

Data-driven decision making

Dirk U. Wulff, *The R Bootcamp*

GMFH @ Bern, 2019

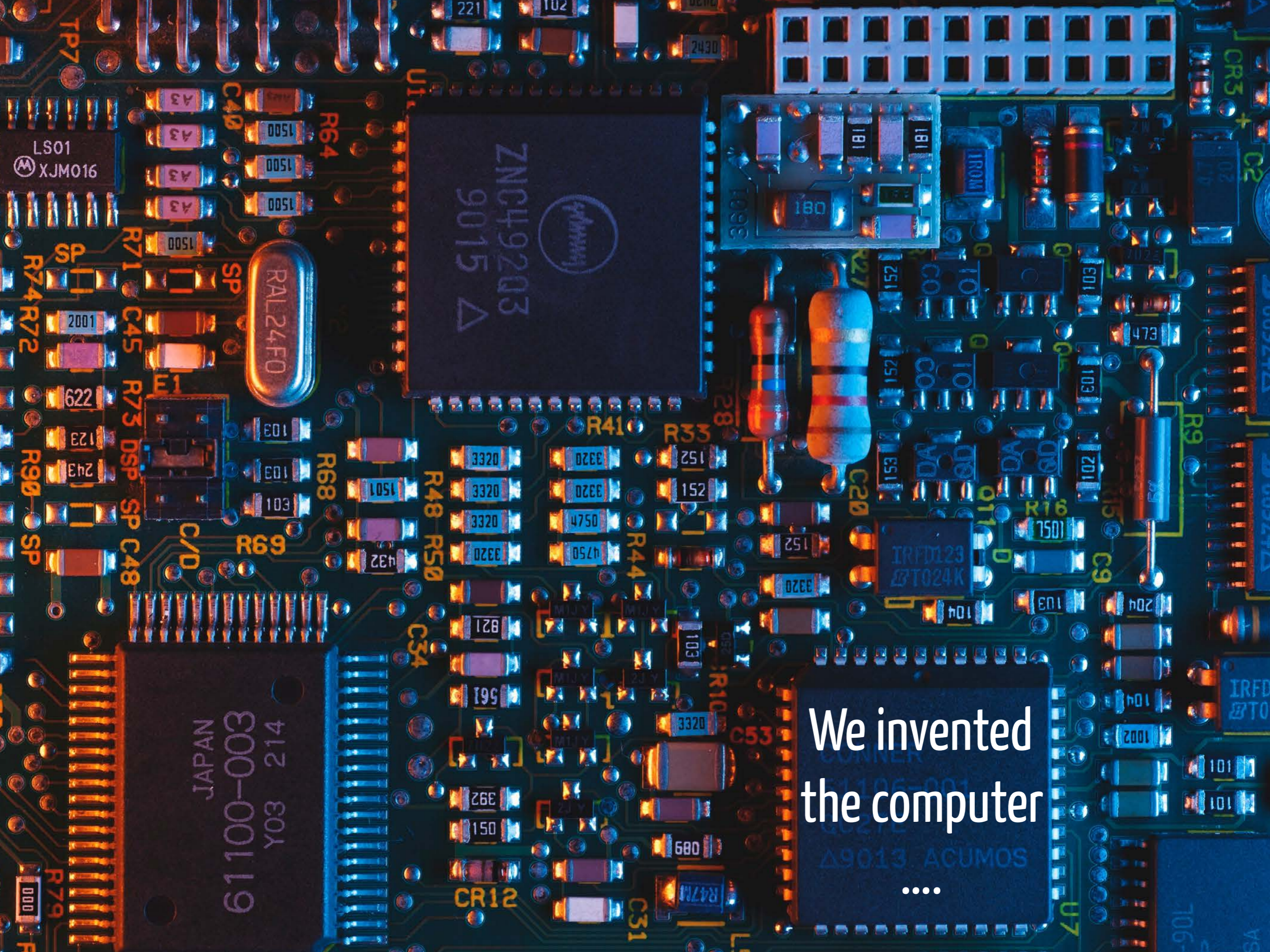


We've been to the
moon...



We learned to treat many diseases...





We invented
the computer

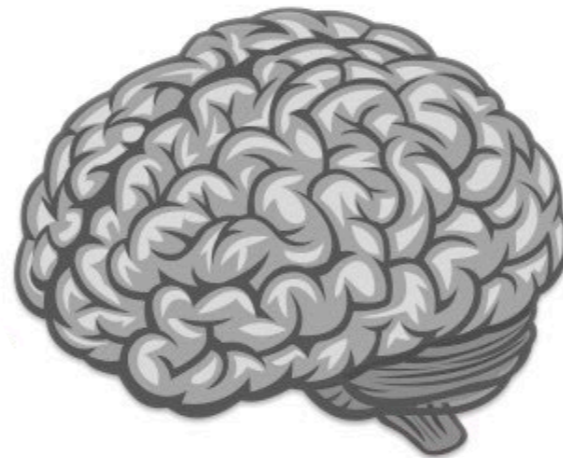
...

Still, we are kind
of stupid!



The cognitive **decision making** system

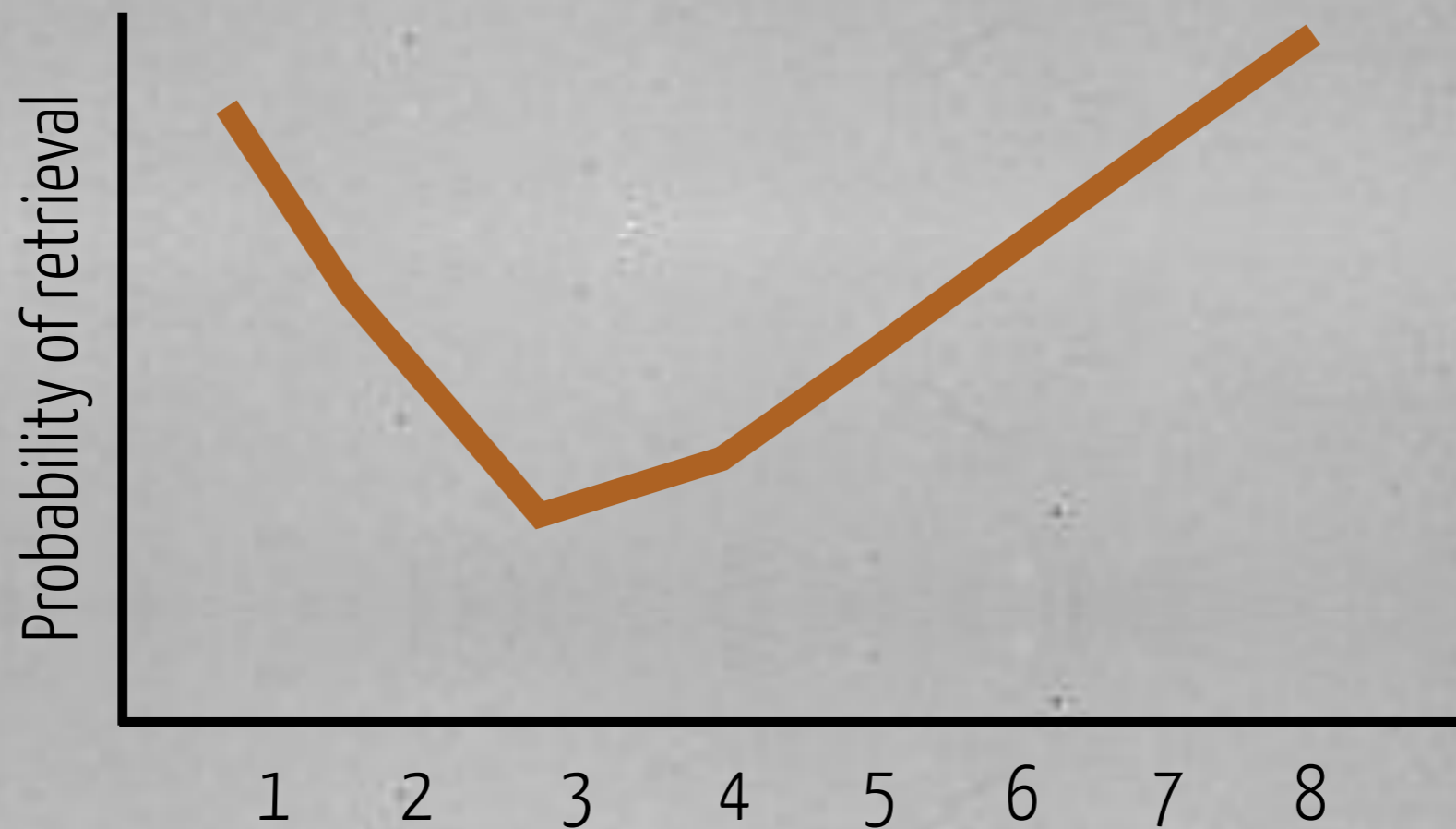
1010111010011010
0001010111101010
1011010010010101
0011101101010000
0101111010010101
0010100010101111
0101010110100100
1010010010101001
1101101010000010



Response

analysis - data - economic - process -
income - context - estimate - export

Can you recall the words?



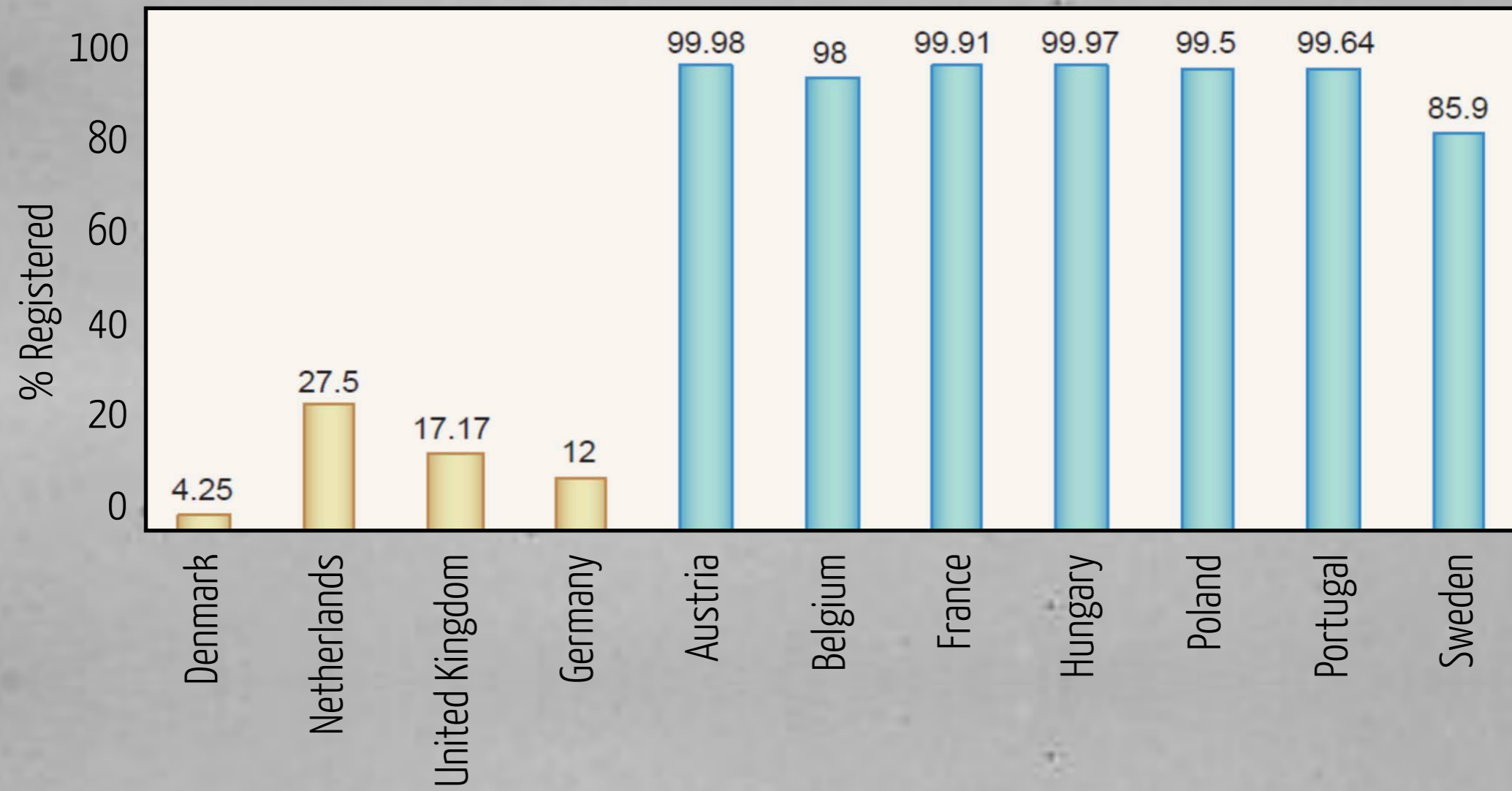
Murdock (1962) JEP

A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball.

How much does the ball cost?

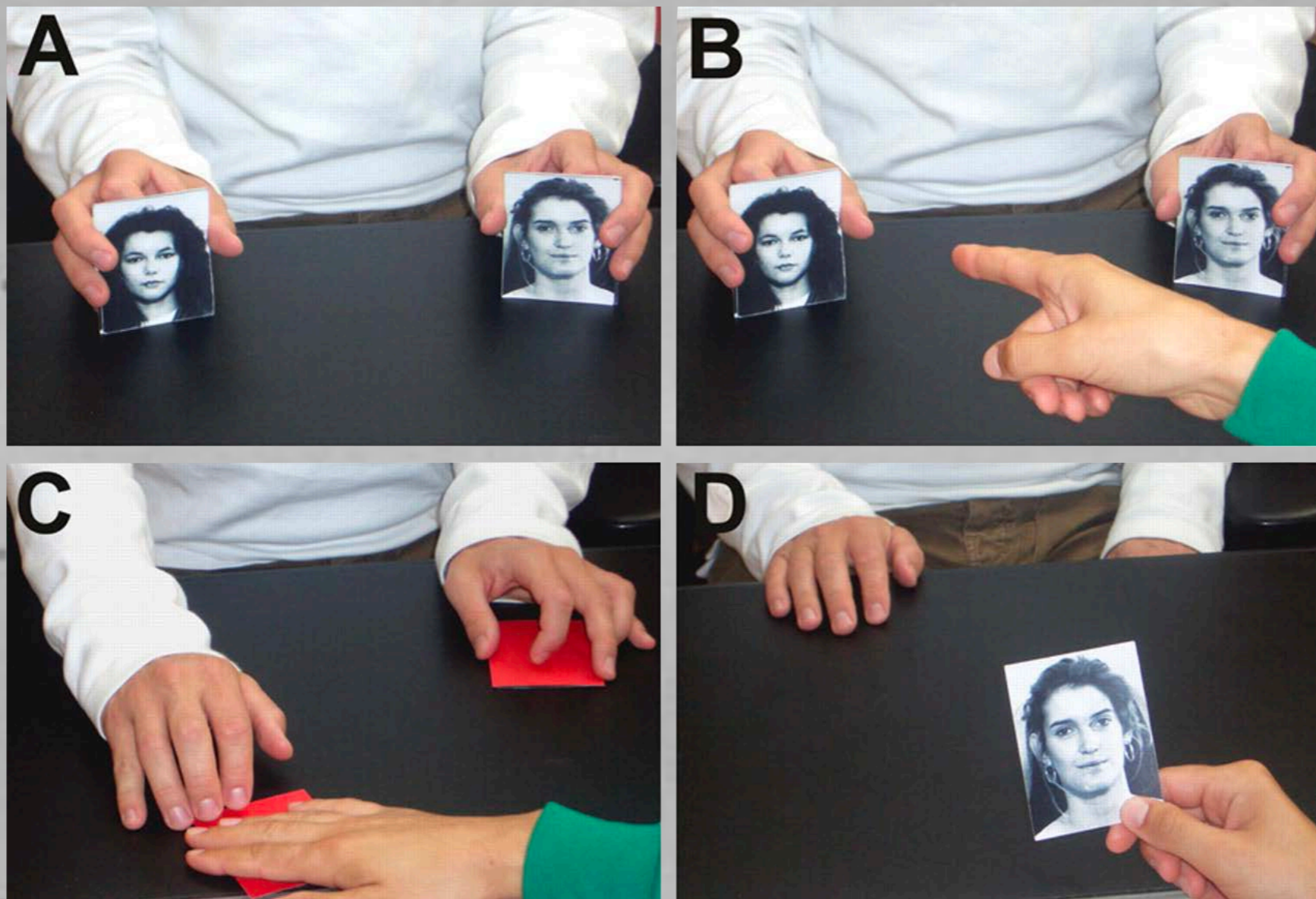


Registered for **organ donation**?



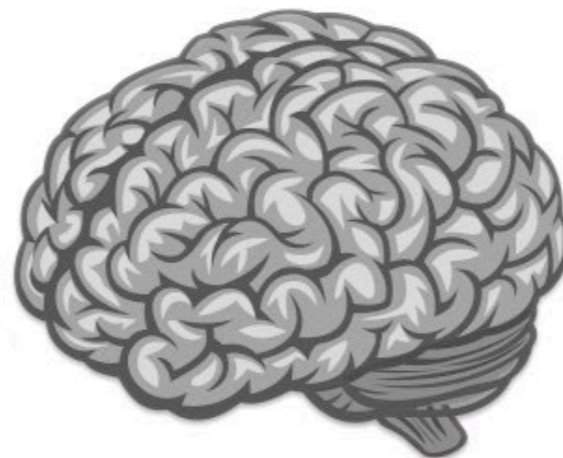
Person selects left picture.

Why did you choose the picture on the right?



Human decision making

1010111010011010
0001010111101010
1011010010010101
0011101101010000
0101111010010101
0010100010101111
0101010110100100
1010010010101001
1101101010000010



Perfect



Chance

Really that
stupid?





Broadly stated, the task is to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist.

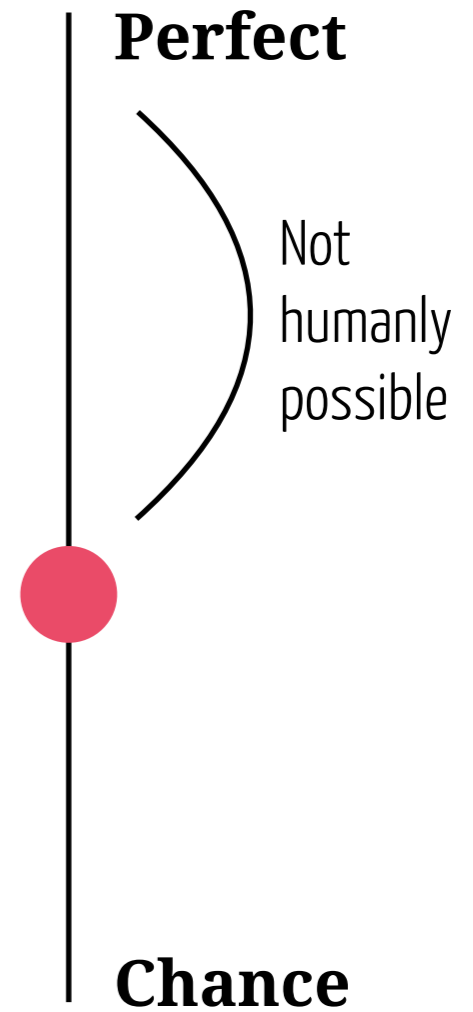
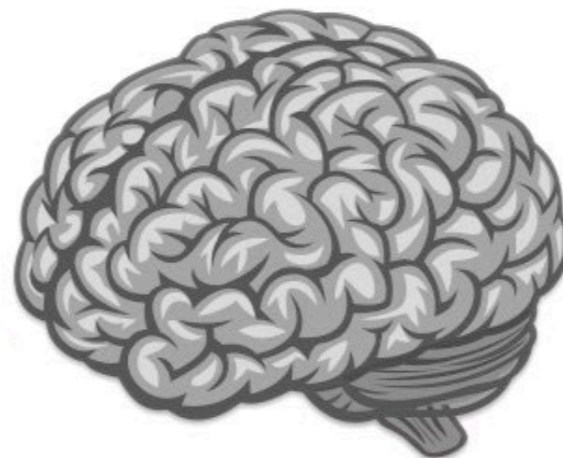
Simon (1955, p. 241)

Herbert Simon

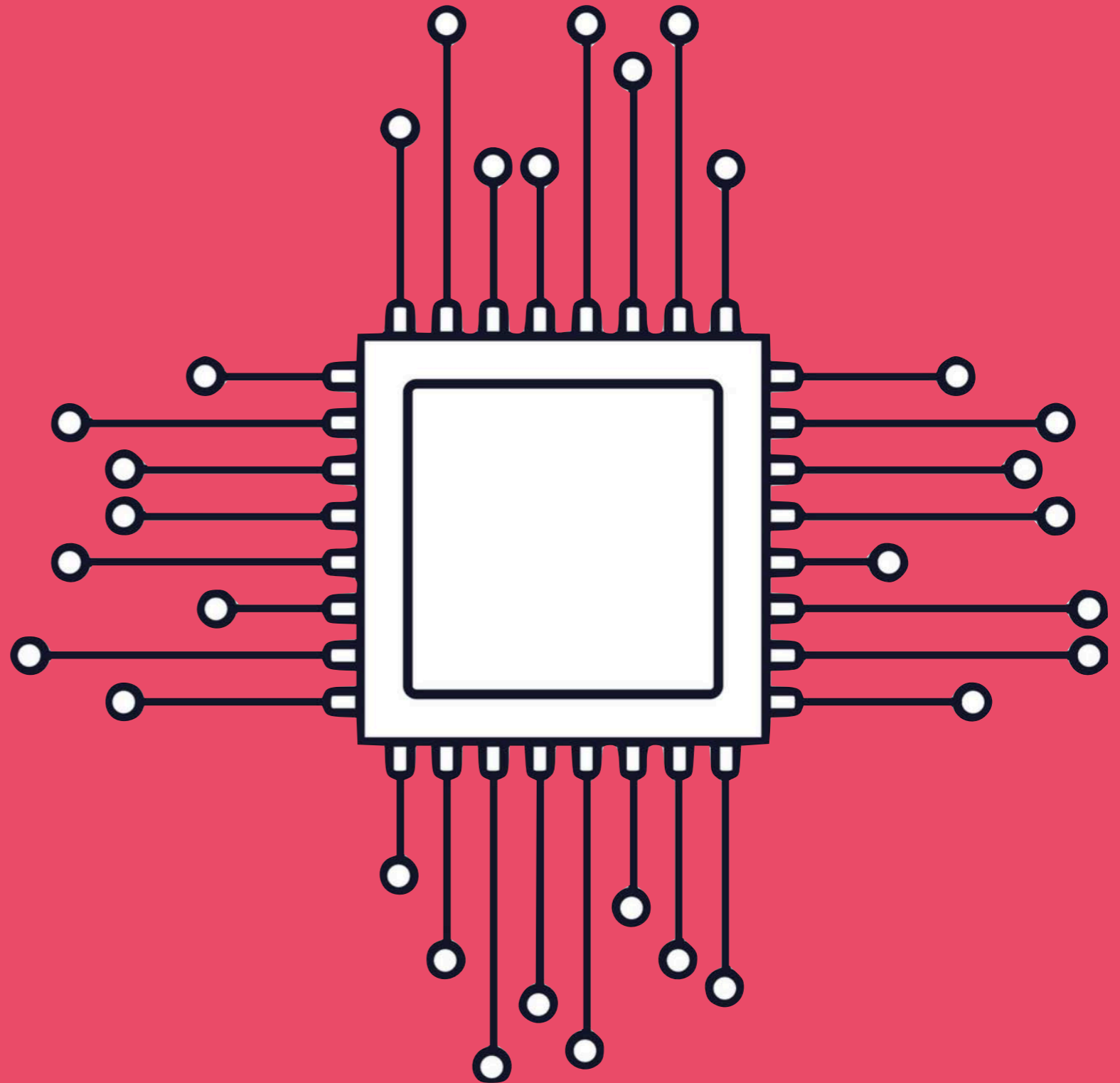
1916 - 2001

Human decision making

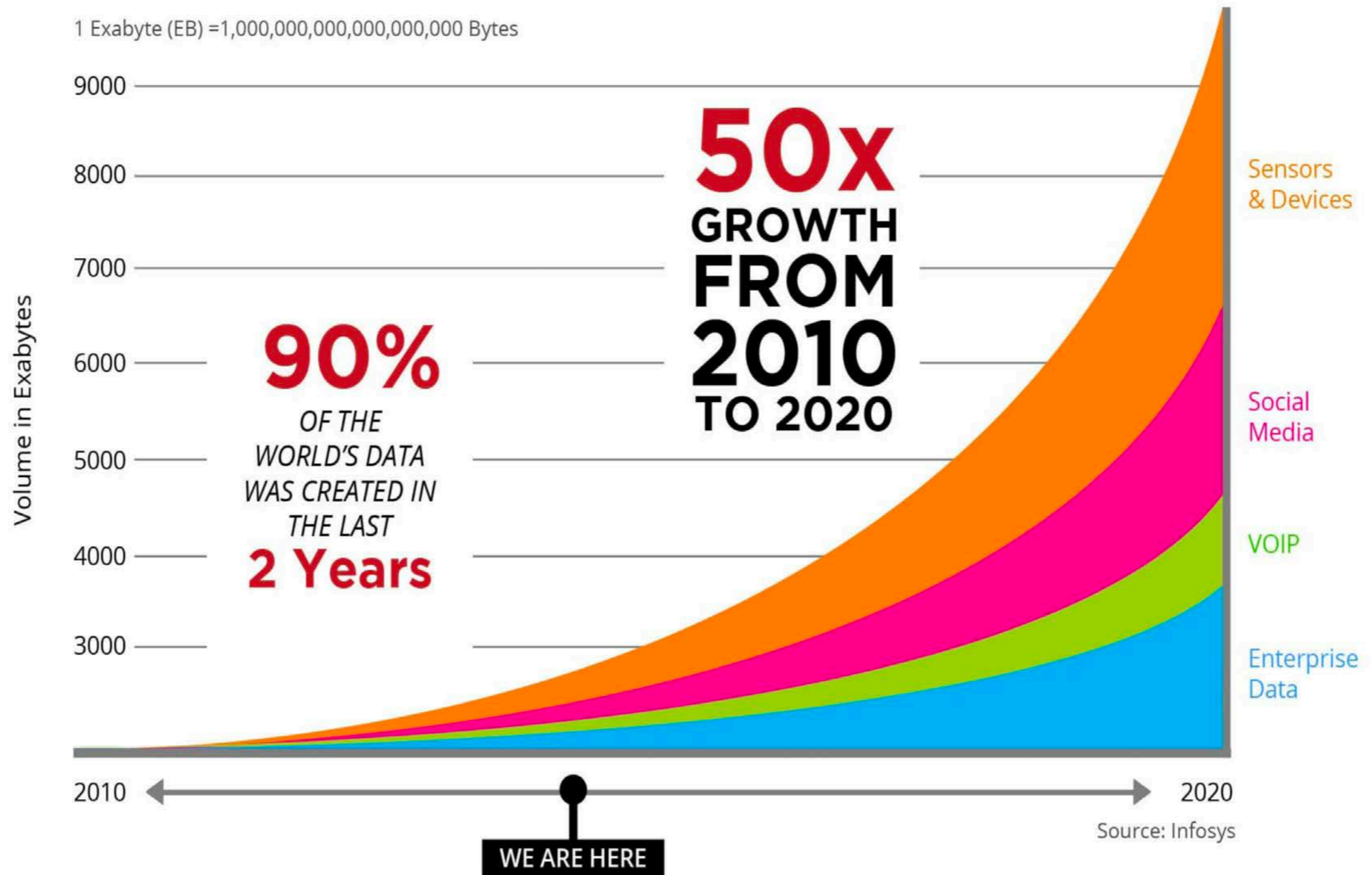
1010111010011010
0001010111101010
1011010010010101
0011101101010000
0101111010010101
0010100010101111
0101010110100100
1010010010101001
1101101010000010



Still, computers
do better!



Data explosion



Computing explosion



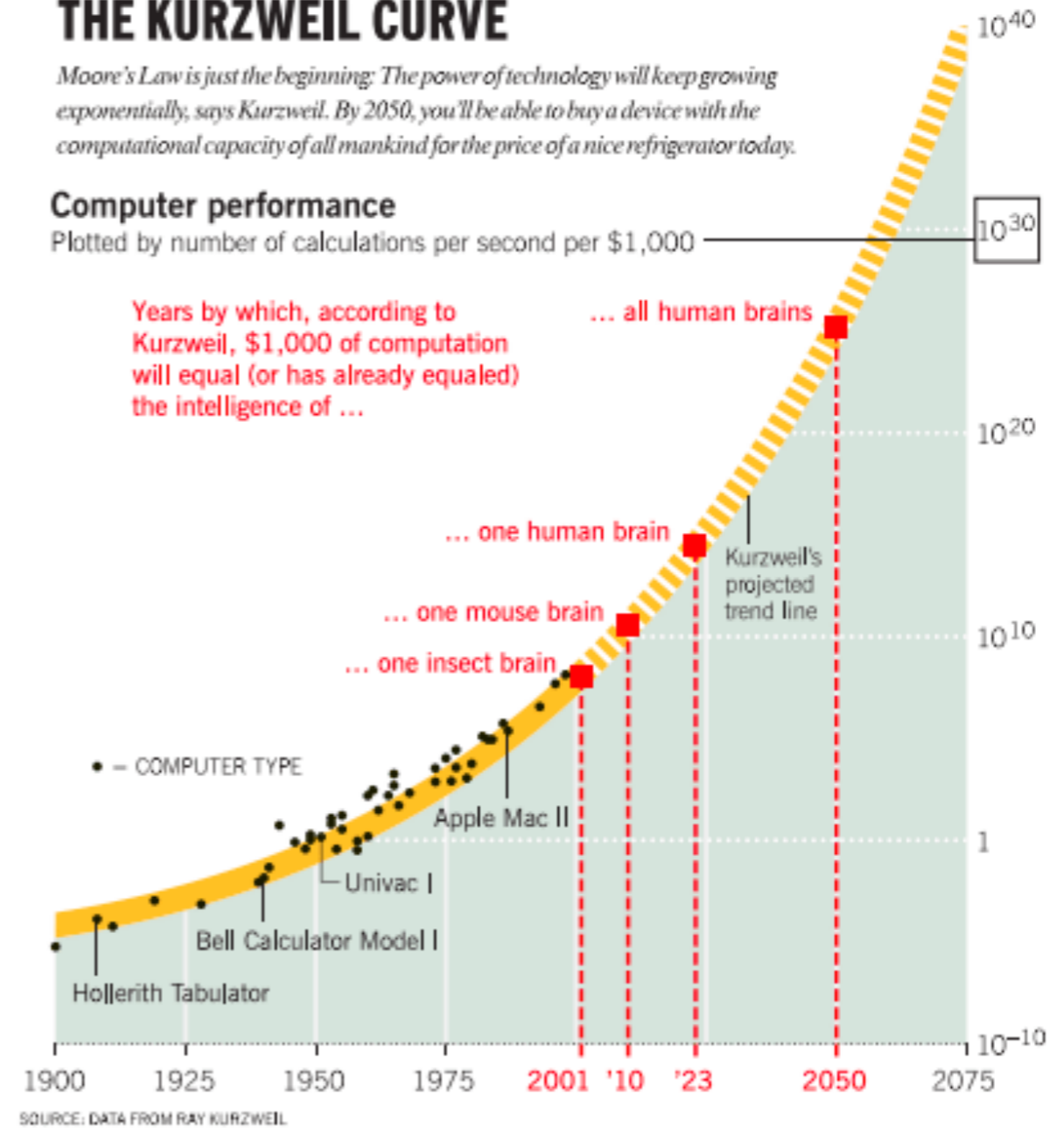
Ray Kurzweil

THE KURZWEIL CURVE

Moore's Law is just the beginning: The power of technology will keep growing exponentially, says Kurzweil. By 2050, you'll be able to buy a device with the computational capacity of all mankind for the price of a nice refrigerator today.

Computer performance

Plotted by number of calculations per second per \$1,000

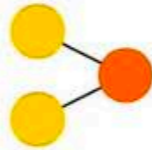


A mostly complete chart of Neural Networks

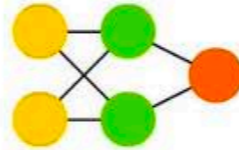
©2016 Fjodor van Veen - asimovinstitute.org

-  Backfed Input Cell
-  Input Cell
-  Noisy Input Cell
-  Hidden Cell
-  Probablistic Hidden Cell
-  Spiking Hidden Cell
-  Output Cell
-  Match Input Output Cell
-  Recurrent Cell
-  Memory Cell
-  Different Memory Cell
-  Kernel
-  Convolution or Pool

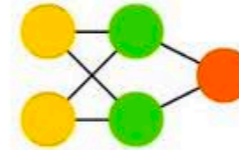
Perceptron (P)



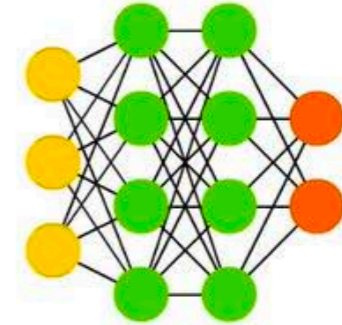
Feed Forward (FF)



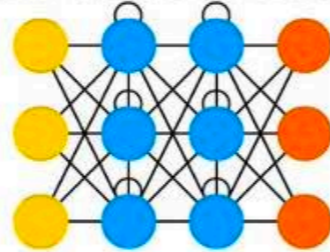
Radial Basis Network (RBF)



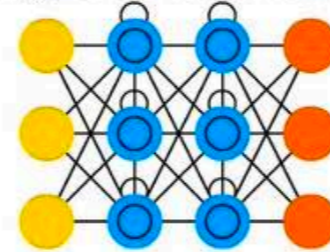
Deep Feed Forward (DFF)



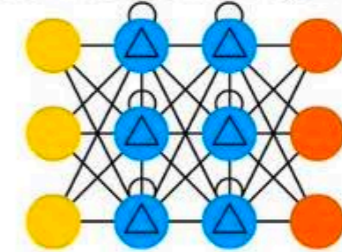
Recurrent Neural Network (RNN)



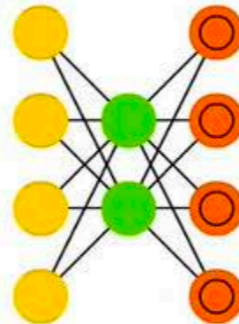
Long / Short Term Memory (LSTM)



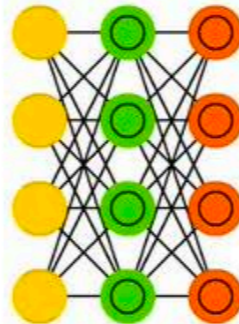
Gated Recurrent Unit (GRU)



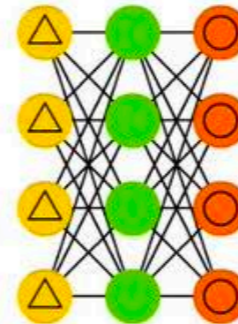
Auto Encoder (AE)



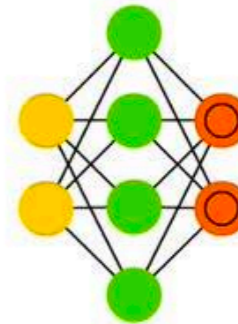
Variational AE (VAE)



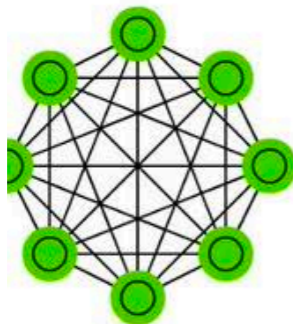
Denosing AE (DAE)



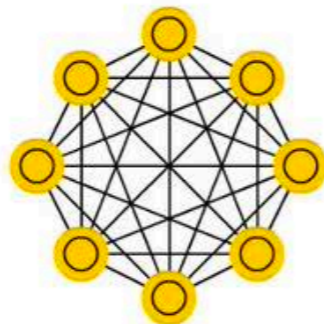
Sparse AE (SAE)



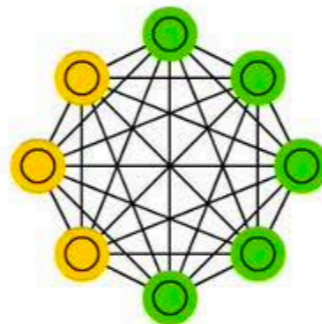
Markov Chain (MC)



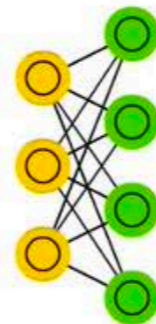
Hopfield Network (HN)



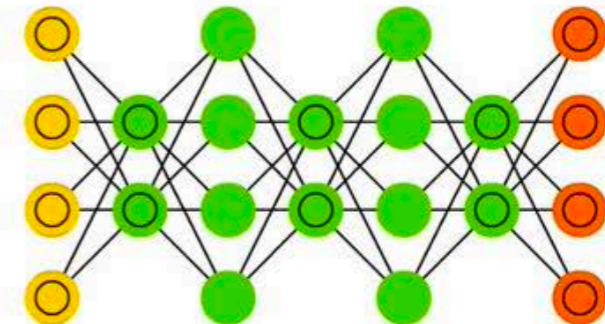
Boltzmann Machine (BM)



Restricted BM (RBM)



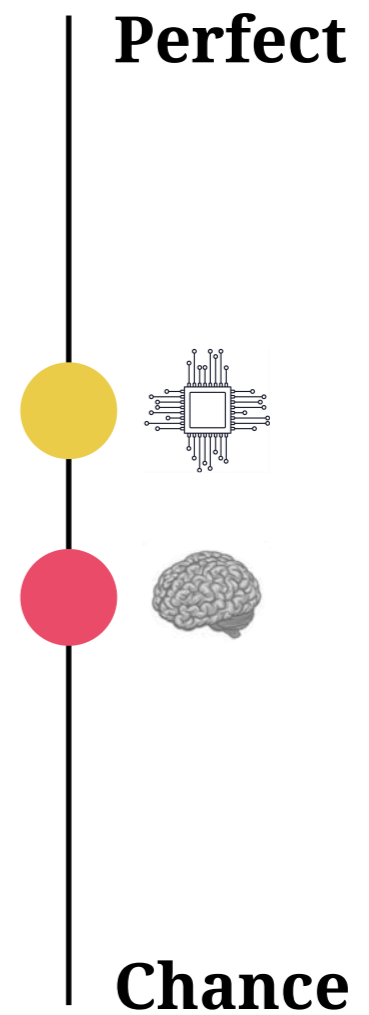
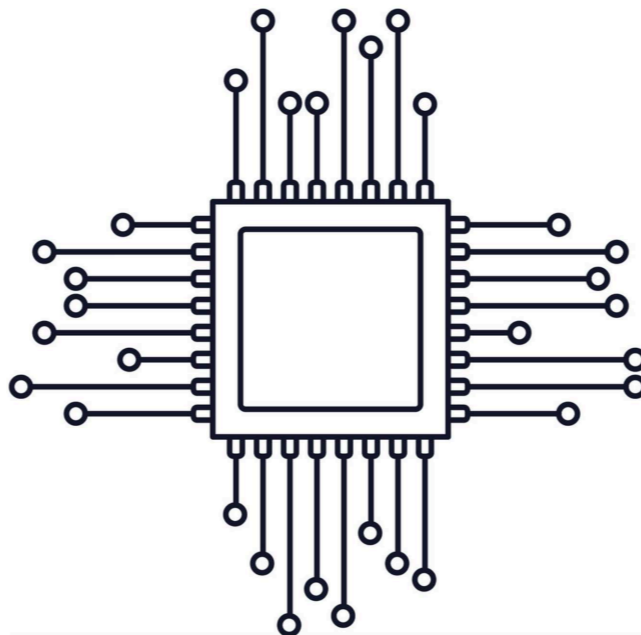
Deep Belief Network (DBN)



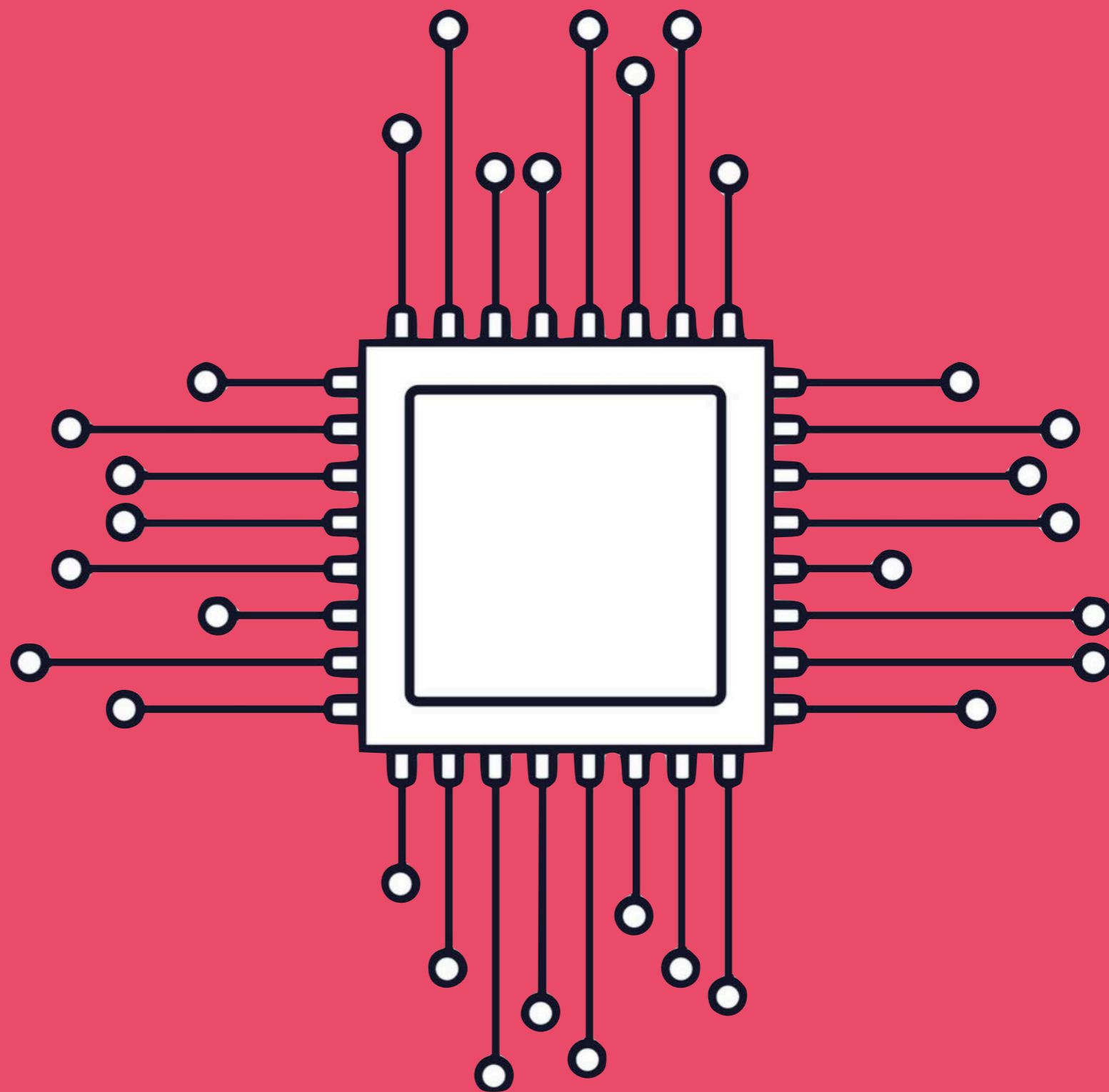
from The Asimov Institute

Machine decision making

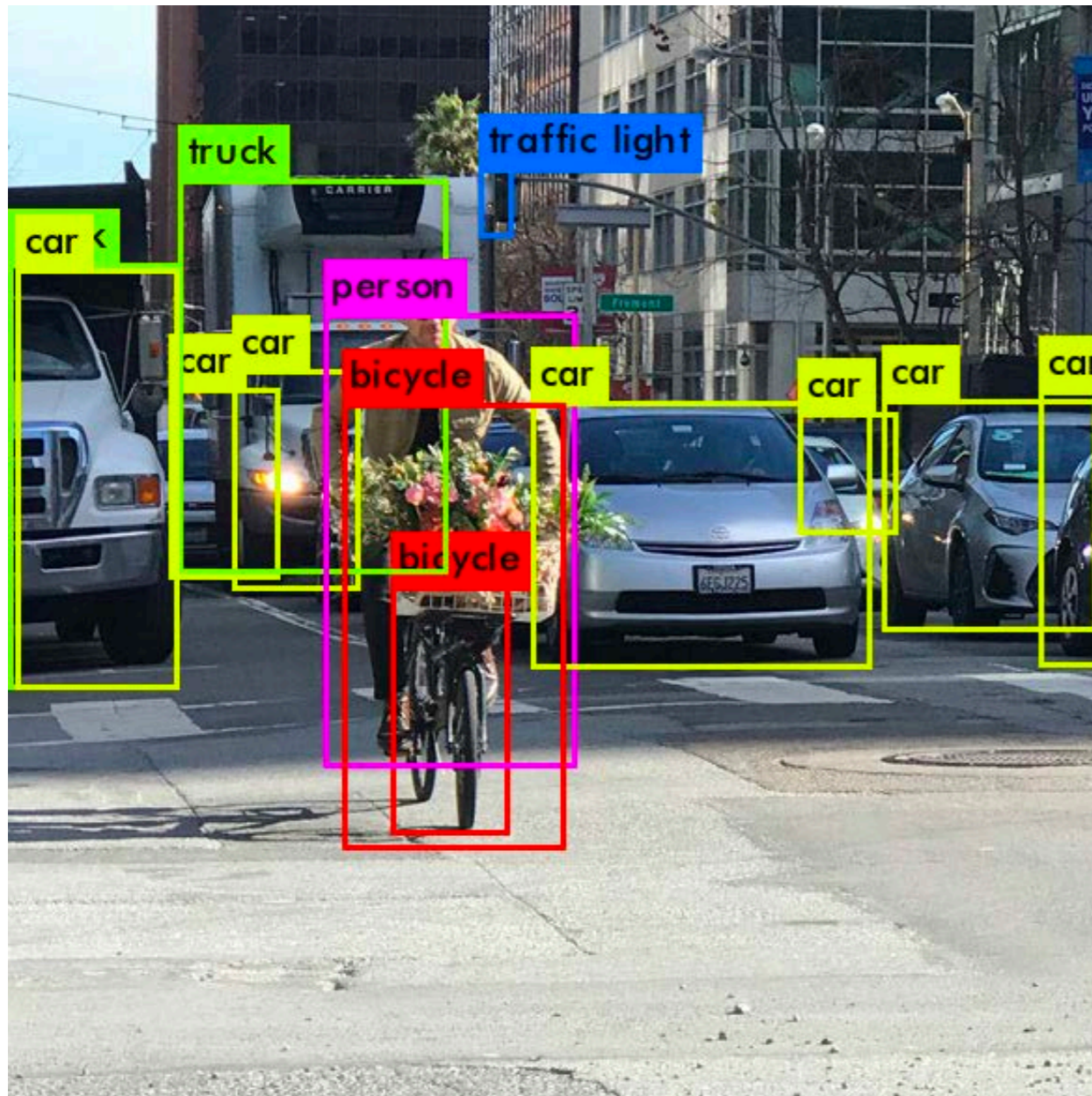
1010111010011010
0001010111101010
1011010010010101
0011101101010000
0101111010010101
0010100010101111
0101010110100100
1010010010101001
1101101010000010



What computers
can do!

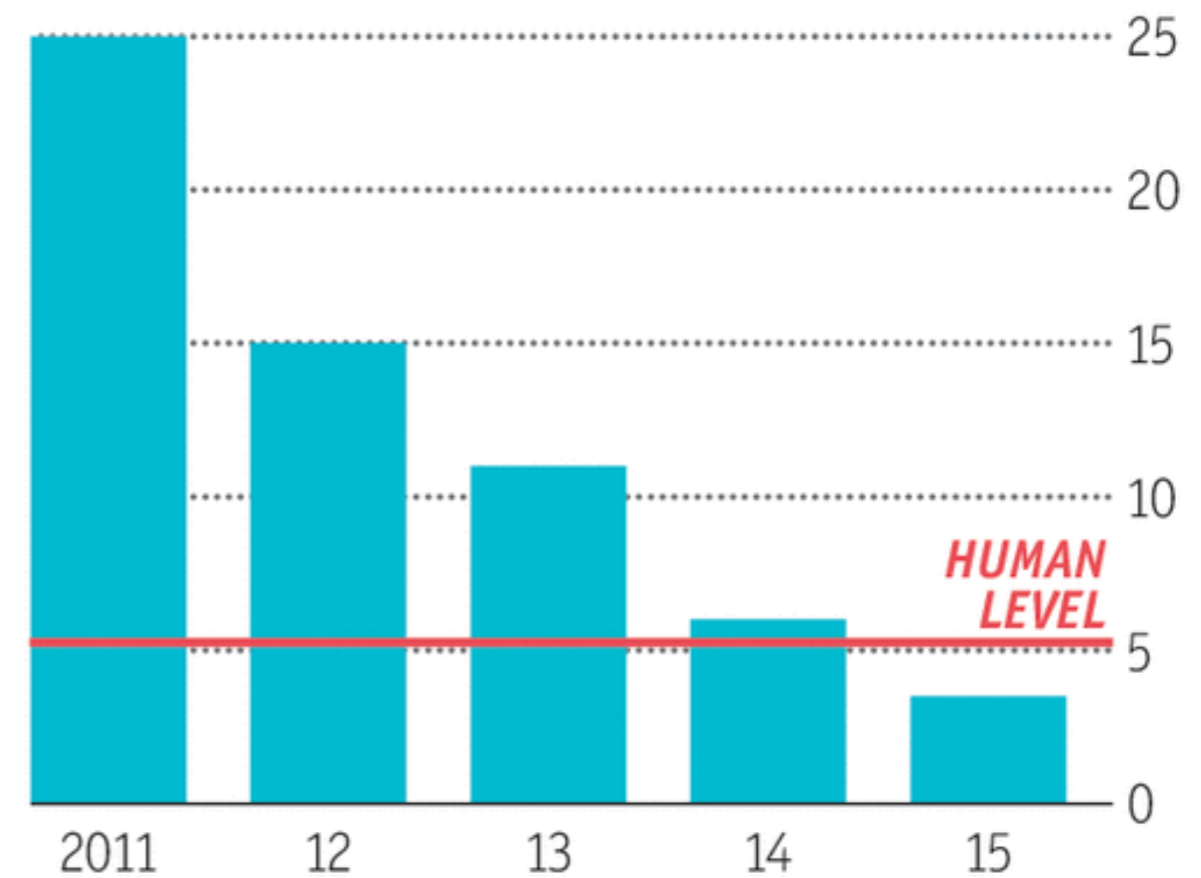


Object recognition



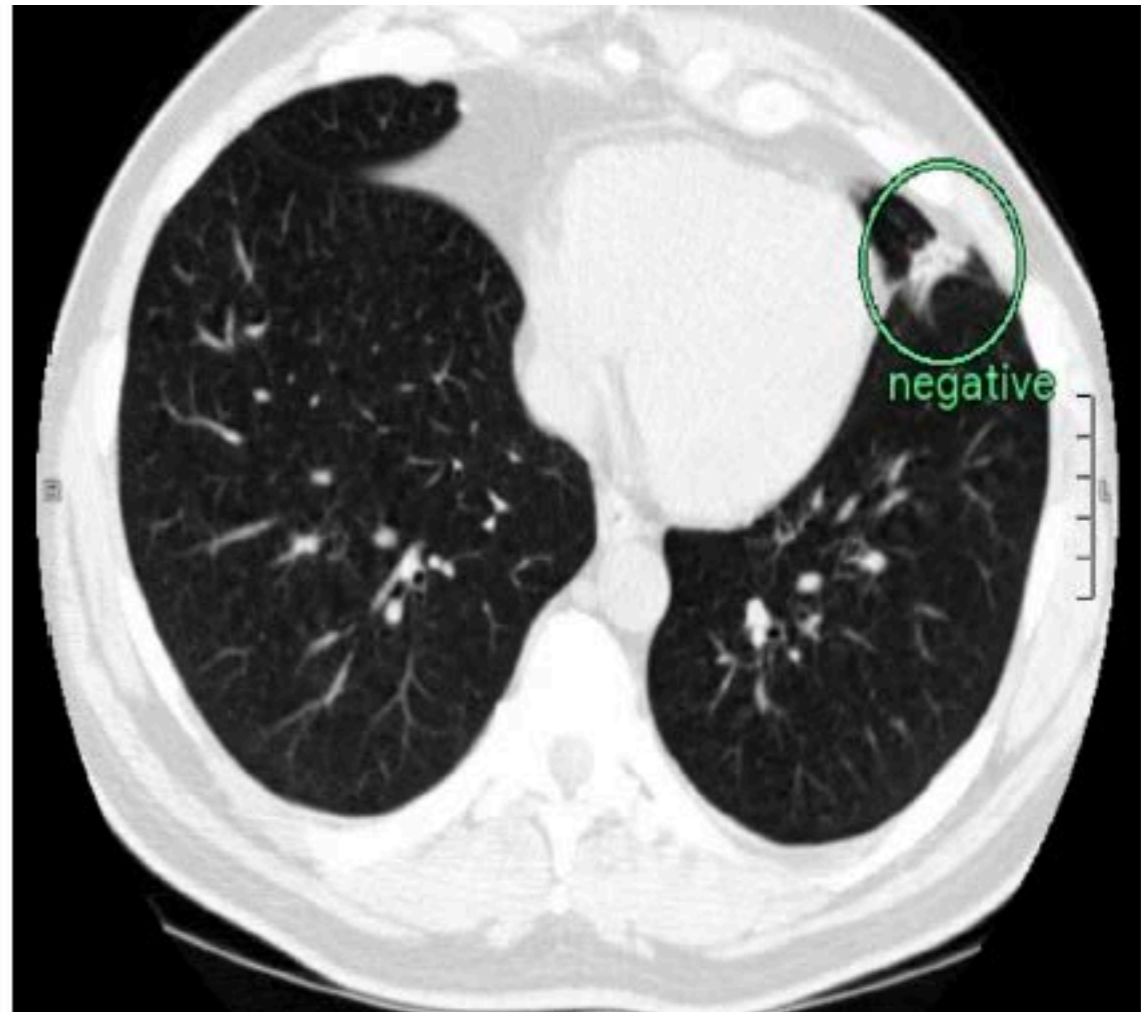
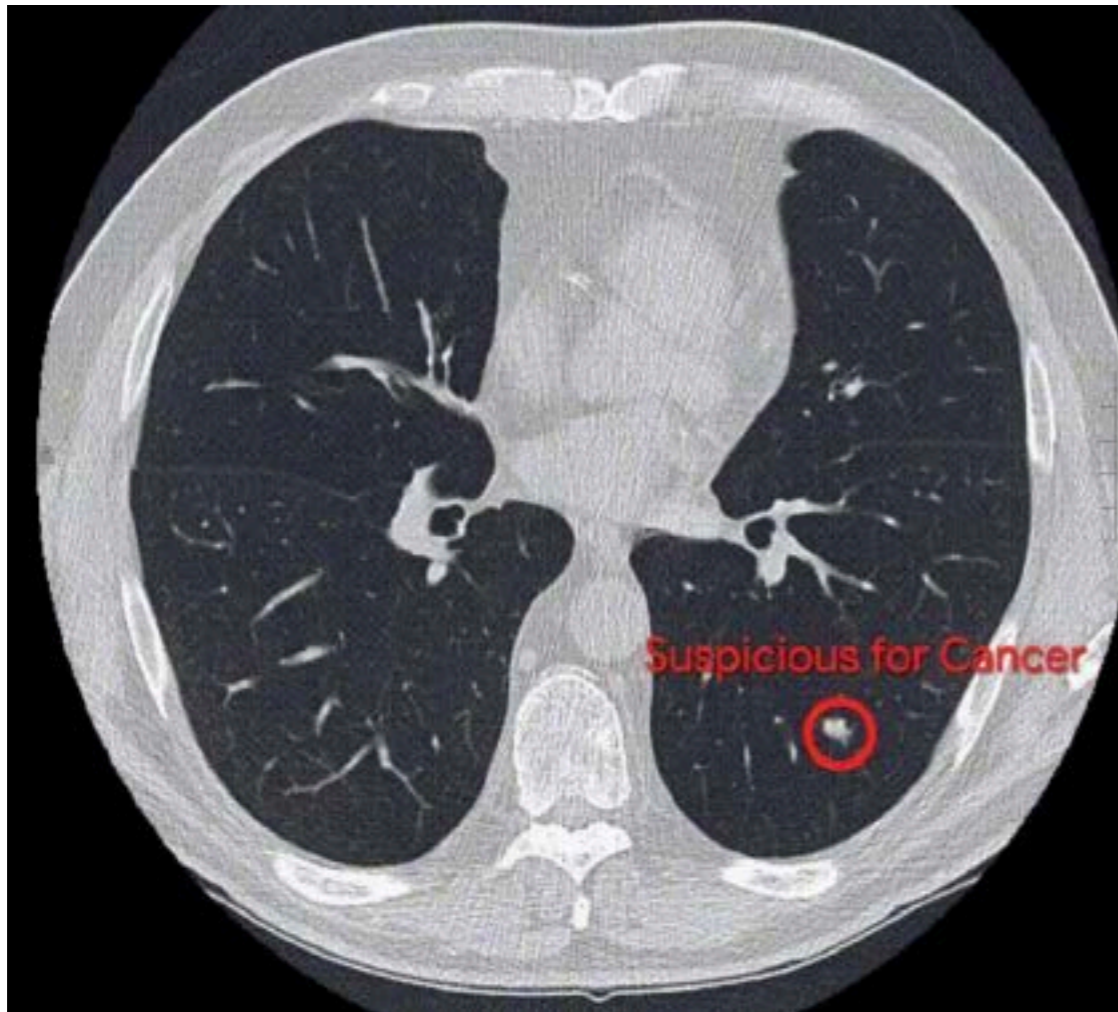
Ever cleverer

Error rates on ImageNet Visual Recognition Challenge, %

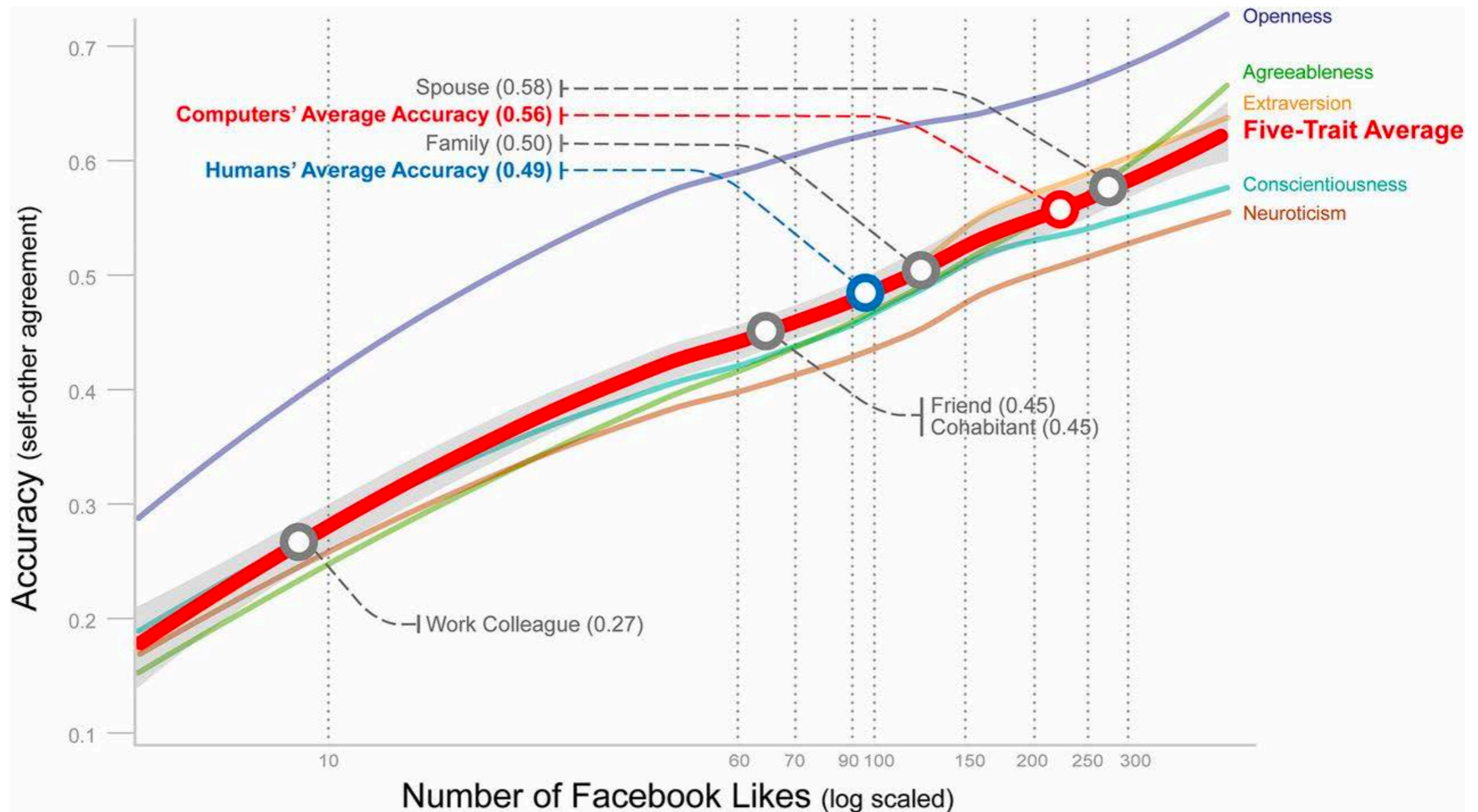


Sources: ImageNet; Stanford Vision Lab

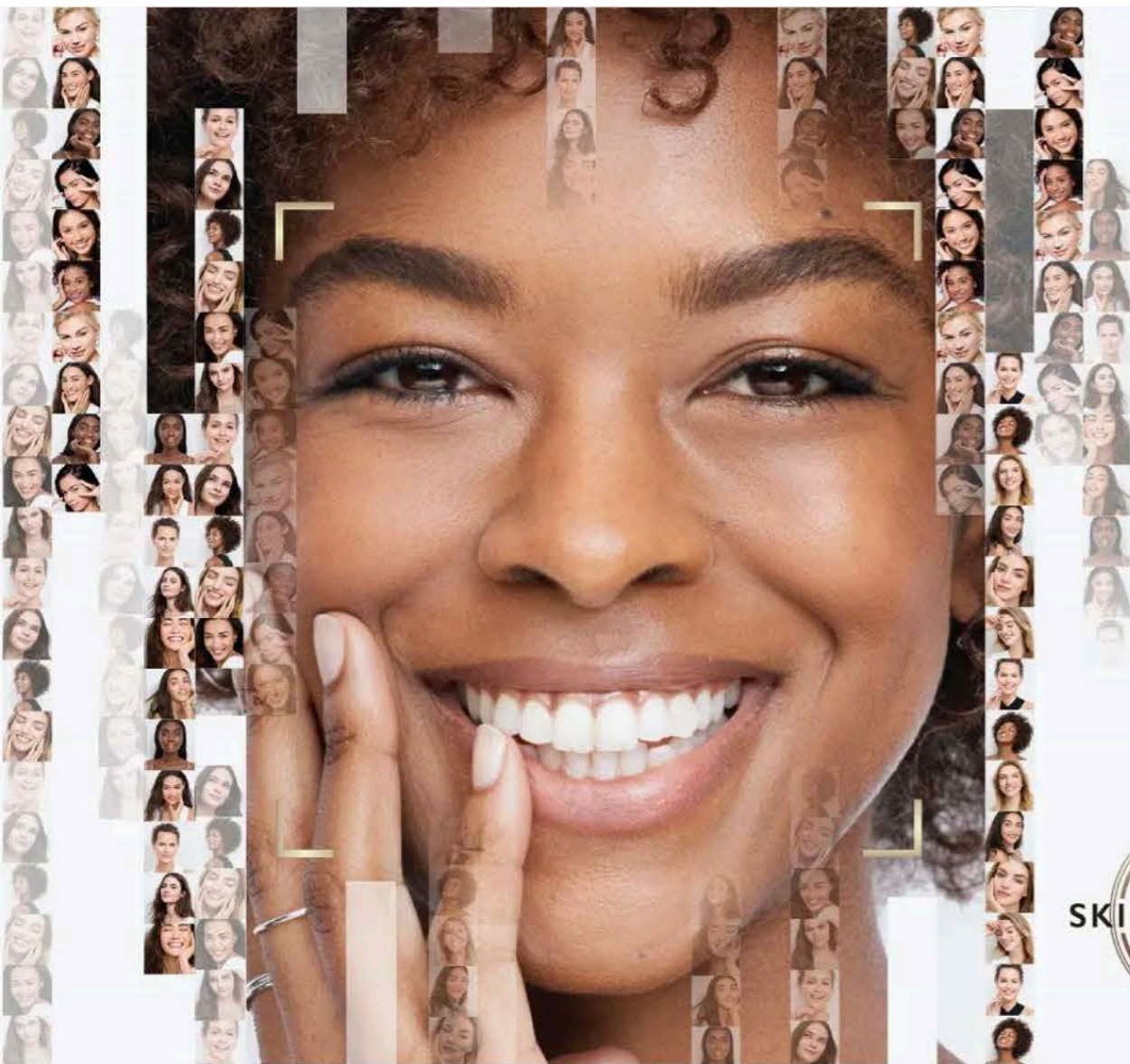
Cancer screening



Personality prediction



Digital marketing



SKINADVISOR

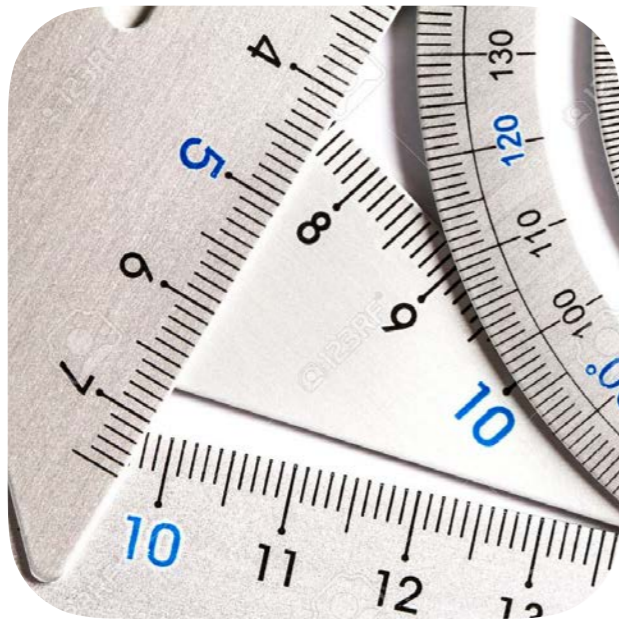
*Are you ready for your
best skin day?*

TRY IT NOW



Productivity benefits

Precise



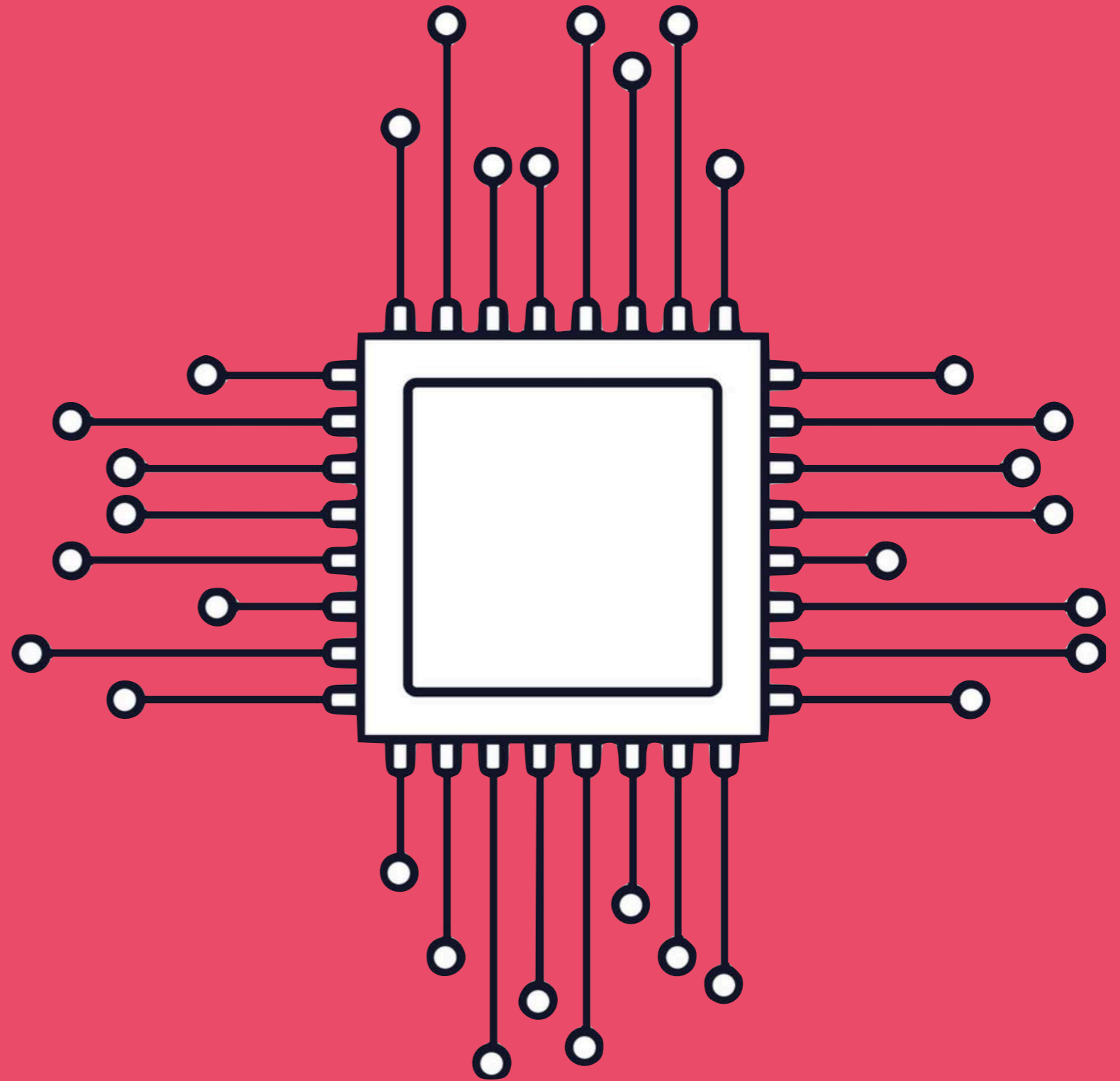
Reproducible



Objective



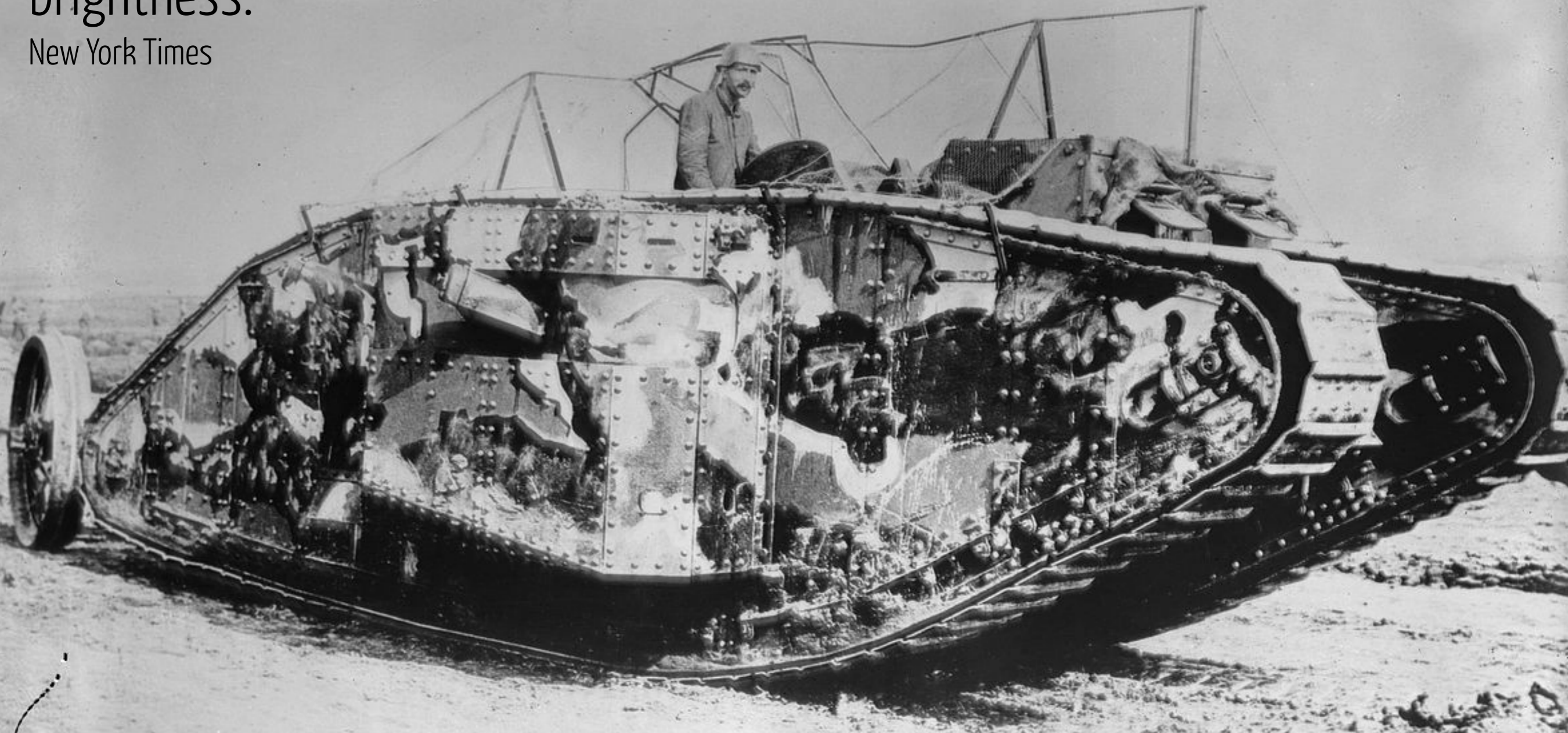
Yes, computers can be
stupid, too!



An urban myth?!

"The Army trained a program to differentiate American tanks from Russian tanks with 100% accuracy. Only later did analysts realize that the American tanks had been photographed on a sunny day and the Russian tanks had been photographed on a cloudy day. The computer had learned to detect brightness."

New York Times

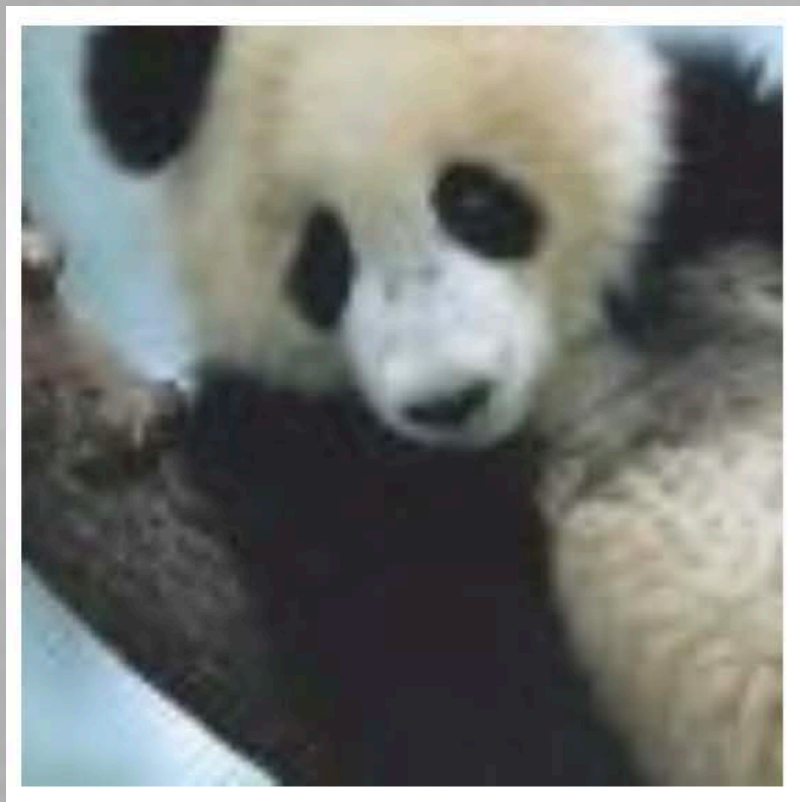


Fooling neural networks

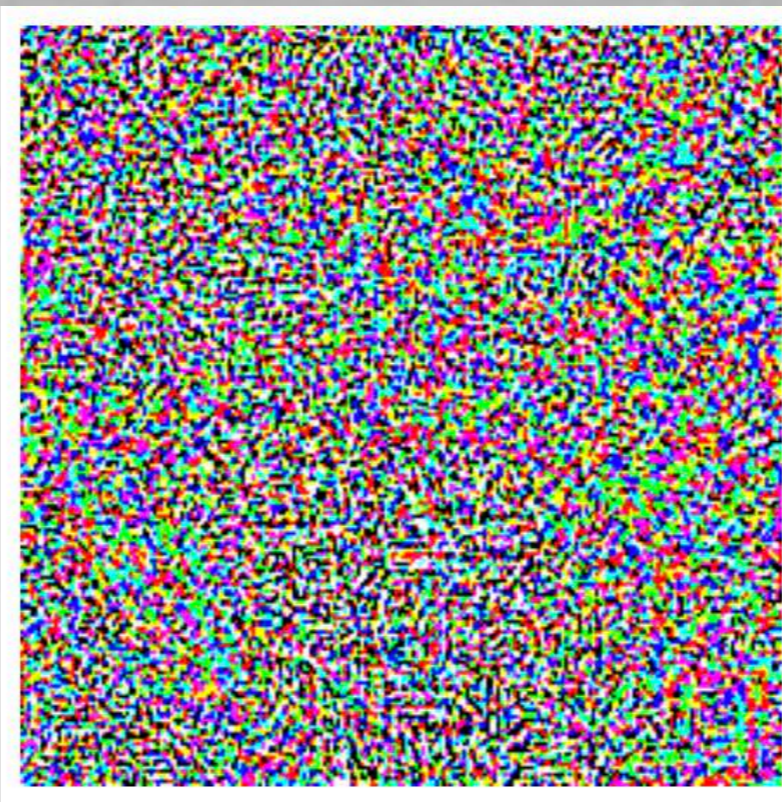
“Panda”
57.7%

“Nematode”
8.2%

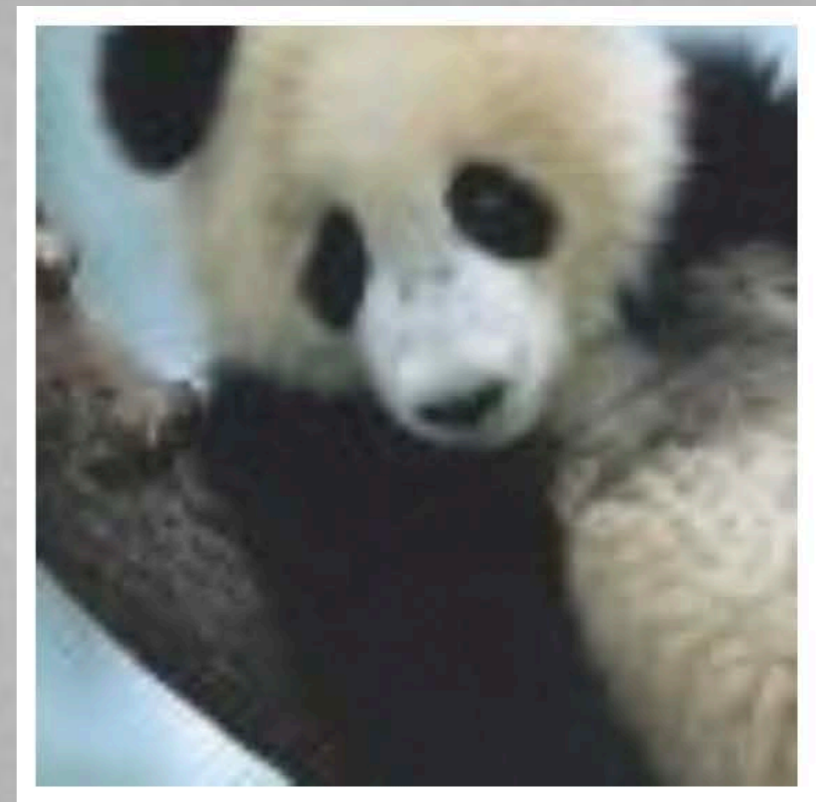
“Gibbon”
99.3%



+ .007 ×

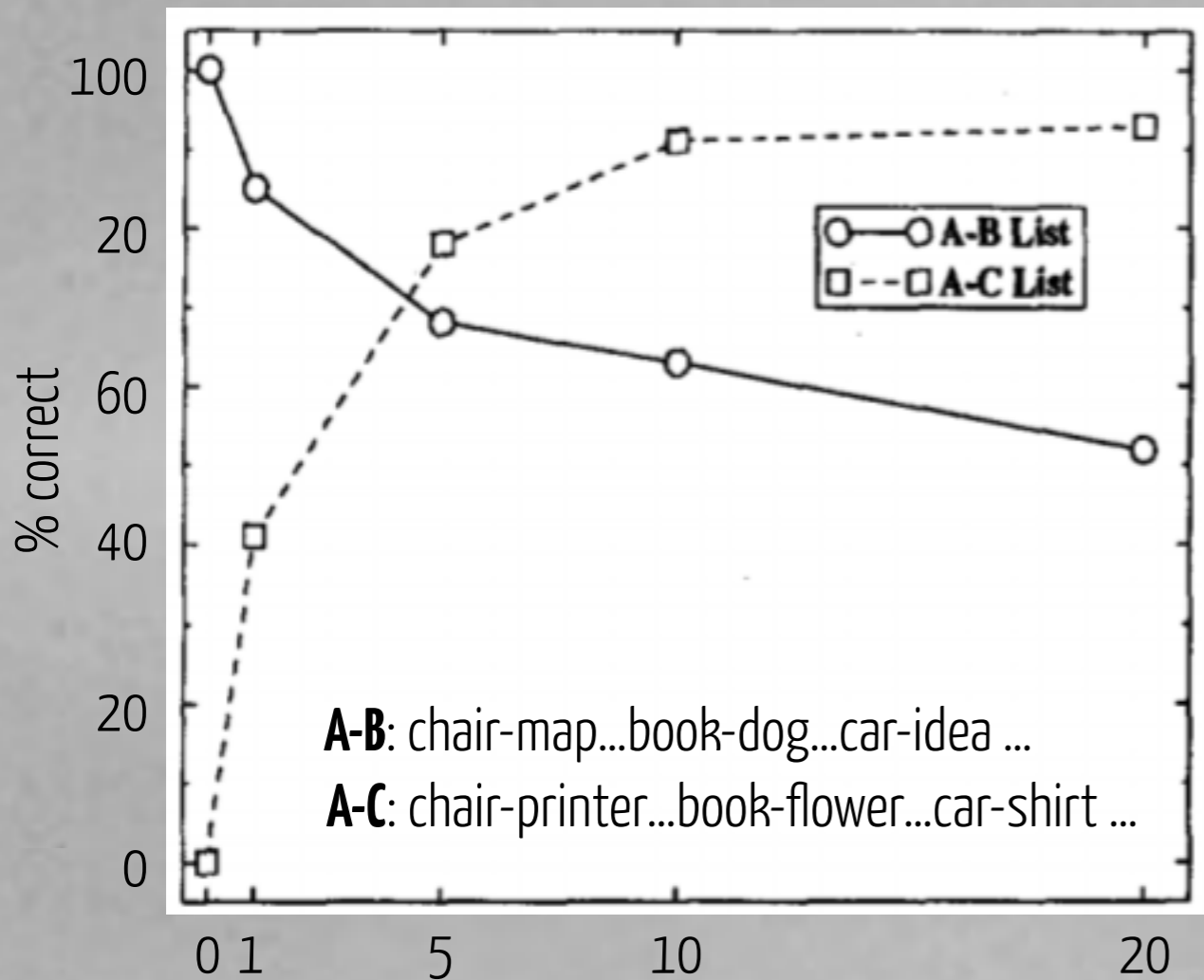


=

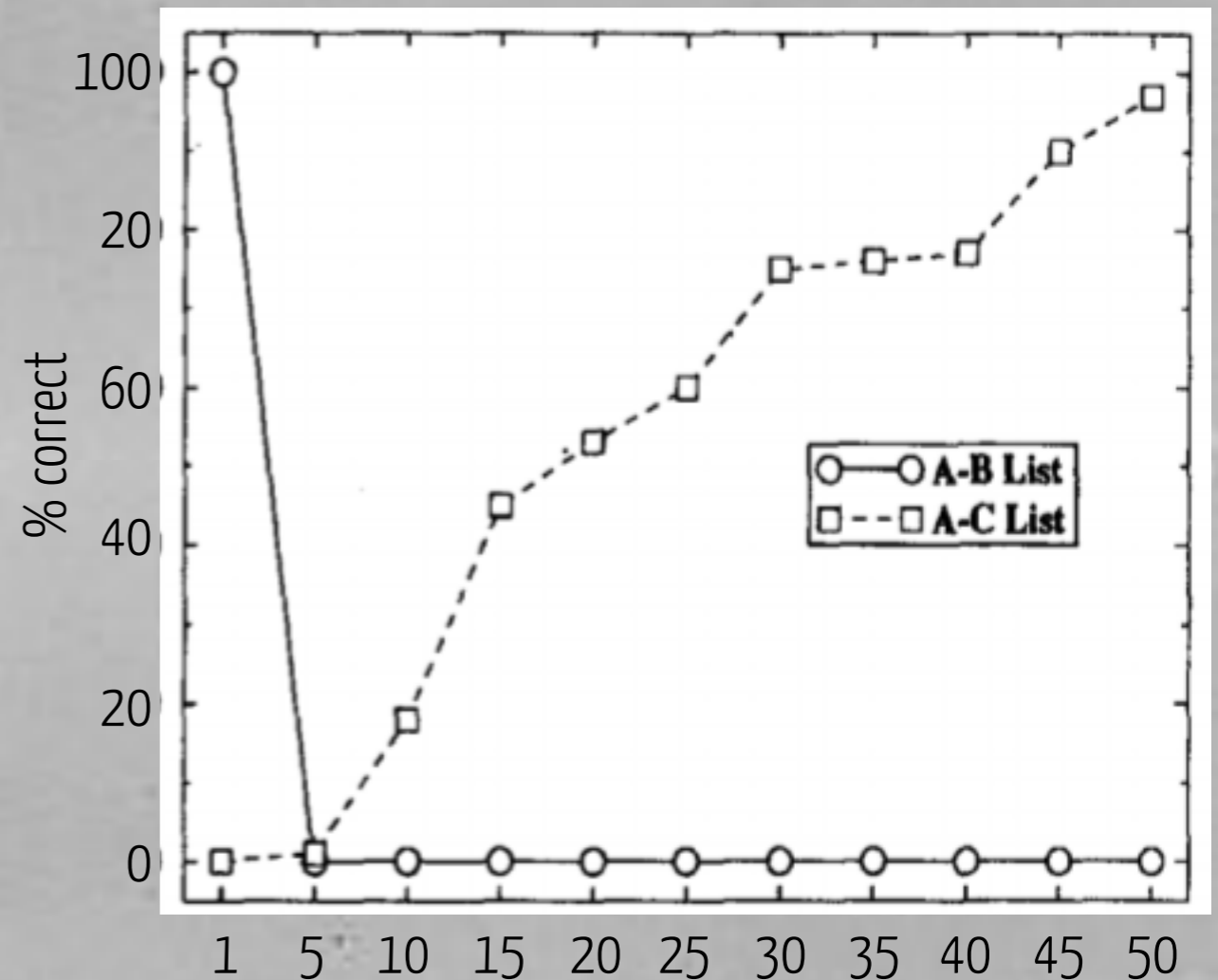


Catastrophic interference

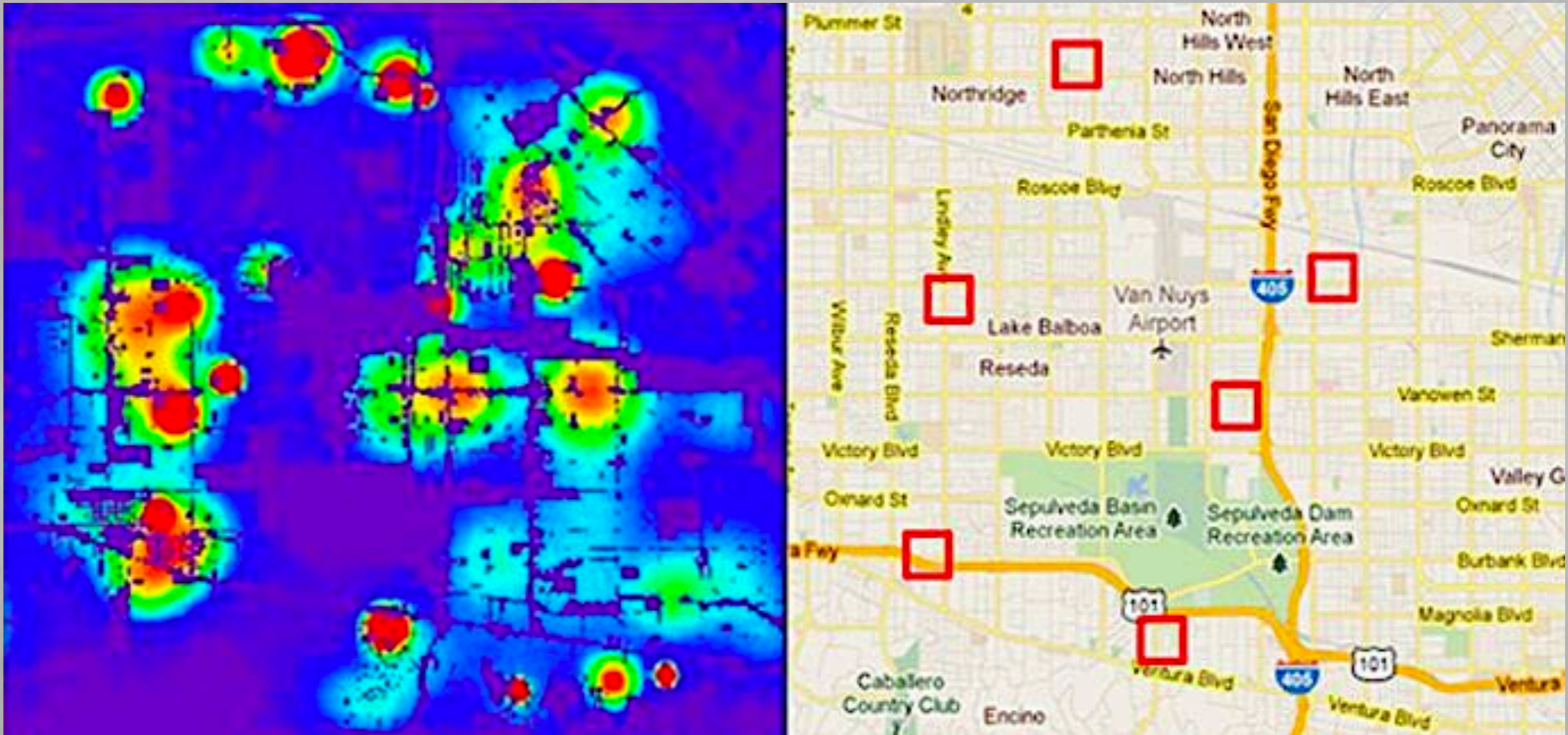
Human



Model

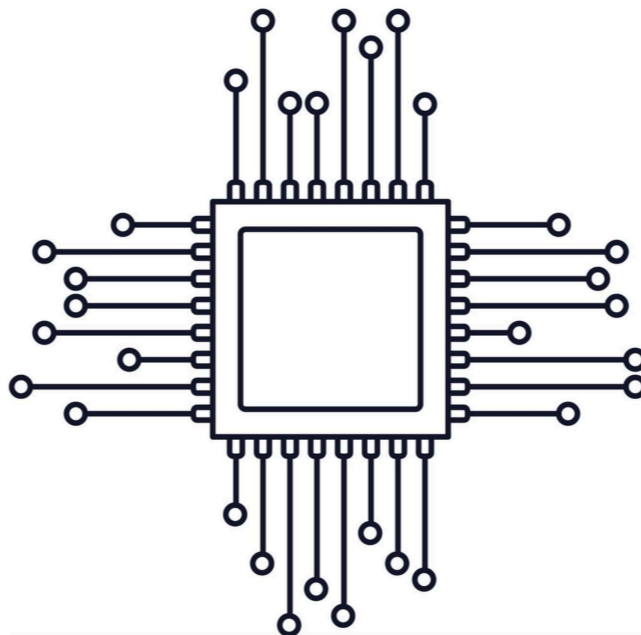


Repeating the past



Machine decision making

1010111010011010
0001010111101010
1011010010010101
0011101101010000
0101111010010101
0010100010101111
0101010110100100
1010010010101001
1101101010000010



Perfect

Not
"machinely"
possible



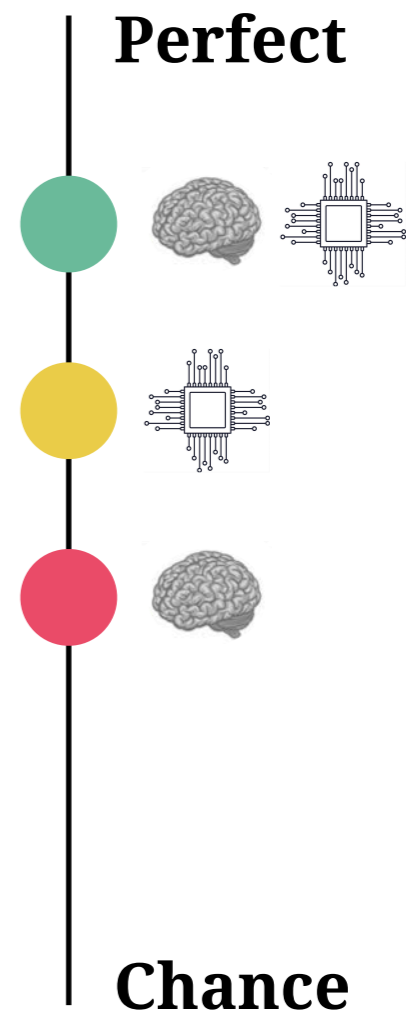
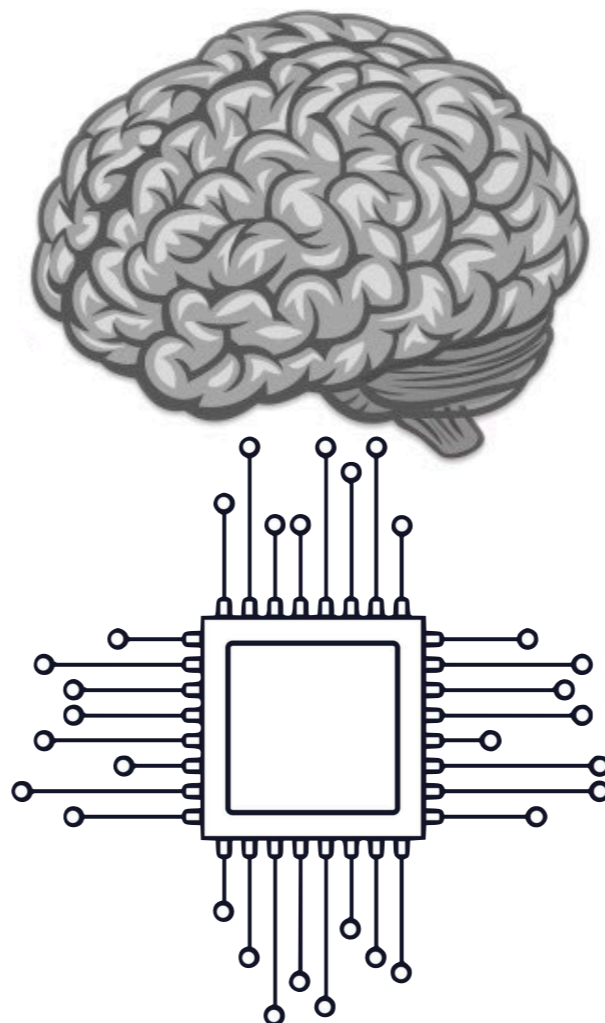
Chance

Implications for educators

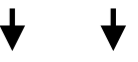


Human-machine decision making

1010111010011010
0001010111101010
1011010010010101
0011101101010000
0101111010010101
0010100010101111
0101010110100100
1010010010101001
1101101010000010



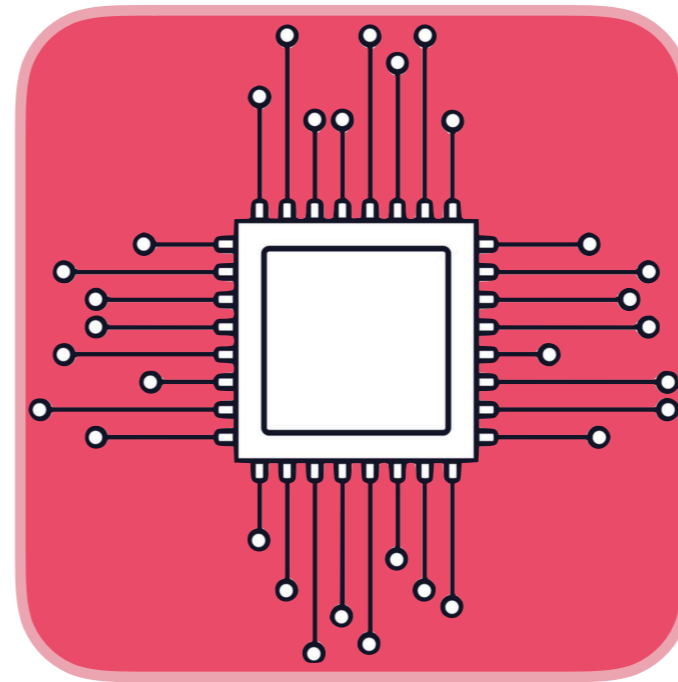
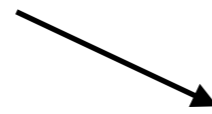
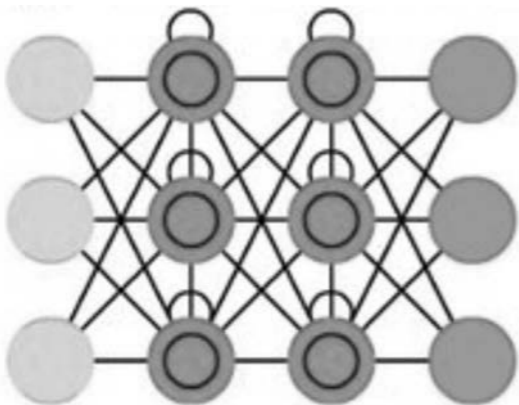
Humans' roles



1. Select the world



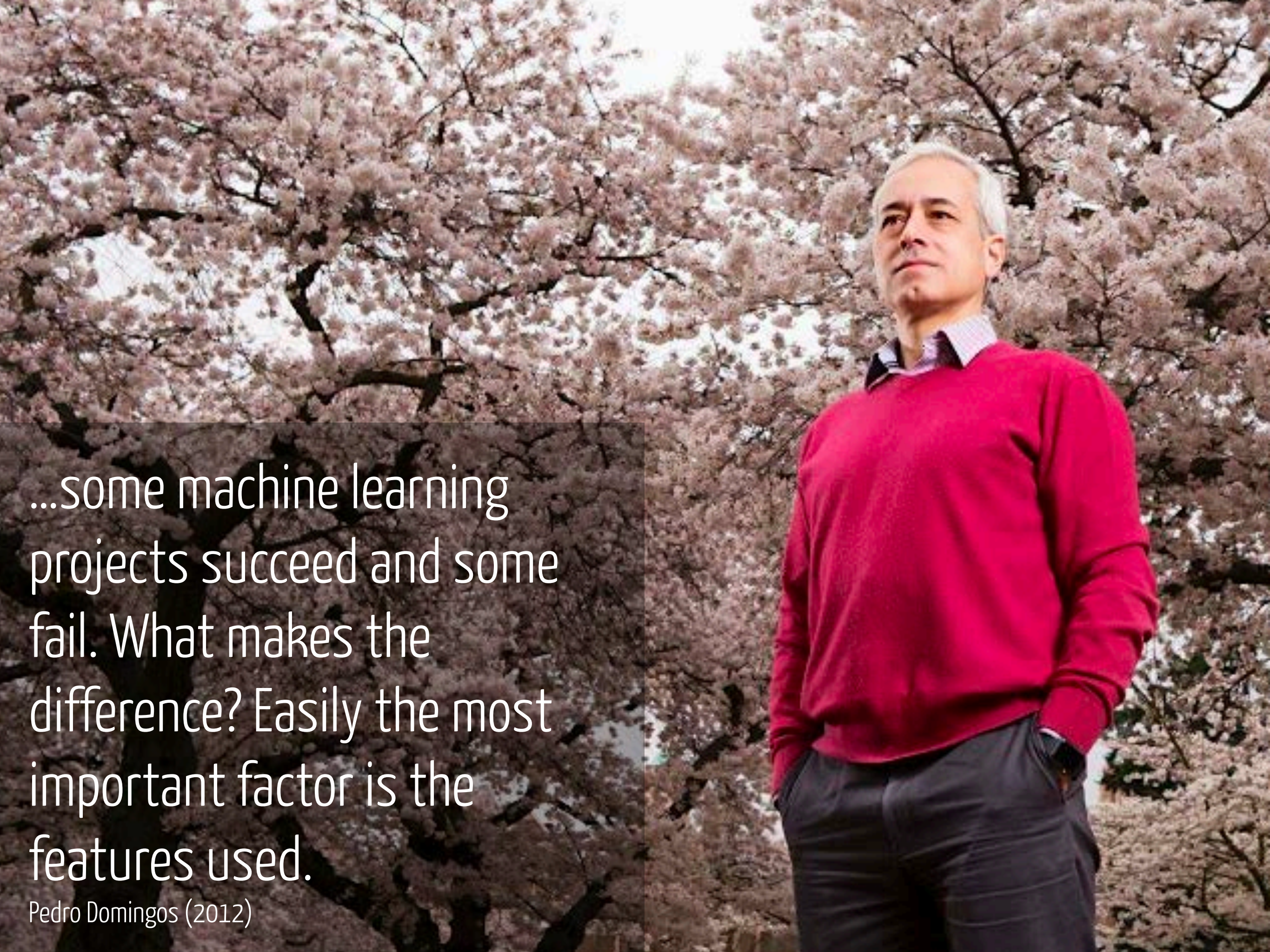
2. Select algorithms



3. Monitor output

Outcome

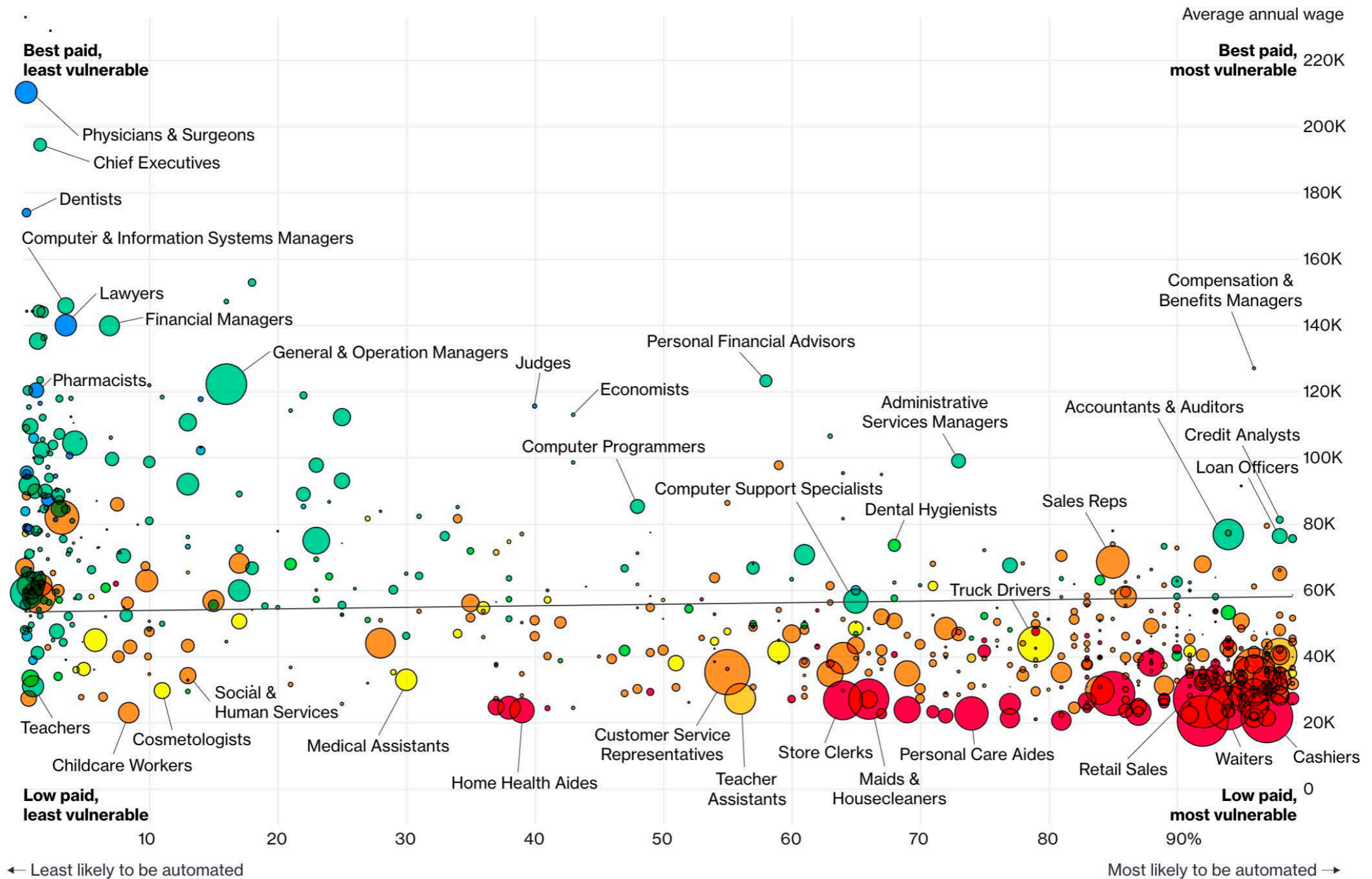


A photograph of Pedro Domingos, a man with grey hair, wearing a red sweater and dark trousers, standing in a field of cherry blossom trees. He is looking upwards and to the right. The background is filled with the dense, light pink blossoms of the trees.

...some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used.

Pedro Domingos (2012)

A changing economy



Essential **math** concepts

Data

- **Measurement**
- Independence
- Designs

Estimation

- Probability
- Combinatorics
- Distributions
- **Loss-functions**
- Numerics
- Bayes theorem

Prediction

- **Overfitting**
- Complexity
- Bias &
Variance

Conclusions

- 1) Humans make errors, and so do machines
- 2) The best decision making is done by humans and machines
- 3) Students must be educated on the workings and pitfalls of machine decision making ... at least until singularity

TheRBootcamp.com



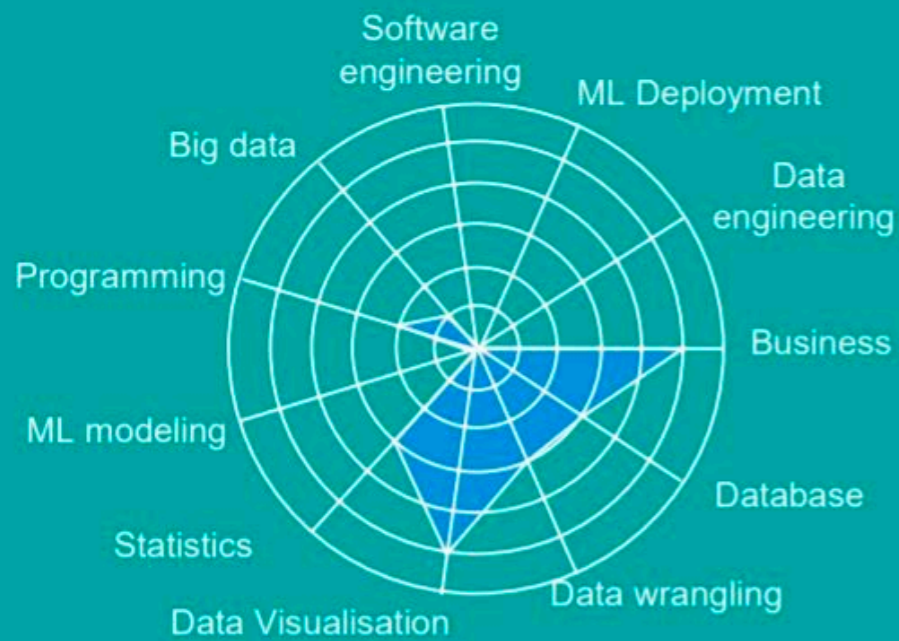


The algorithms we used are very standard for Kagglers. [...] We spent most of our efforts in feature engineering. [...] We were also very careful to discard features likely to expose us to the risk of overfitting our model..

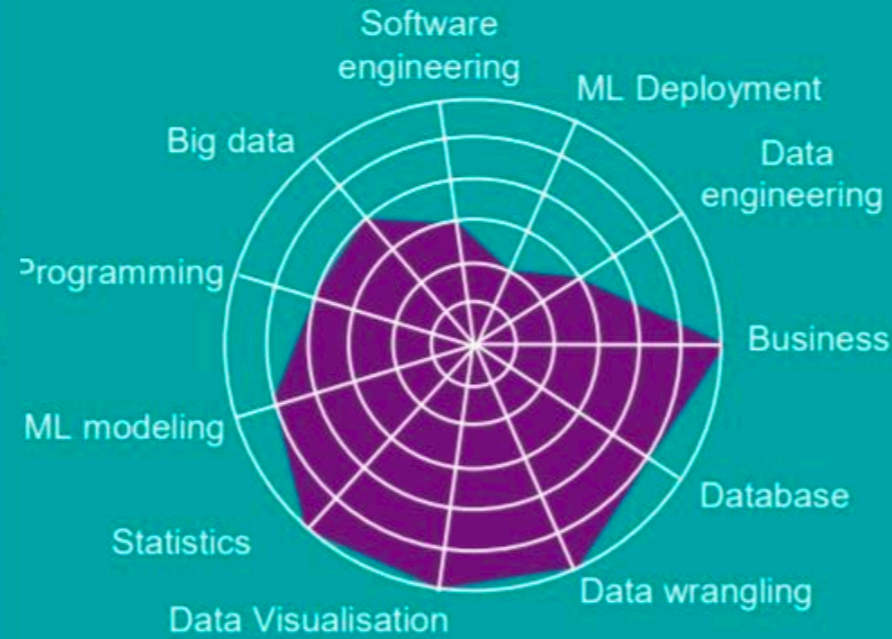
Xavier Conort (2015)

Data jobs

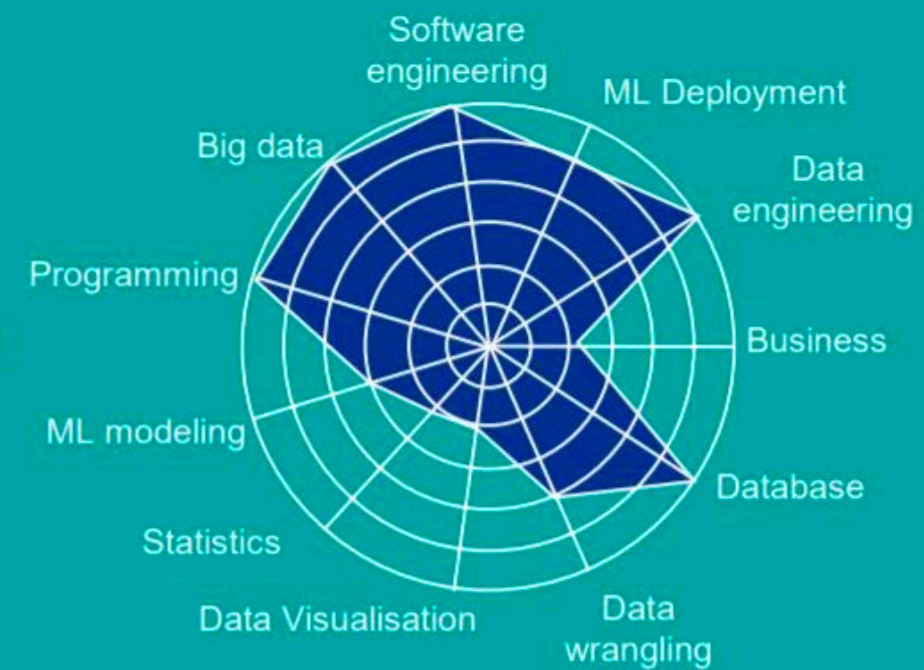
Data analyst



Data scientist



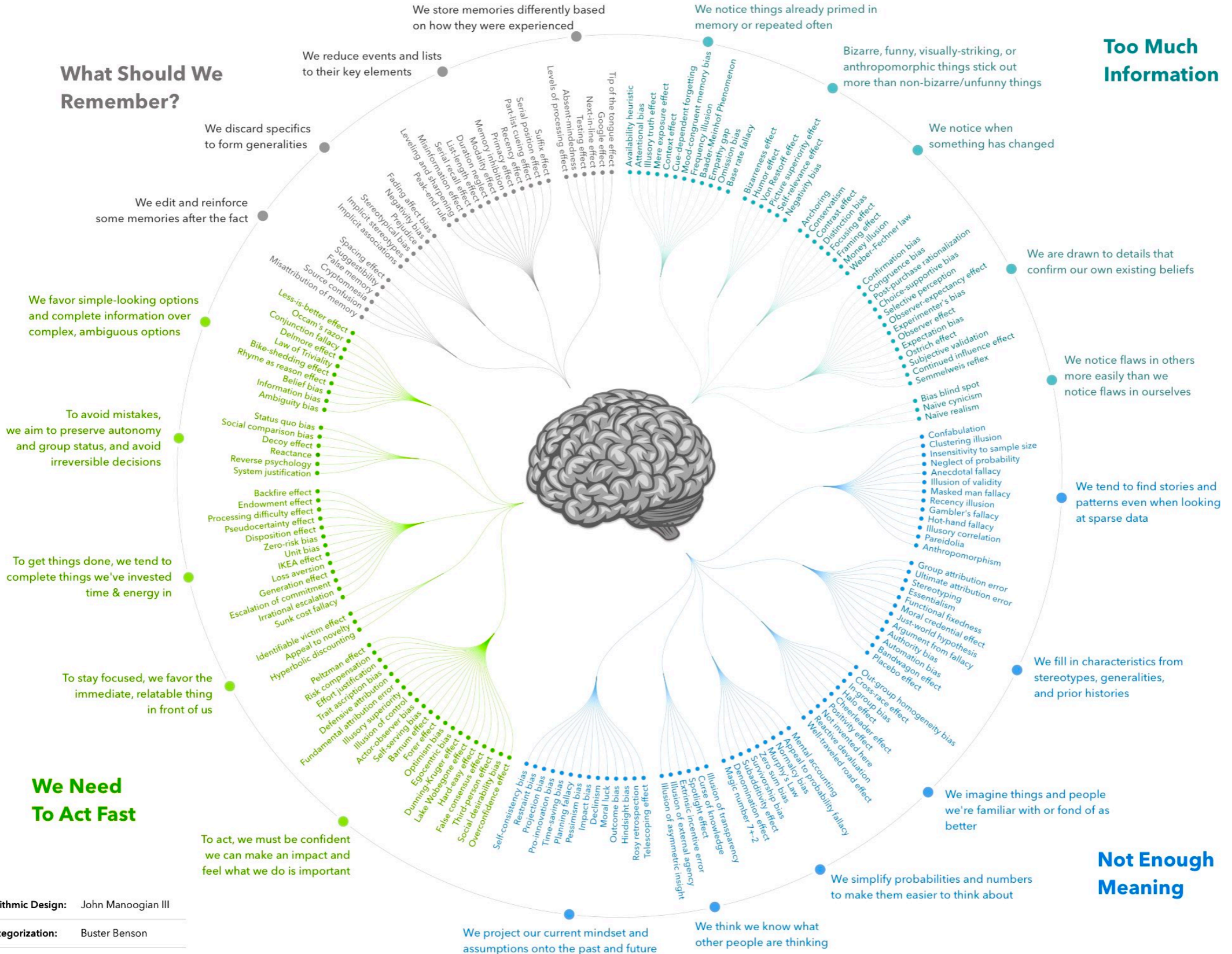
Data engineer



TO START
PRESS ANY KEY



COGNITIVE BIAS CODEX



Visual & Algorithmic Design: John Manoogian III

Concept & Categorization: Buster Benson

List of 188 Cognitive Biases: Wikipedia