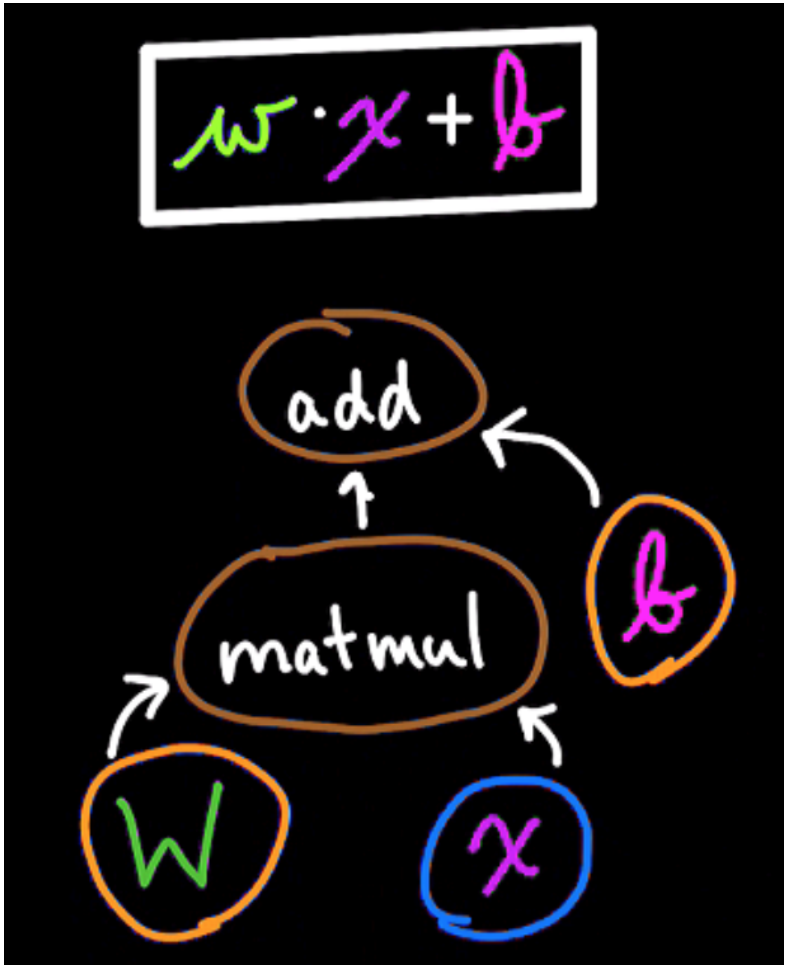
# Outline

# Leading DL Libraries

## A Comparison

| | Caffe | Torch | MXNet | TensorFlow |
|---|---|---|---|---|
| *Language* | Python, Matlab | Lua, C | Python, R, C++<br>Julia, Matlab<br>JavaScript, Go<br>Scala, Perl | Python, R, C++<br>C, Java, Go |
| *Programming Style* | Symbolic | Imperative | Imperative | Symbolic |
| *Parallel GPUs: Data* | Yes | Yes | Yes | Yes |
| *Parallel GPUs: Model* | | Yes | Yes | Yes |
| *Pre-Trained Models* | Model Zoo | ModelZoo | Model Zoo | github.com/tensorflow/models |
| *For RNNs* | | | | Best |
| *High-Level APIs* | | PyTorch | in-built | Keras, TFLearn |

# TensorFlow Graphs



# Neurons in TensorFlow

[LeNet-5 in TF]

# Outline

[LeNet-5 in TF]

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Improving Model Performance

1. Xavier Glorot initialization
2. problem simplification
3. layer architecture
4. cost function
5. avoid overfitting
6. variable learning rate $\eta$
7. epochs
8. regularization parameters, e.g., $\lambda$
9. mini-batch size
10. grid-search automation

# Tuning Hyperparameters

...in `lenet_in_keras.ipynb`:

```
model = Sequential()
model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(28, 28, 1)))
model.add(Conv2D(64, kernel_size=(3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.25))
model.add(Flatten())
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(n_classes, activation='softmax'))
```

...in `lenet_in_tensorflow.ipynb`:

```
# max pooling layer:
pool_size = 2
mp_layer_dropout = 0.25

# dense layer:
n_dense = 128
dense_layer_dropout = 0.5
```

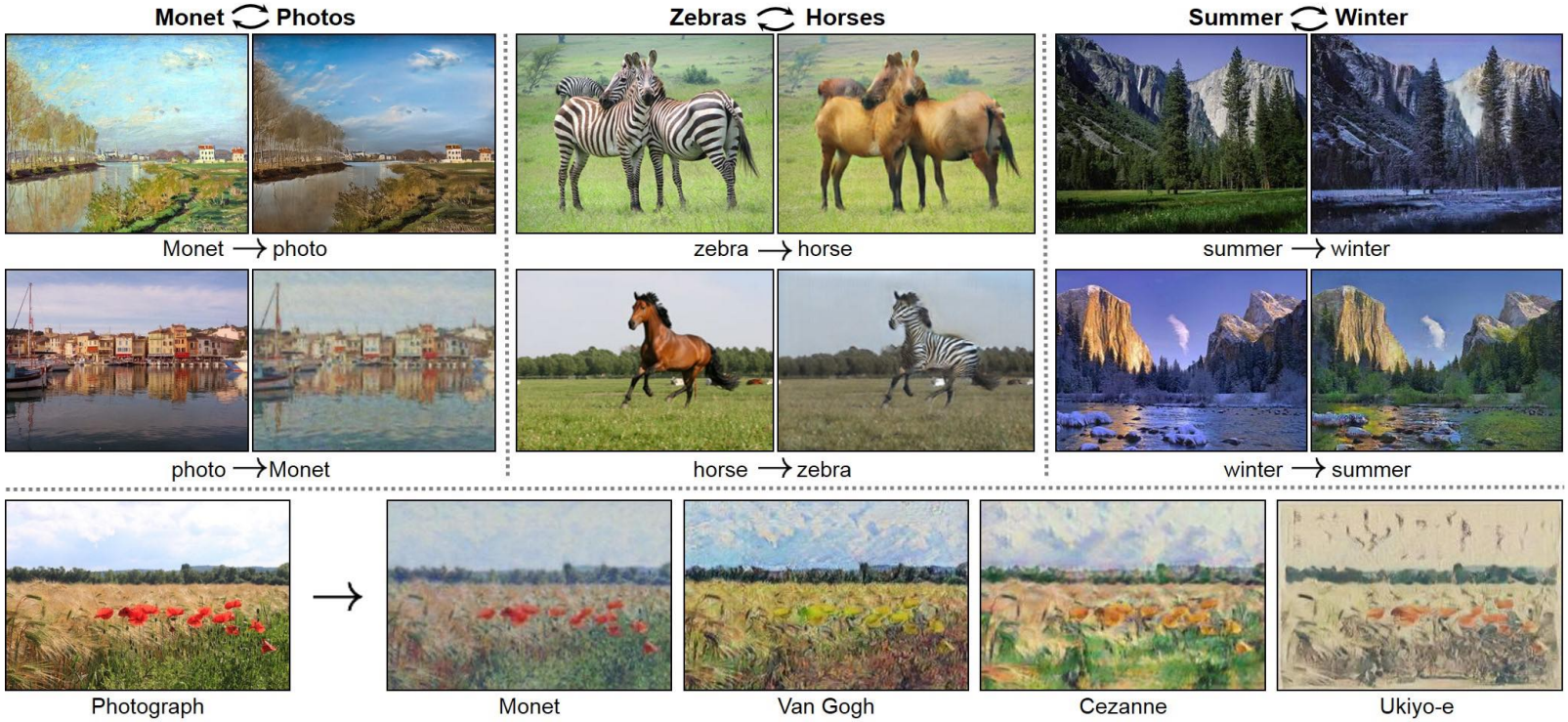
```
# convolutional and max-pooling layers:
conv_1 = conv2d(square_x, weights['W_c1'], biases['b_c1'])
conv_2 = conv2d(conv_1, weights['W_c2'], biases['b_c2'])
pool_1 = maxpooling2d(conv_2, mp_psize)
pool_1 = tf.nn.dropout(pool_1, 1-mp_dropout)

# dense layer:
flat = tf.reshape(pool_1, [-1, weights['W_d1'].get_shape().as_list()[0]])
dense_1 = dense(flat, weights['W_d1'], biases['b_d1'])
dense_1 = tf.nn.dropout(dense_1, 1-dense_dropout)
```

# Your Deep Learning Project IV

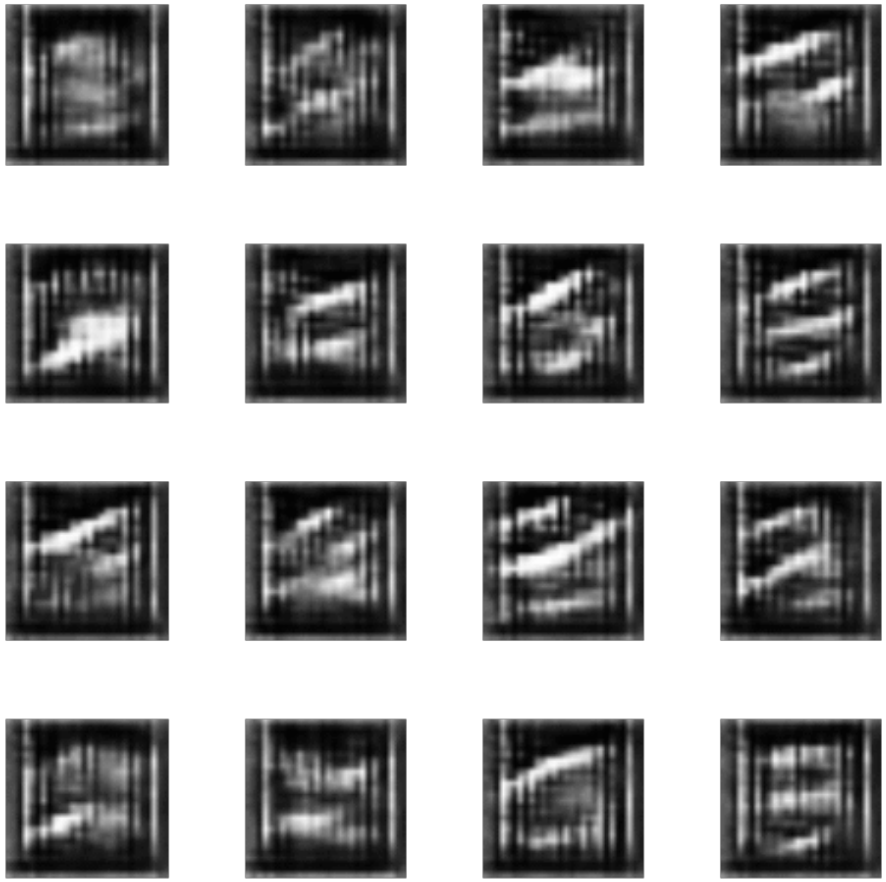
## Improving

# Outline


Monet ⇄ Photos
Zebras ⇄ Horses
Summer ⇄ Winter
Monet → photo
zebra → horse
summer → winter
photo → Monet
horse → zebra
winter → summer
Photograph
Monet
Van Gogh
Cezanne
Ukiyo-e

[Which Face is Real?]

[Quick, Draw!]



[*GAN* notebook]

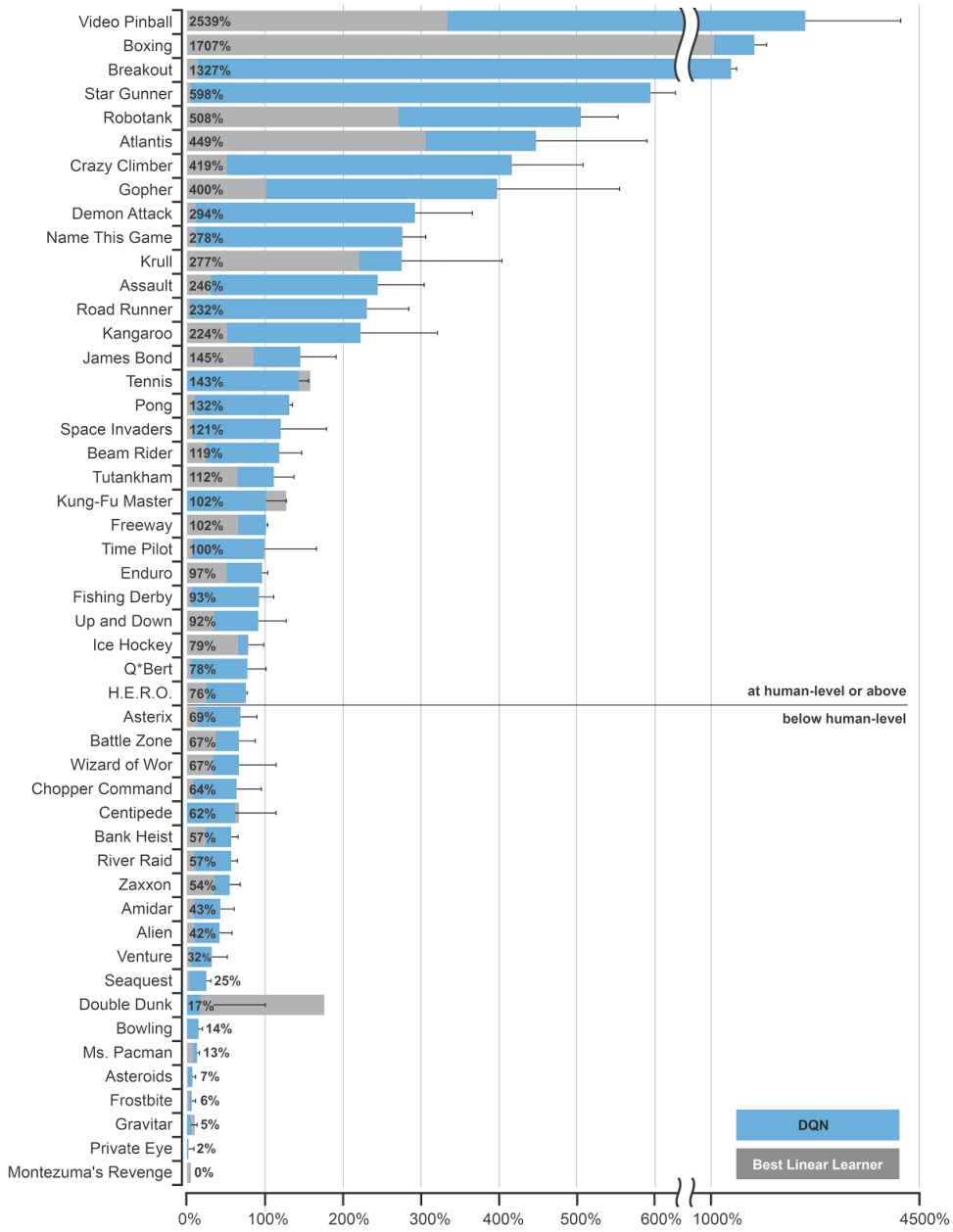
# Outline

# AlphaGO

## Silver et al. (2016)

# Deep Q-Learning

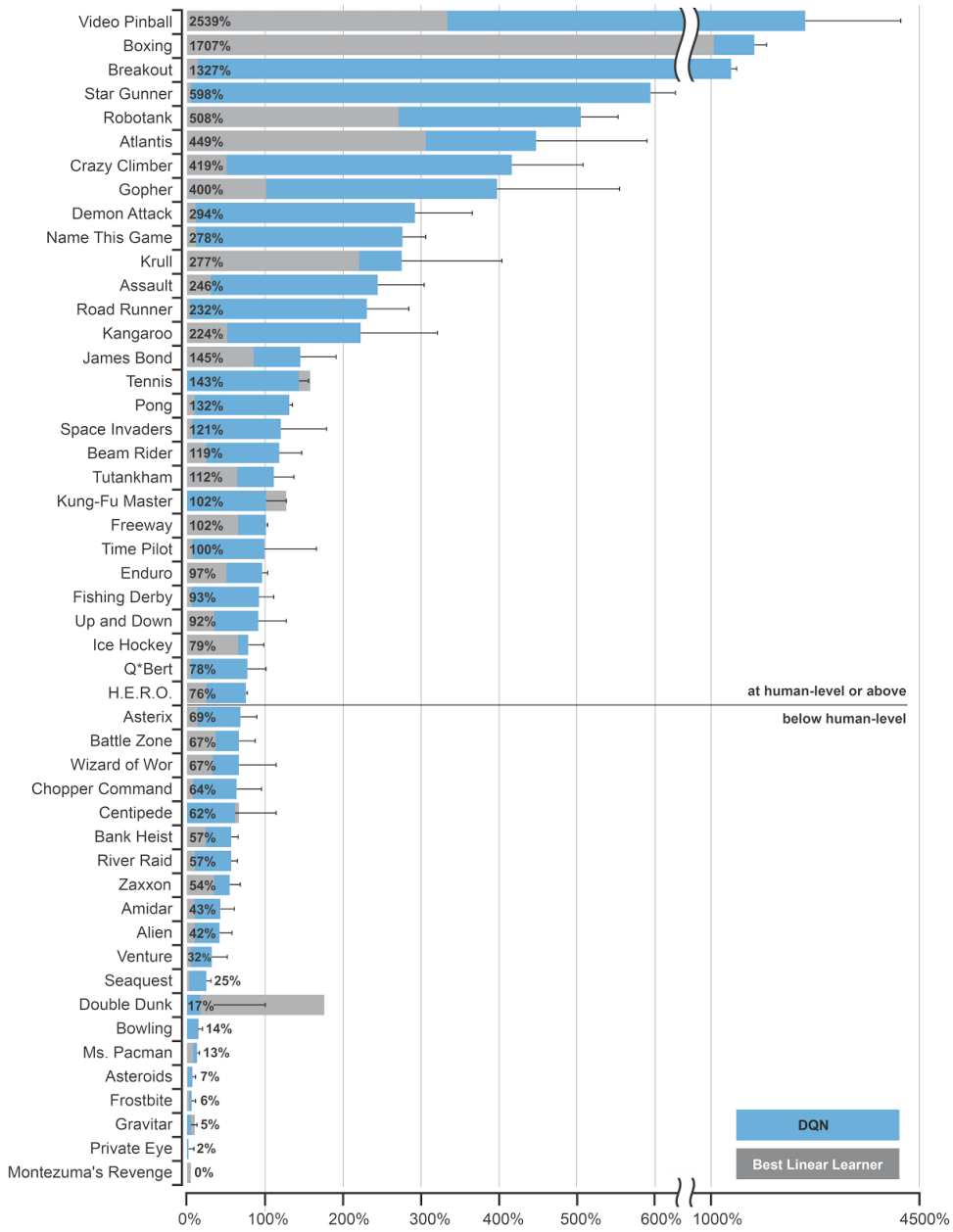
## Mnih et al. (2015)




[Atari Games]

# Deep Q-Learning

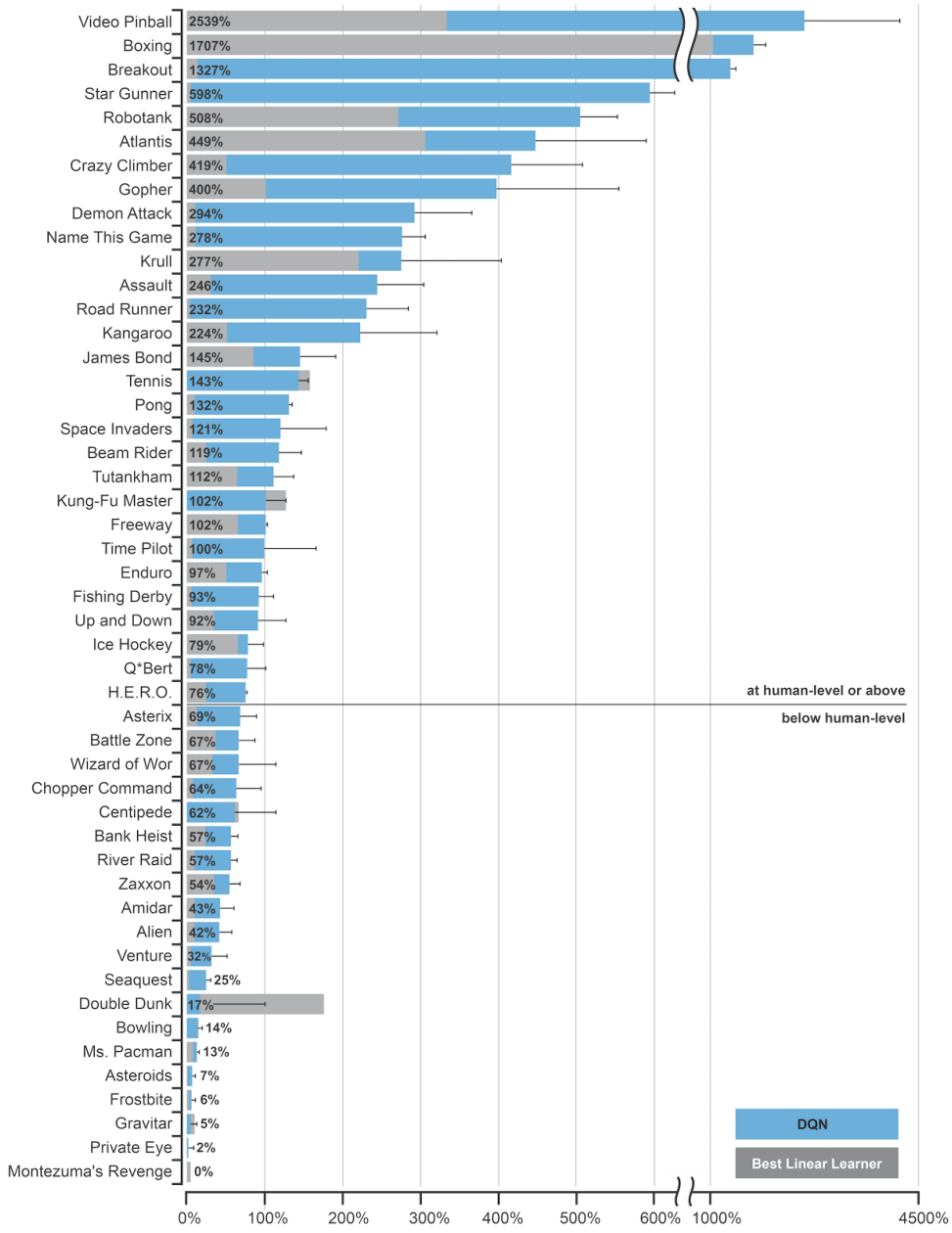
## Mnih et al. (2015)




[Atari Games]

# Deep Q-Learning

## Mnih et al. (2015)




[Atari Games]

[*Deep Q-Learning Network* notebook]

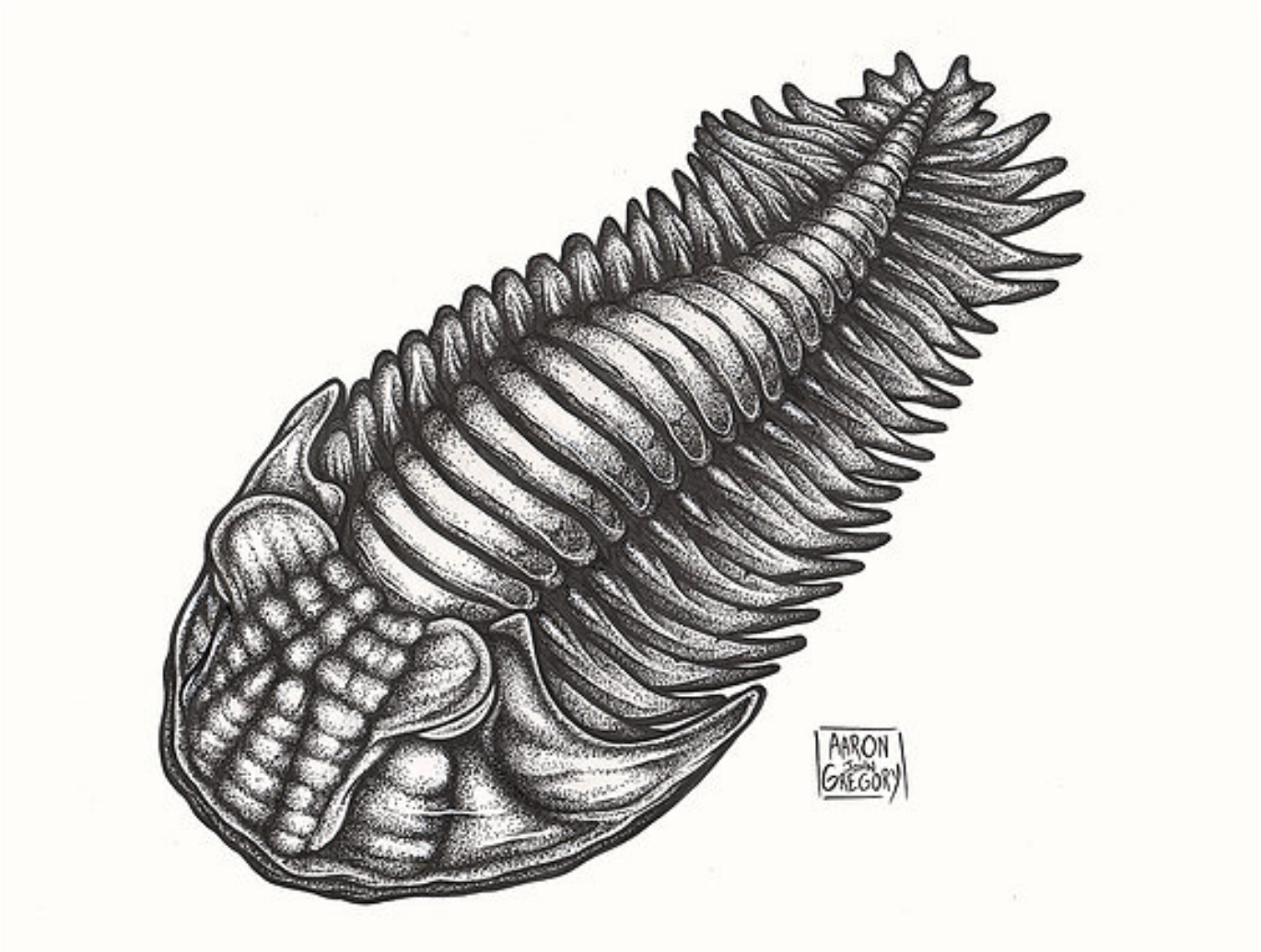
[SLM-Lab]

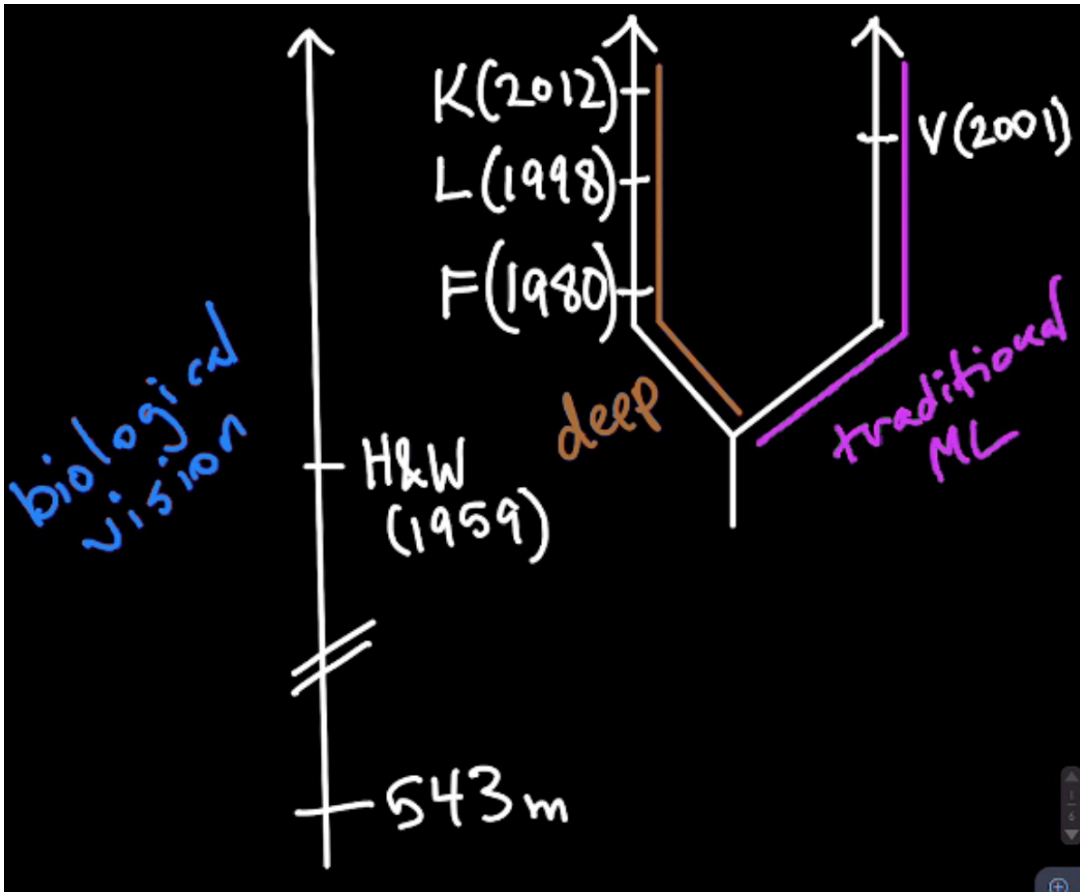
# Your Deep Learning Project V

## Presenting

# The AI Revolution

## Hasn't Even Begun



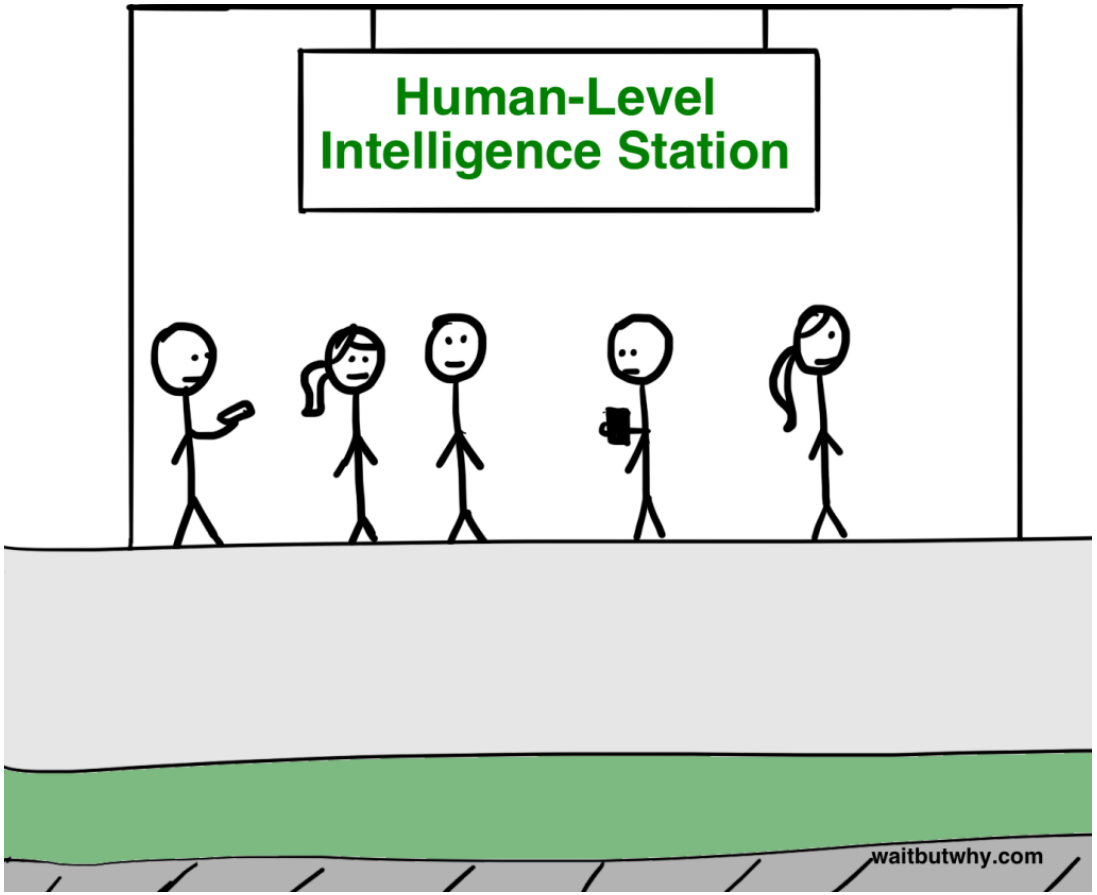


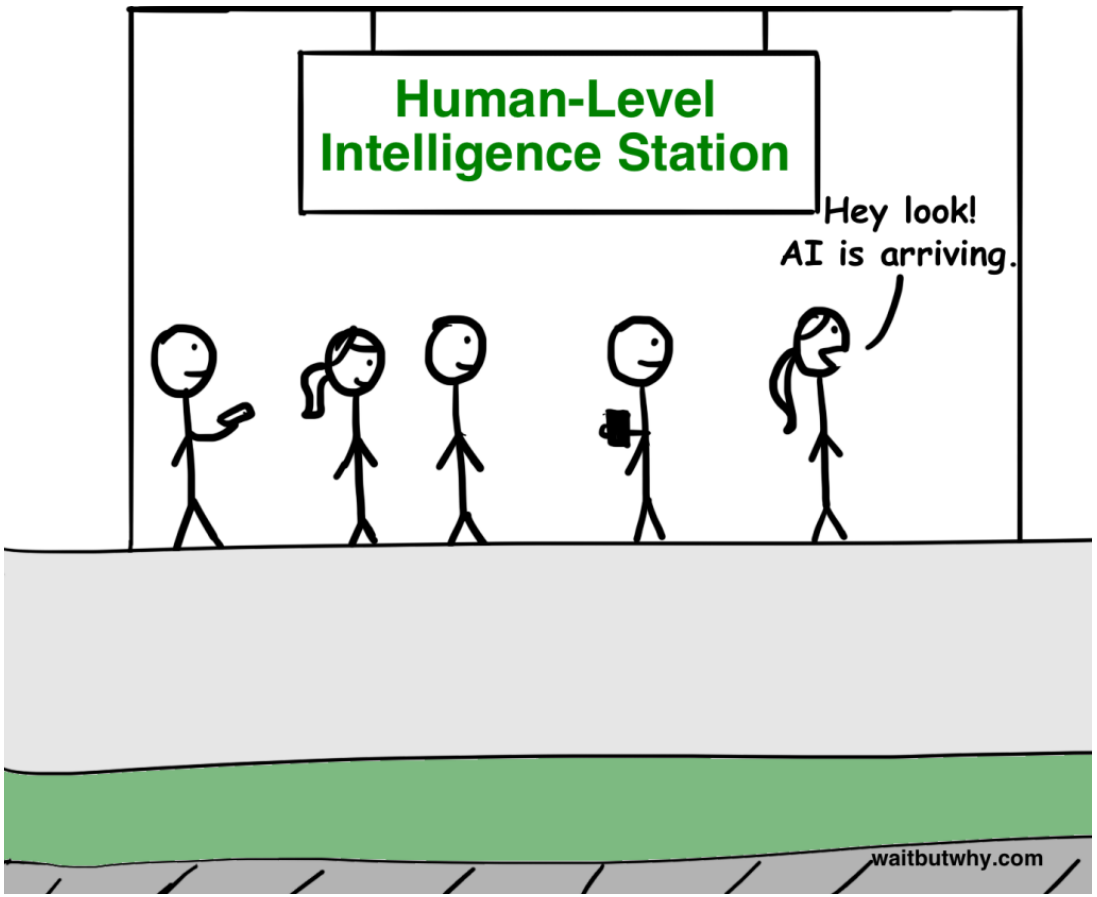
1. data doubling every 18 months
2. processing power cost halving every two years
3. cheap sensors appearing everywhere
4. Deep Learning techniques refined in academia and in industry, shared at light speed

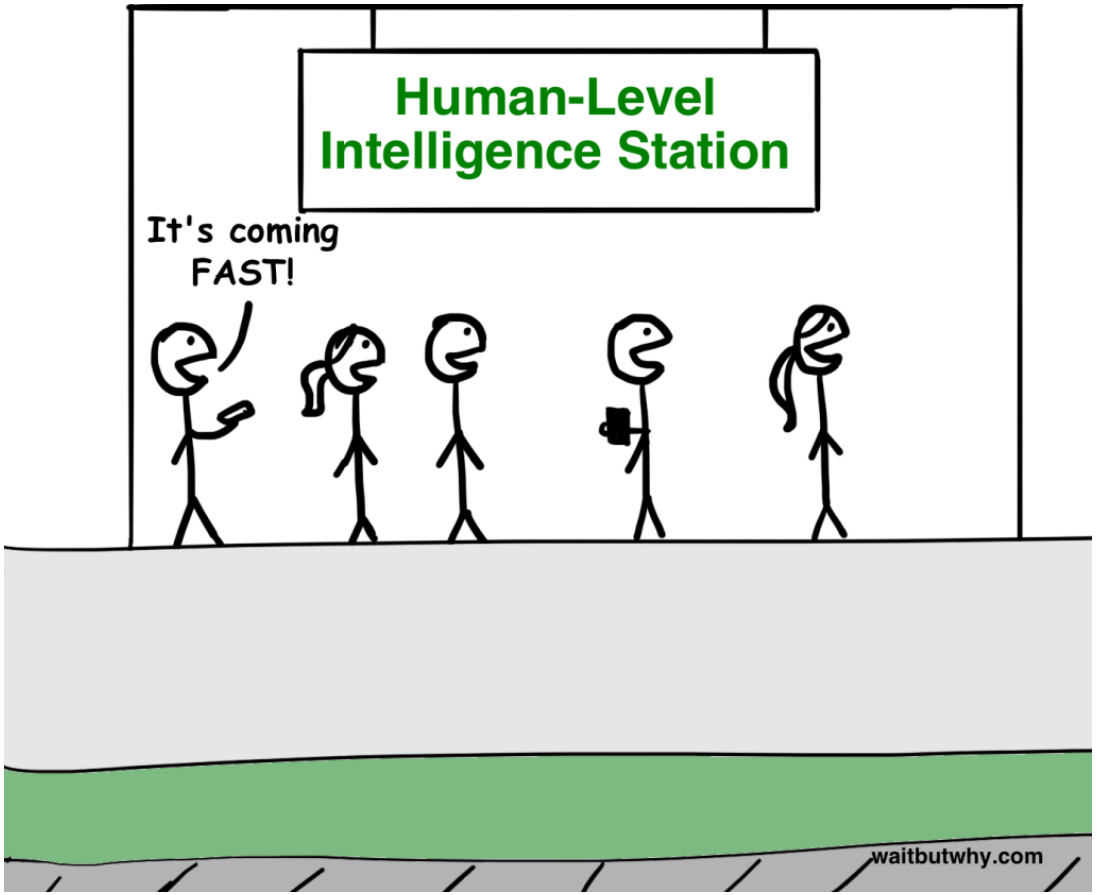
1. data doubling every 18 months
2. processing power cost halving every two years
3. cheap sensors appearing everywhere
4. Deep Learning techniques refined in academia and in industry, shared at light speed

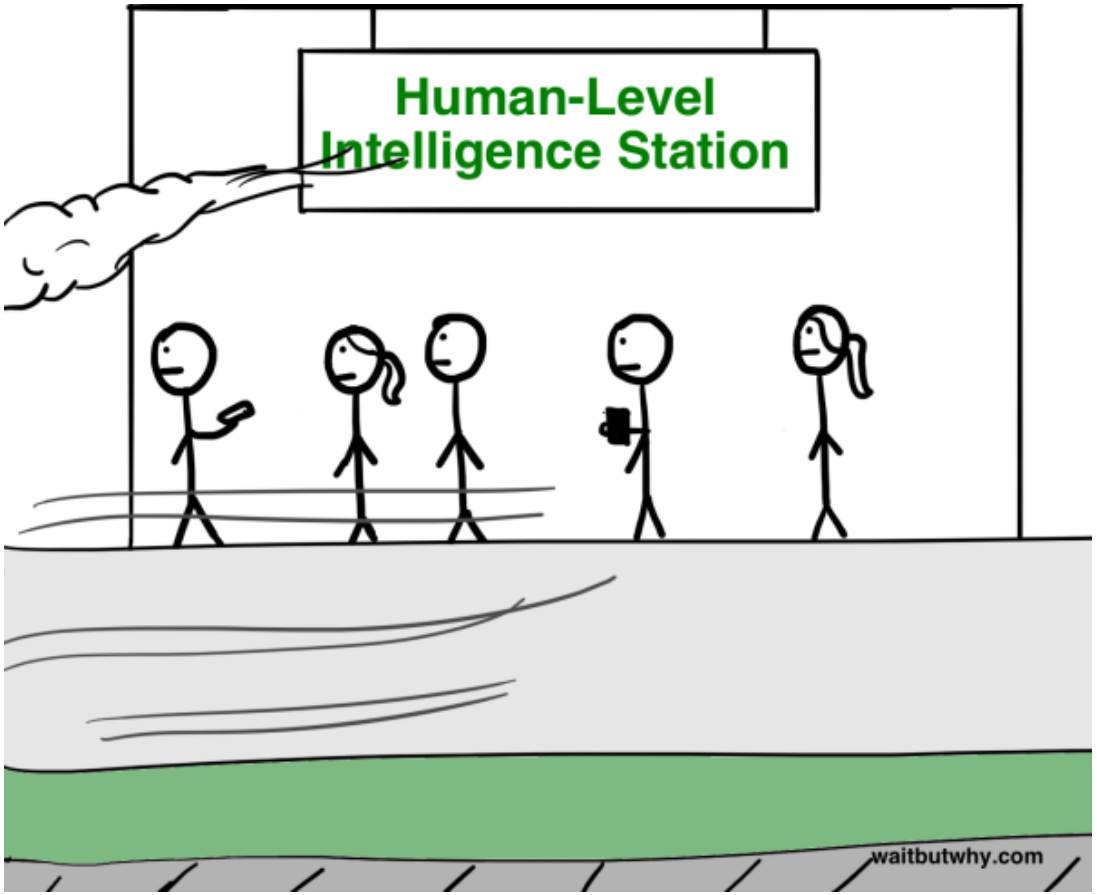
1. data doubling every 18 months
2. processing power cost halving every two years
3. cheap sensors appearing everywhere
4. Deep Learning techniques refined in academia and in industry, shared at light speed

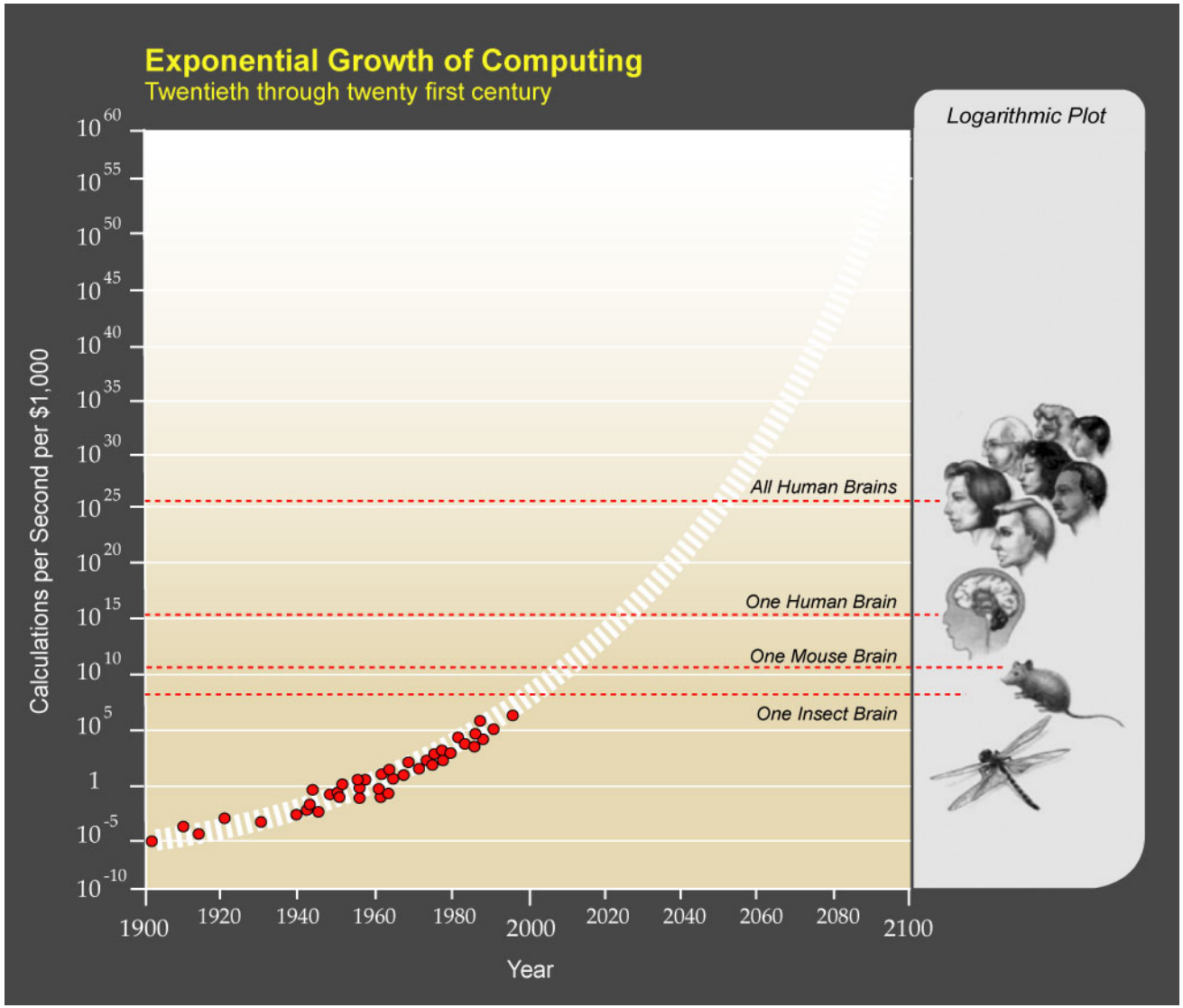
1. data doubling every 18 months
2. processing power cost halving every two years
3. cheap sensors appearing everywhere
4. Deep Learning techniques refined in academia and in industry, shared at light speed

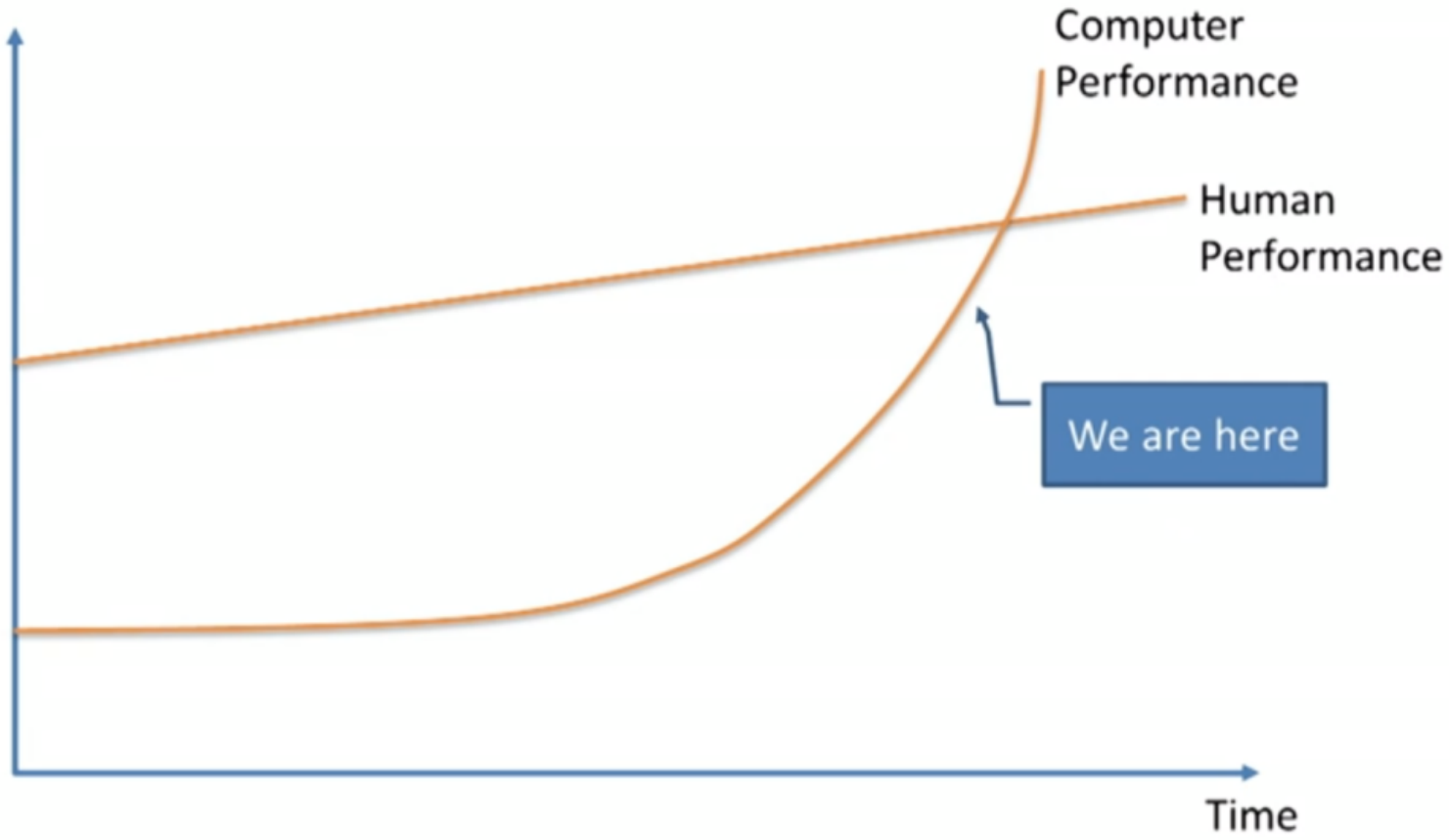


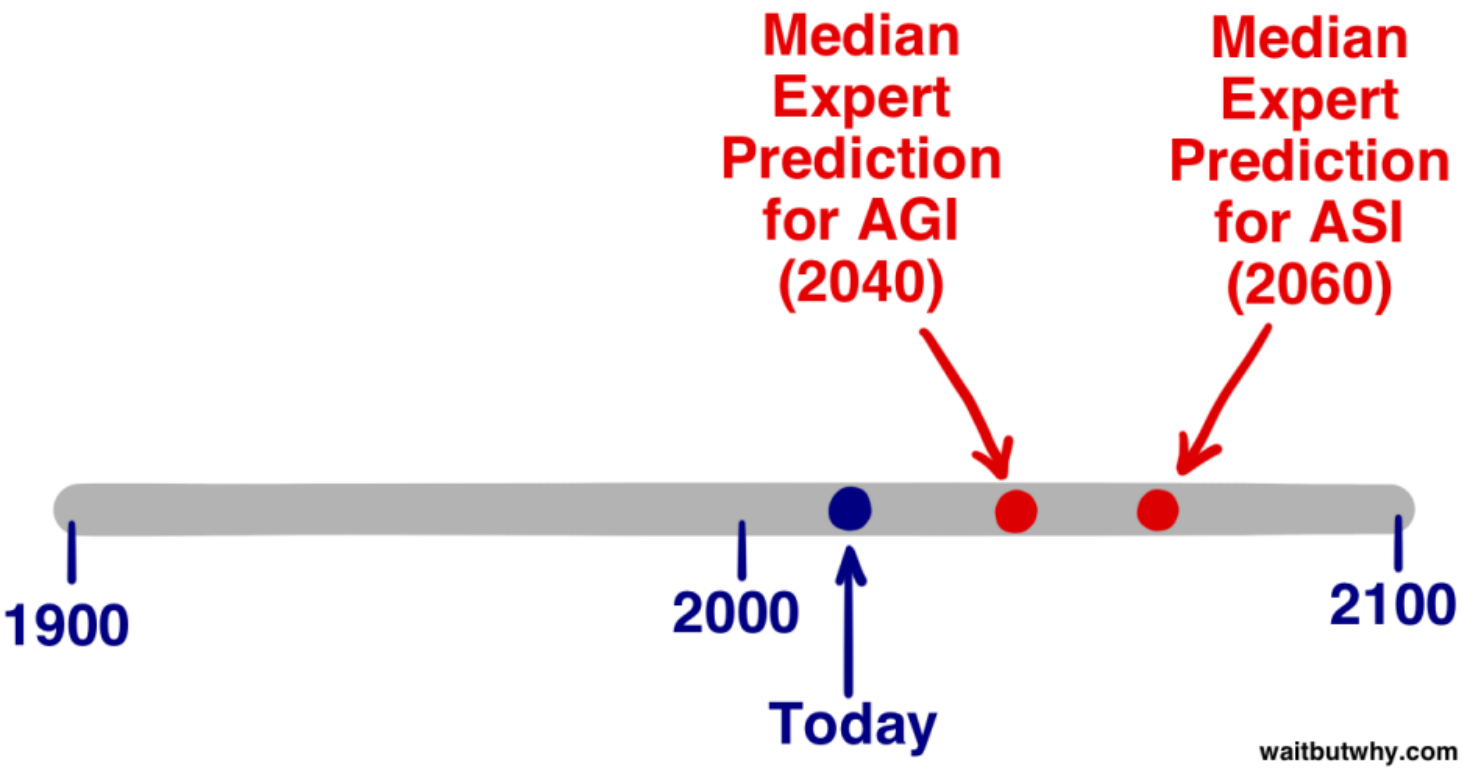

Human-Level Intelligence Station
Hey look! AI is arriving.
waitbutwhy.com

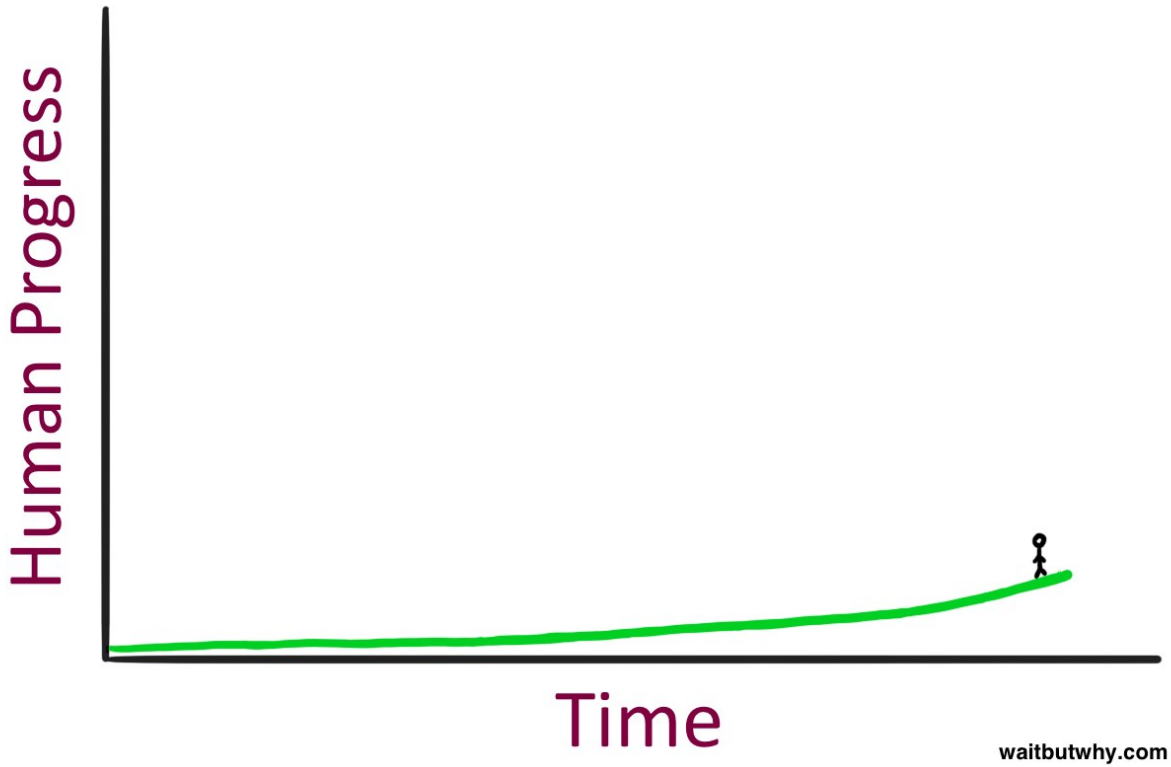


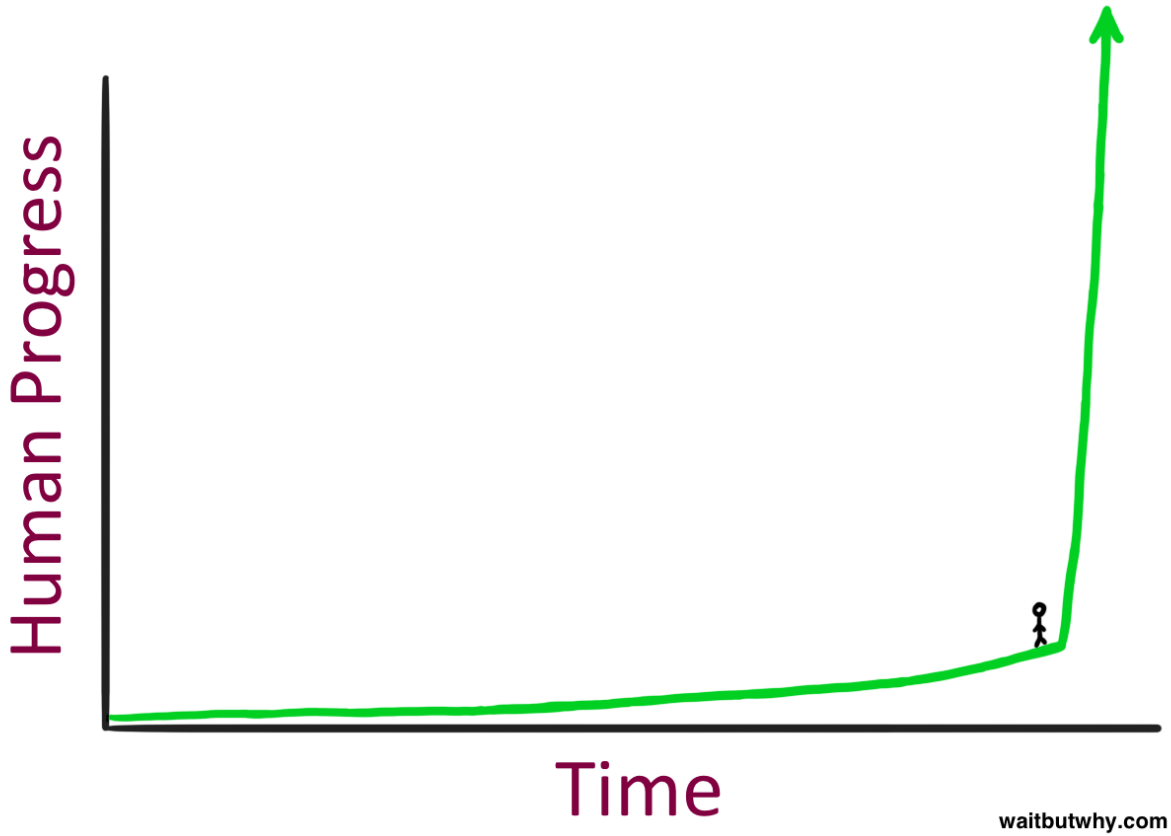

Human-Level
Intelligence Station
waitbutwhy.com








Human Progress
Time
waitbutwhy.com


Human Progress
Time
waitbutwhy.com

# Course Details

- March 16th
- March 23rd
- March 30th
- April 6th
- April 13th

See `nycdatascience.com/courses/deep-learning`

10% tuition discount with `dlt2019`
15% discount for NYCDSA Bootcampers

# Course Details

- March 16th
- March 23rd
- March 30th
- April 6th
- April 13th

See `nycdatascience.com/courses/deep-learning`

10% tuition discount with `dlt2019`
15% discount for NYCDSA Bootcampers

# Course Details

- March 16th
- March 23rd
- March 30th
- April 6th
- April 13th

See `nycdatascience.com/courses/deep-learning`

10% tuition discount with `dlt2019`
15% discount for NYCDSA Bootcampers

# Course Details

- March 16th
- March 23rd
- March 30th
- April 6th
- April 13th

See `nycdatascience.com/courses/deep-learning`

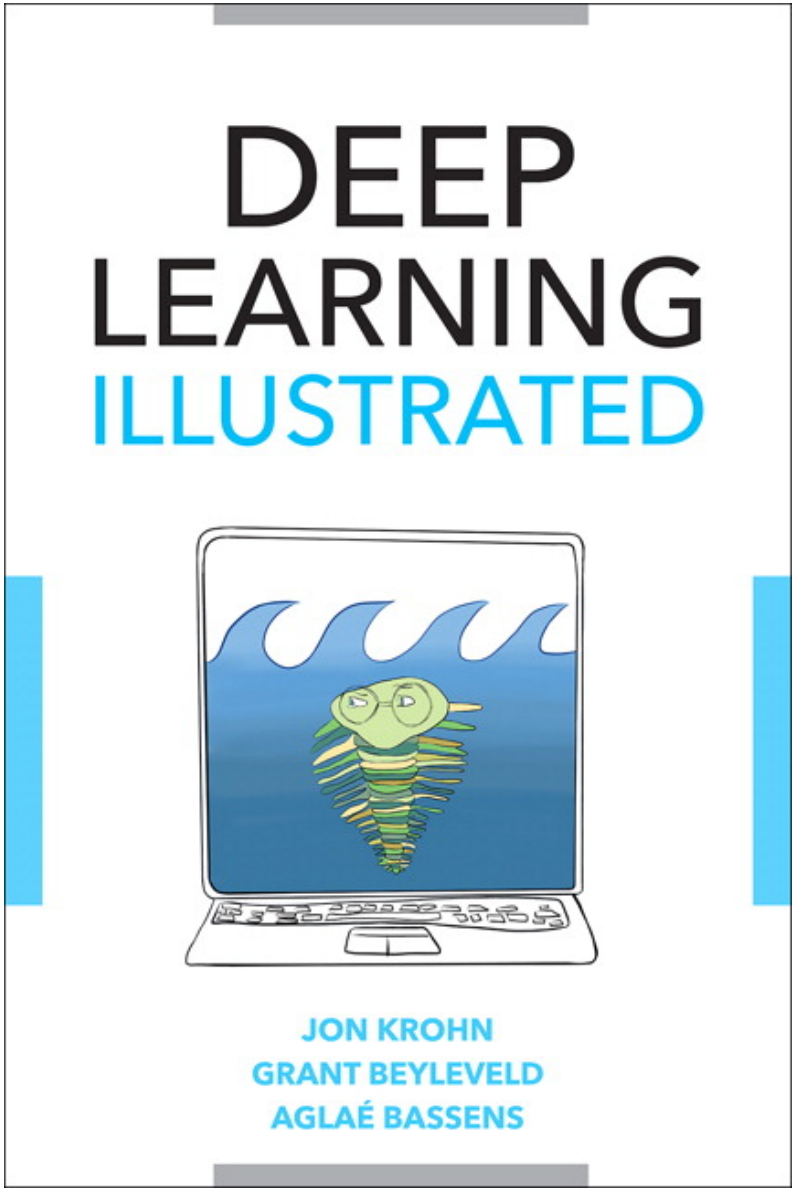
10% tuition discount with `dlt2019`
15% discount for NYCDSA Bootcampers

# Course Details

- March 16th
- March 23rd
- March 30th
- April 6th
- April 13th

See `nycdatascience.com/courses/deep-learning`

10% tuition discount with `dlt2019`
15% discount for NYCDSA Bootcampers




Available for pre-order now