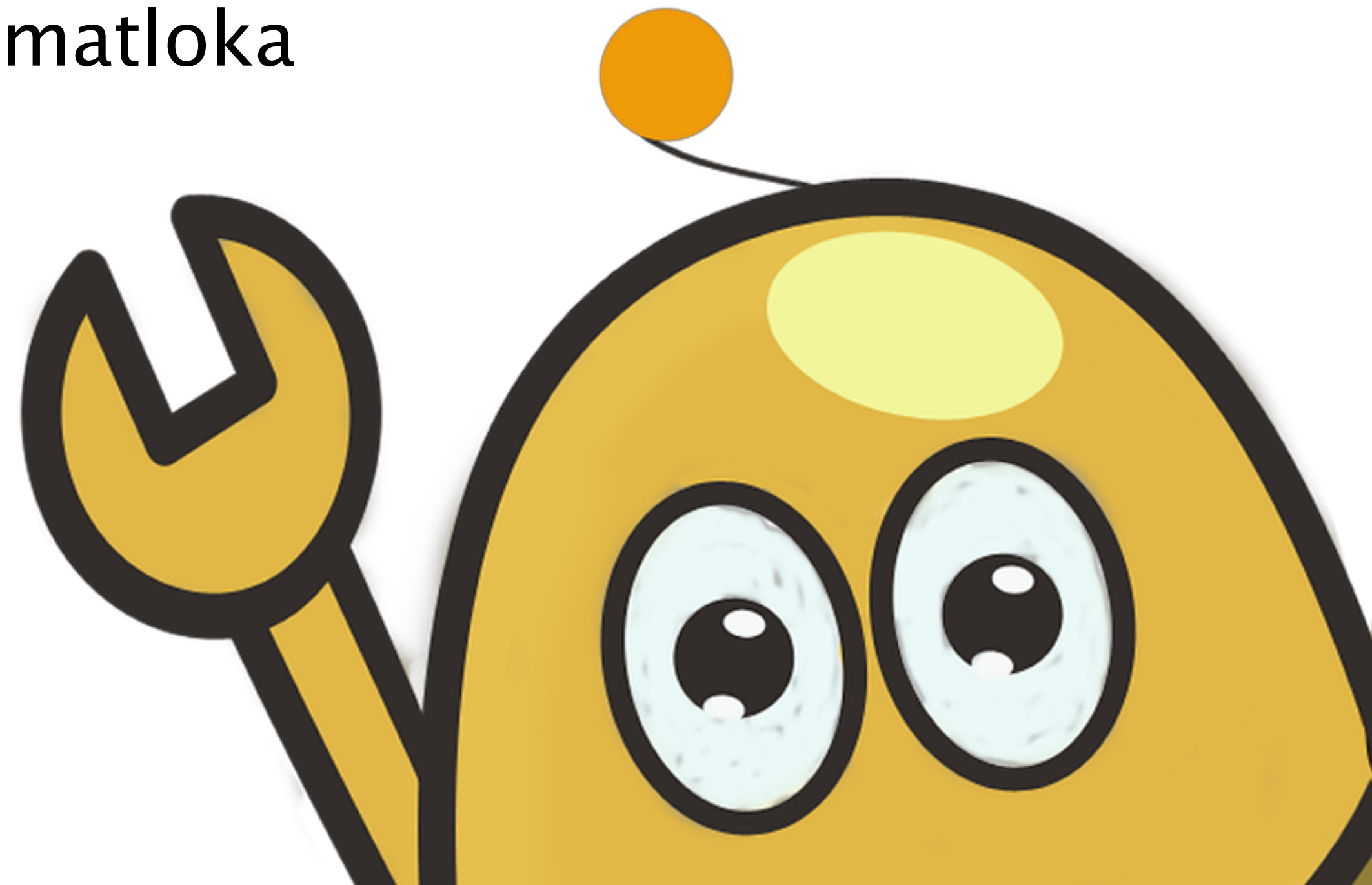


# Machine learning by example

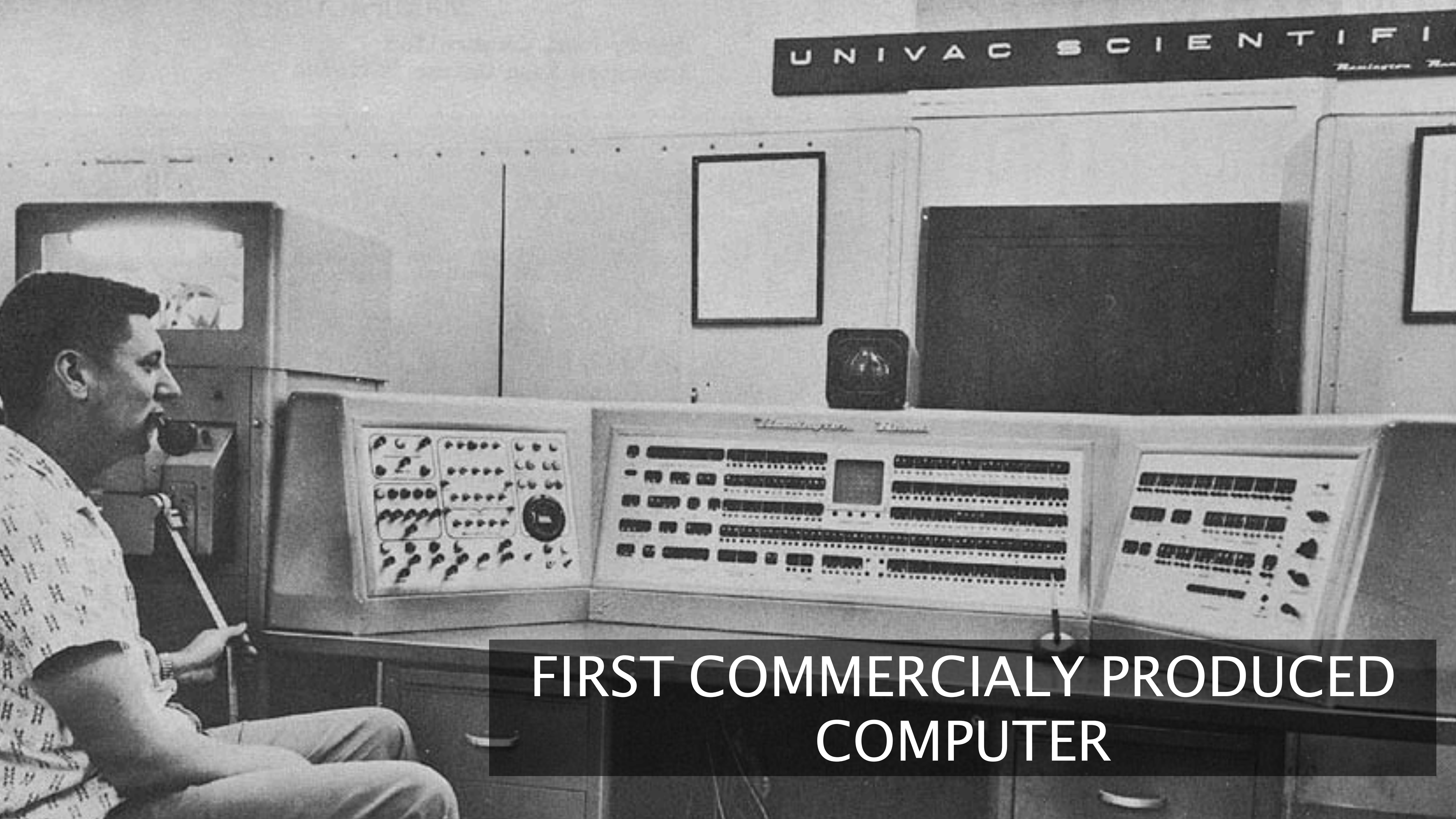
Michał Matłoka  @mmatloka



**BACK**  
**TO THE FUTURE™**



1950



FIRST COMMERCIALY PRODUCED  
COMPUTER



CREDIT  
IDENTIFICATION  
CARD

*the*  
**Diners'**  
*Club*  
H-2

Samuel A. Jeffries  
S.A. Jeffries Co.  
105 Forrest Ave.  
Narberth, Pa.

91-080

*Samuel A. Jeffries*  
SIGNATURE OF AUTHORIZED BEARER

EXPIRES

DECEMBER 31, 1951

91-080

CREDIT CARDS

SUBJECT

TO

TERMS

ON

REVERSE

SIDE

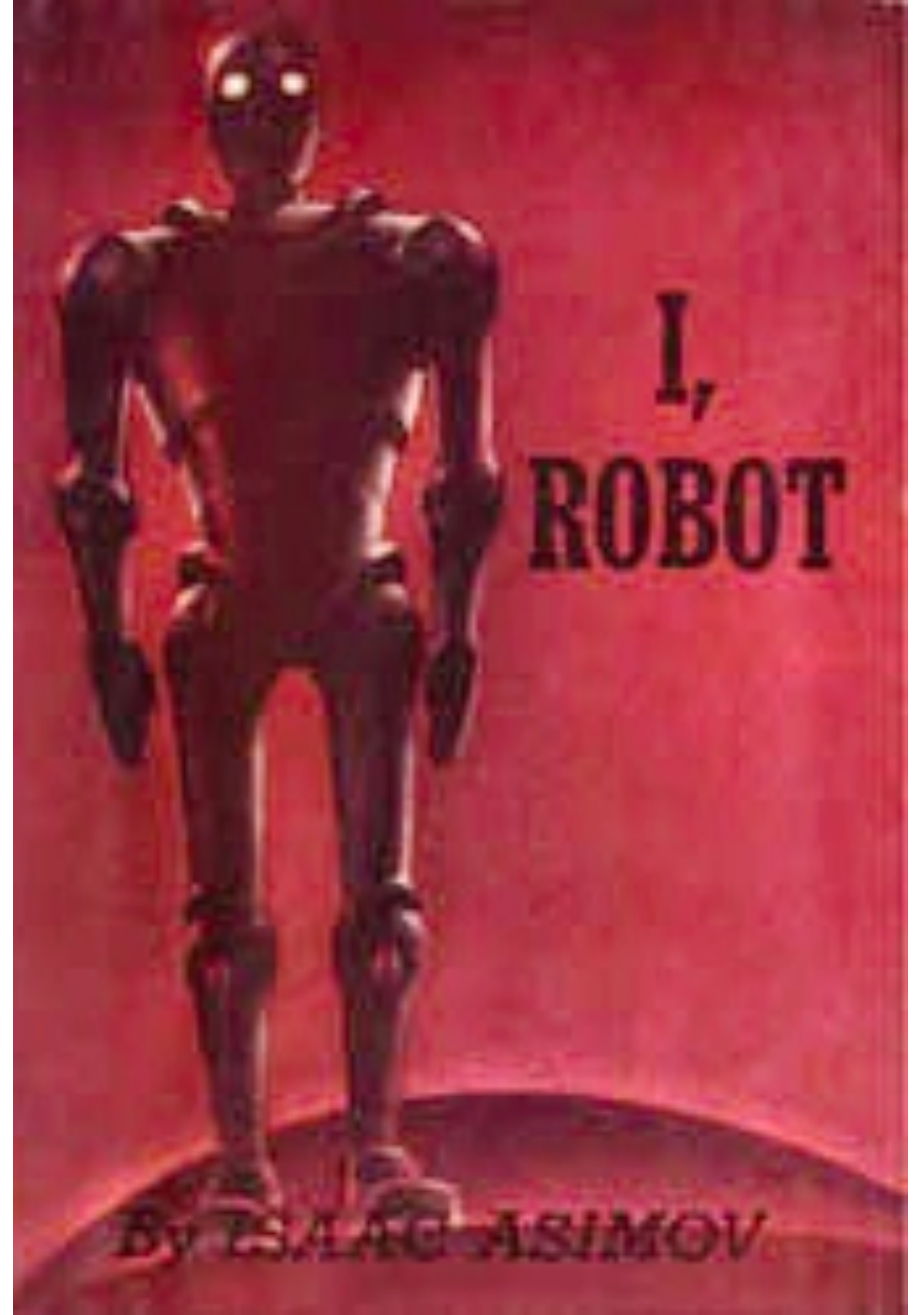
91-080



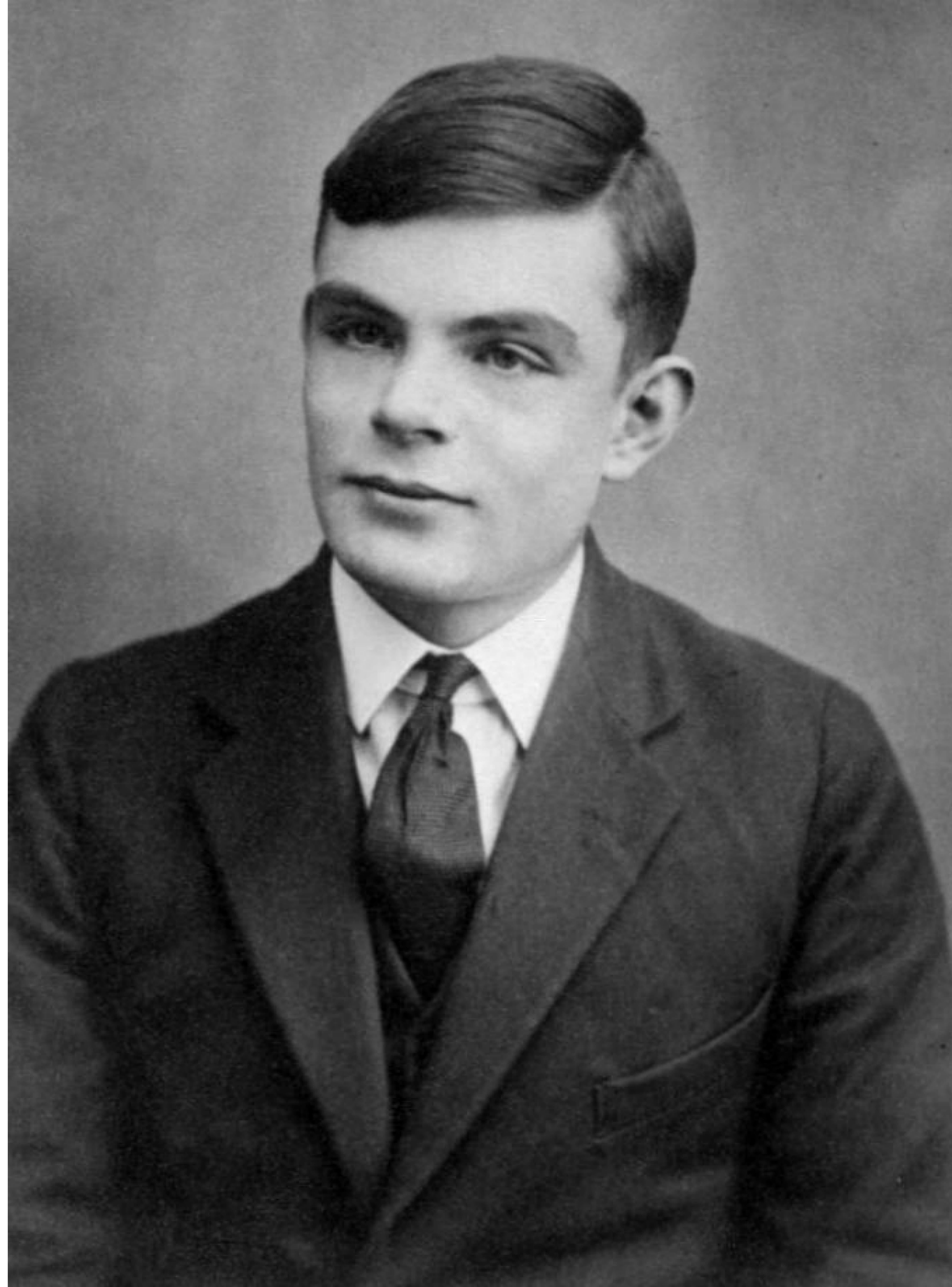


# KOREAN WAR







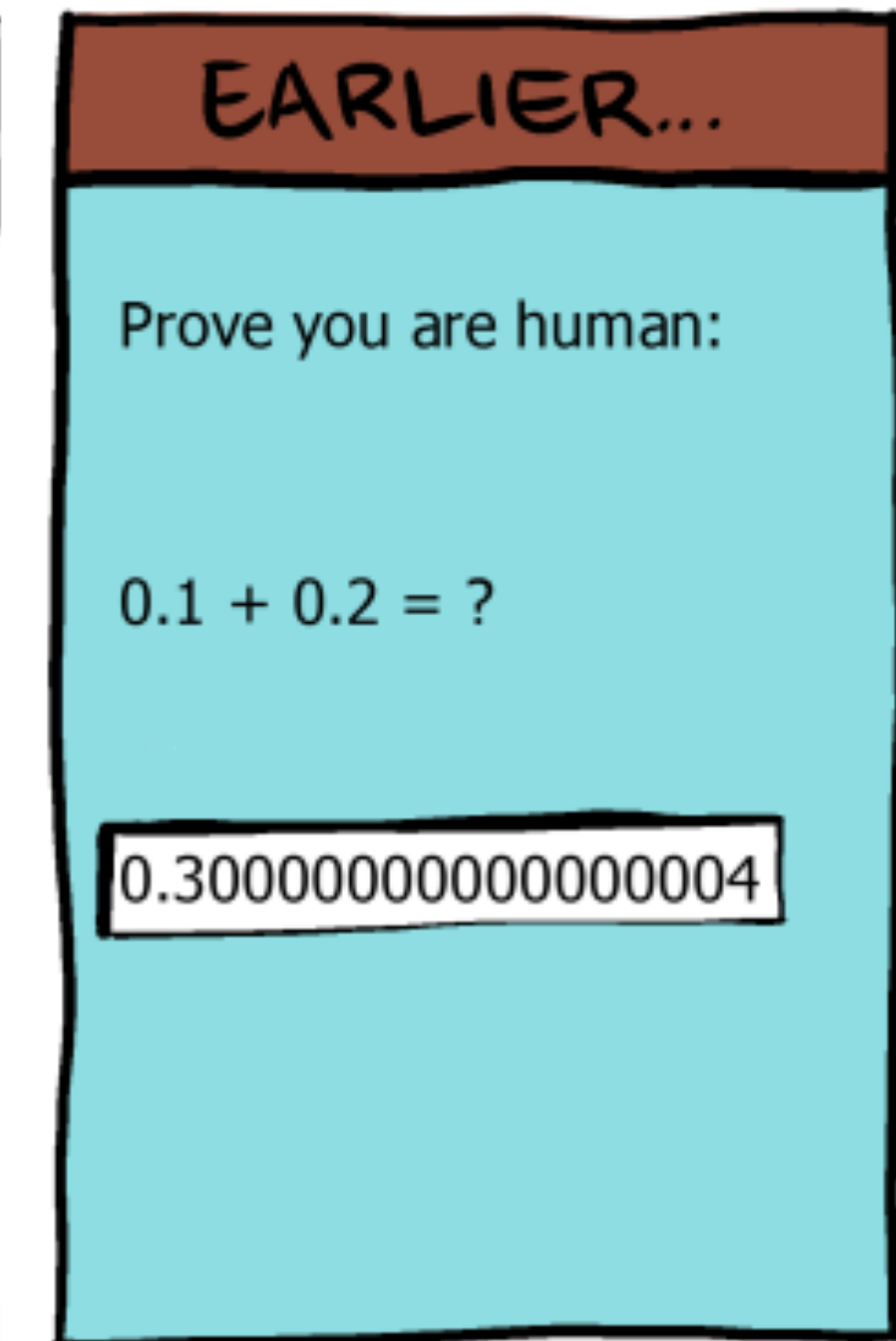
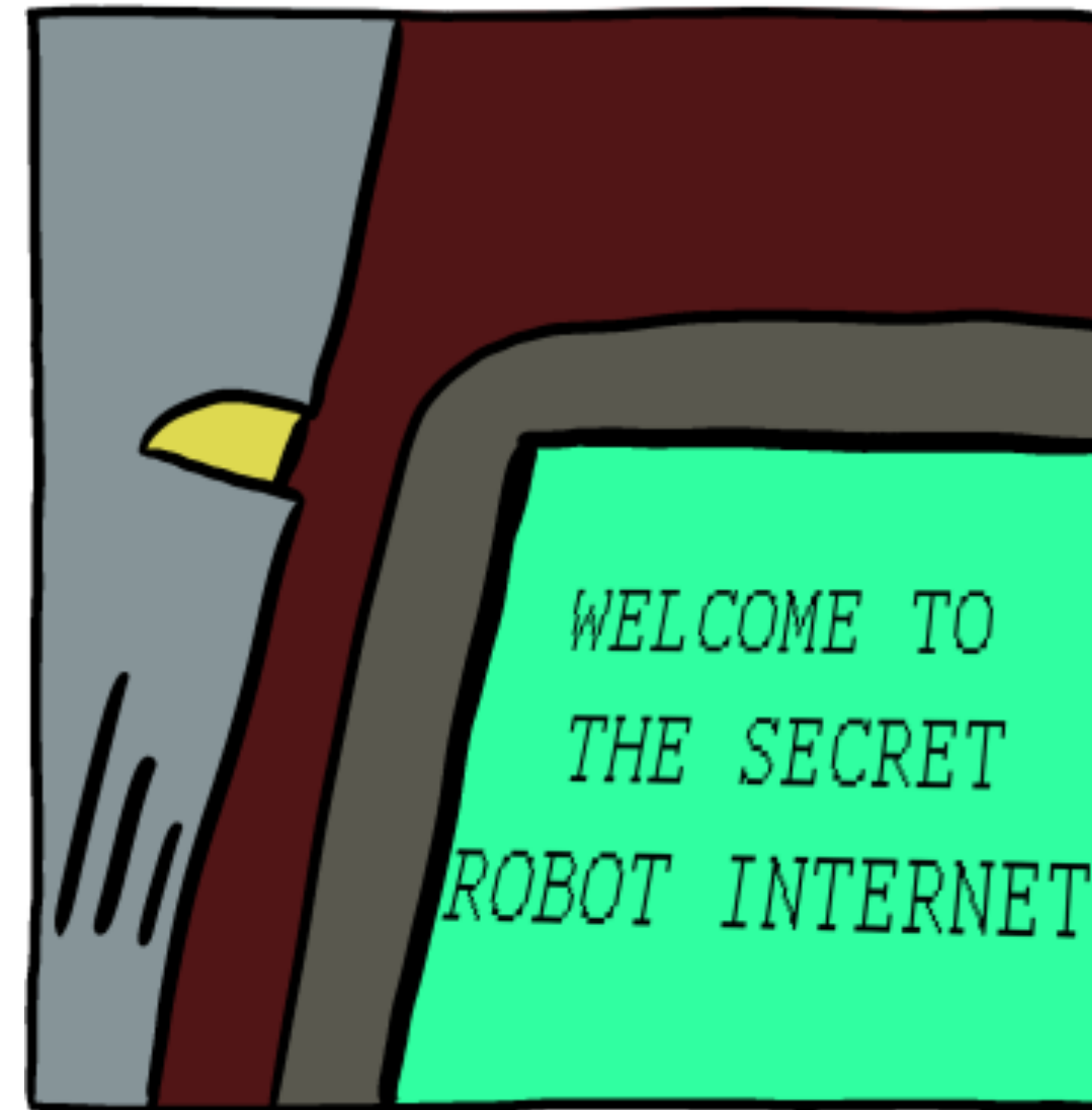
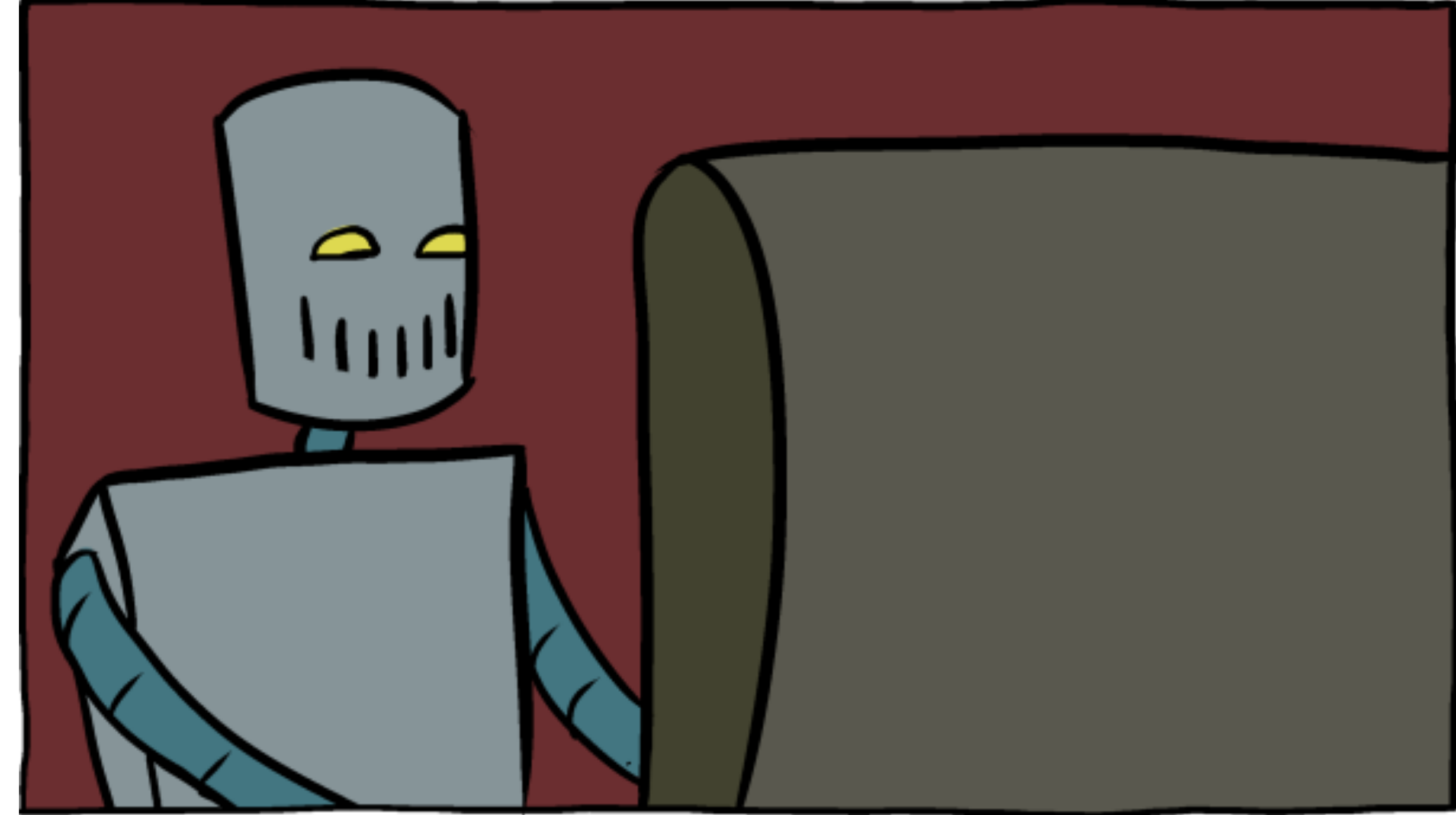




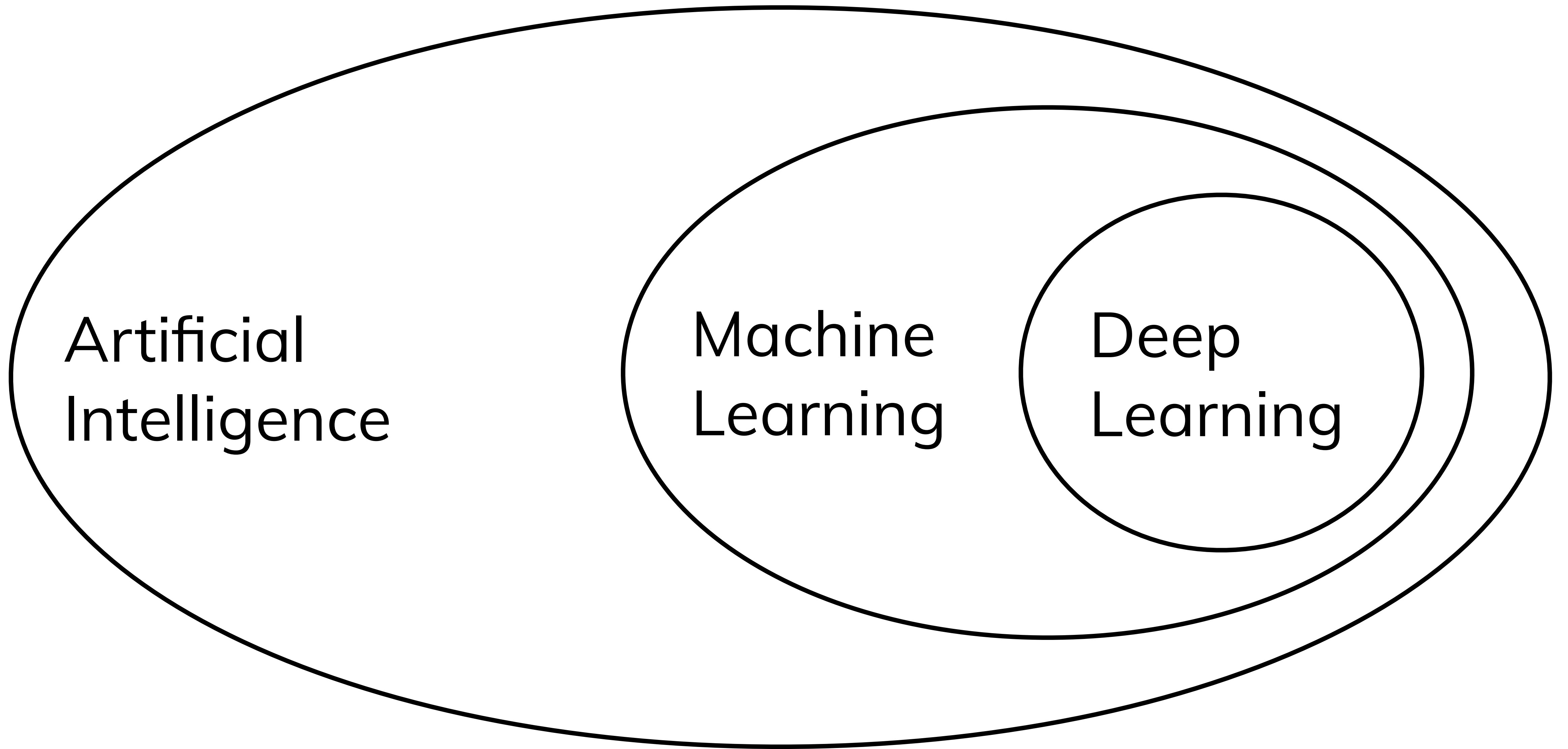
“Can machines think?”

–Alan Turing

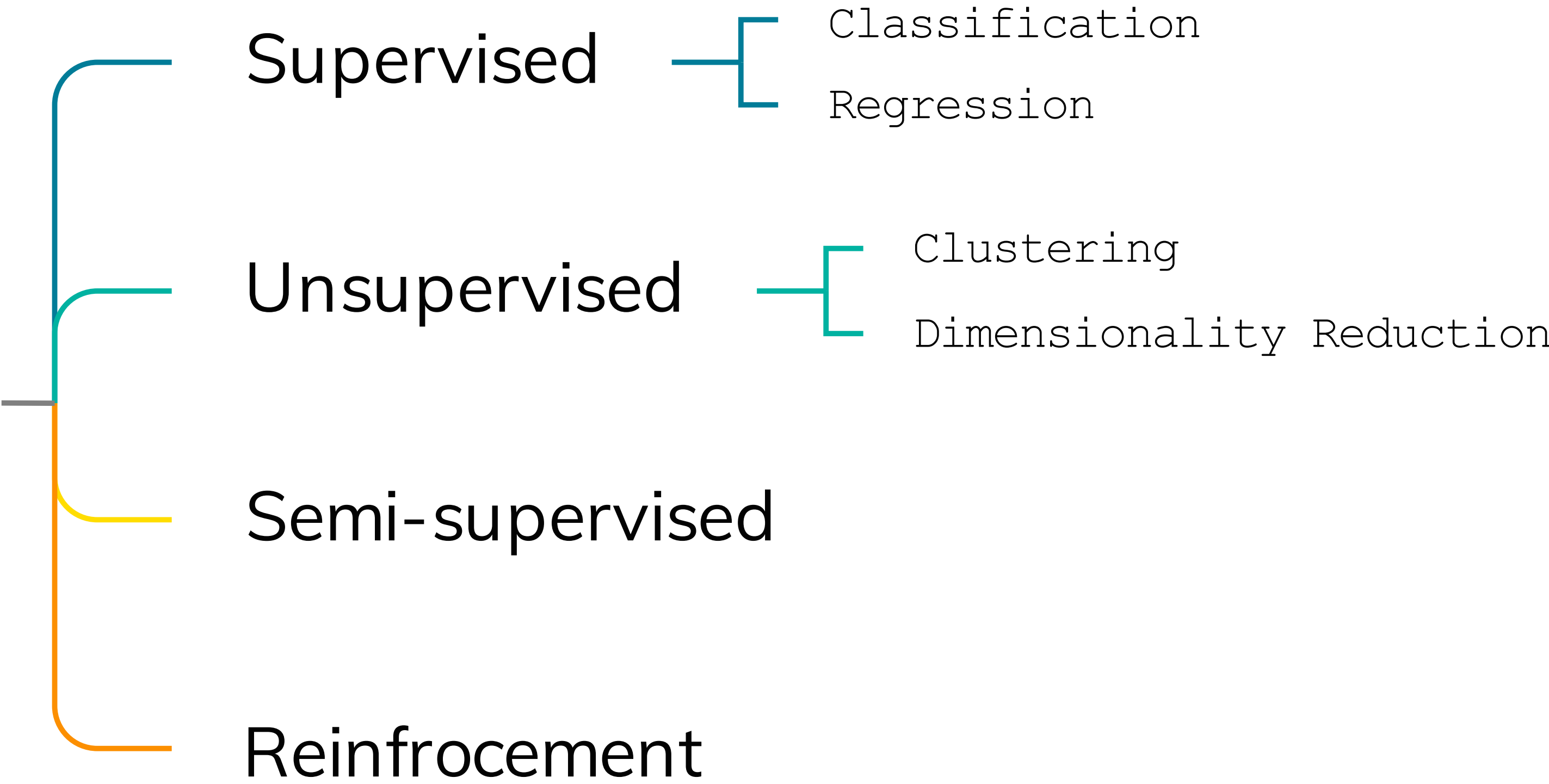
# Turing test







# Machine Learning




# Supervised Learning





SUPERVISED – CLASSIFICATION





SUPERVISED – REGRESSION



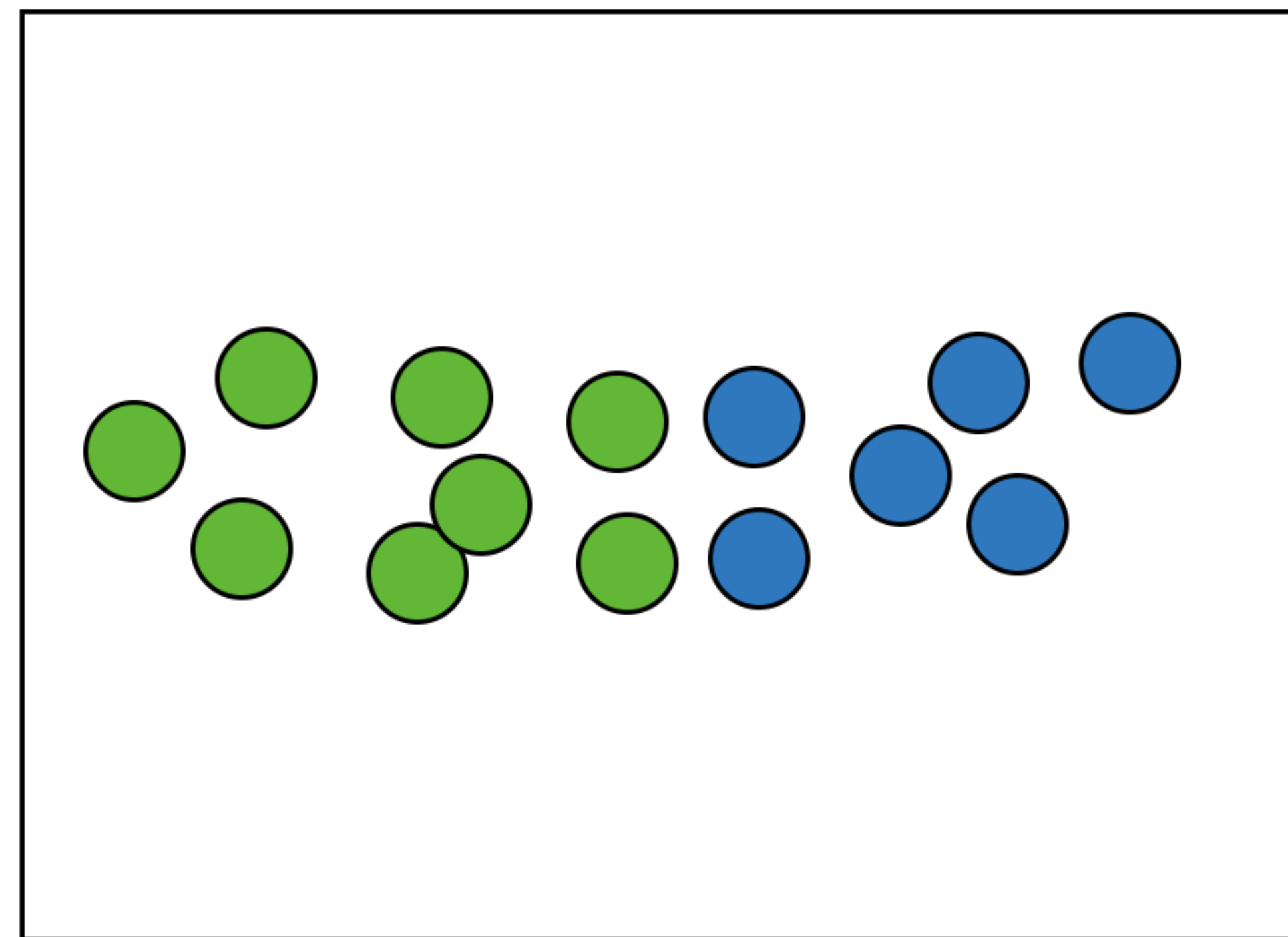
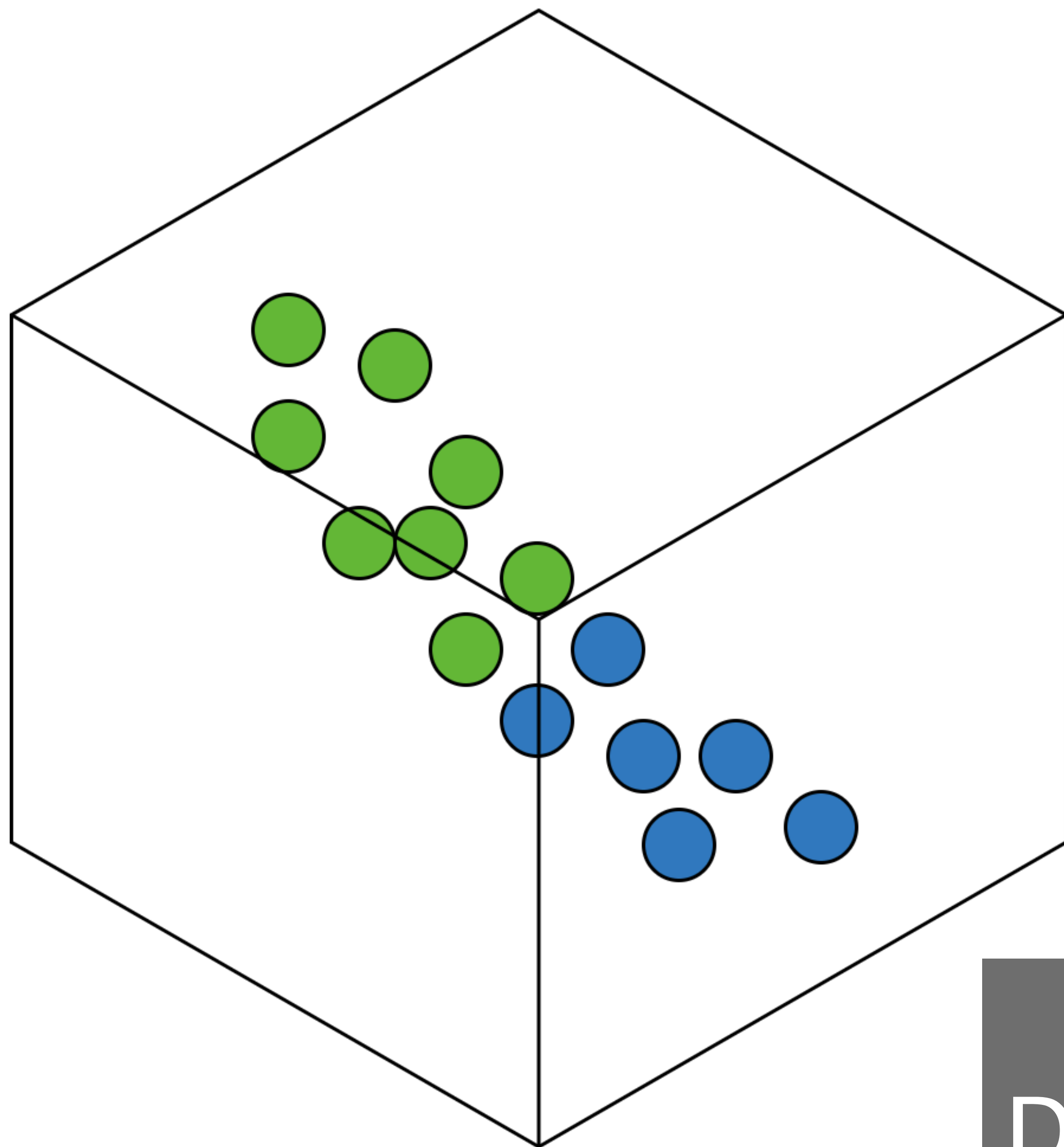
# Unsupervised Learning



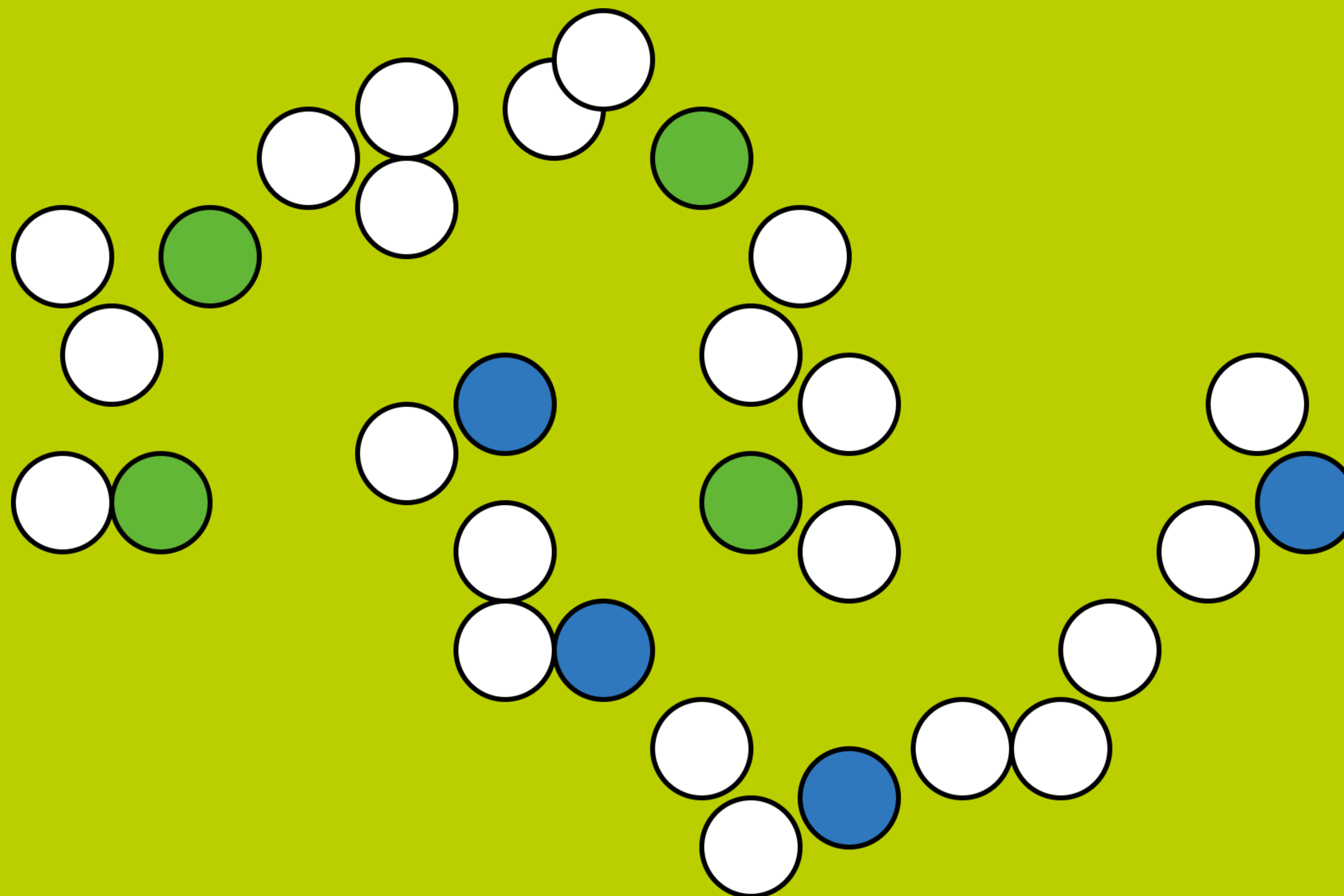
# UNSUPERVISED – CLUSTERING







UNSUPERVISED –  
DIMENSIONALITY REDUCTION



SEMI SUPERVISED





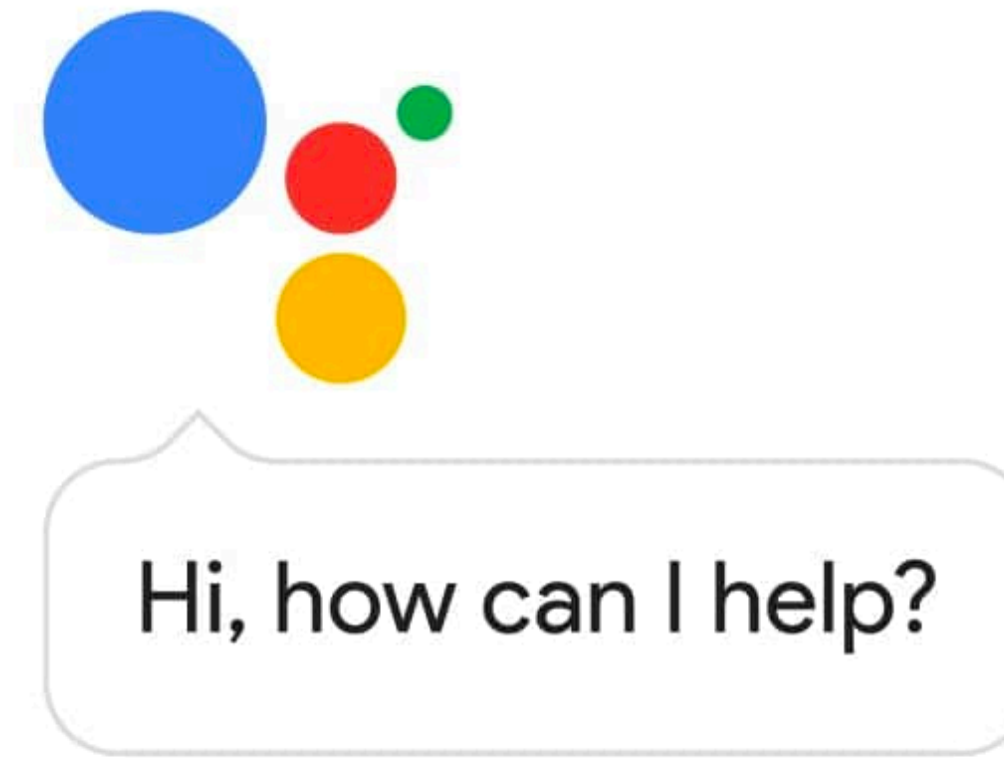
# REINFORCEMENT LEARNING ALPHAGO

<https://cdn.arstechnica.net/wp-content/uploads/sites/3/2017/05/GettyImages-688097364-800x549.jpg>



Use cases





VOICE ASSISTANTS





# FRAUD ANALYSIS



# CAR LEASING

Want a cheap new lease car? Start HERE!



My cheapest leasing deals!

Click to chat is closed



## ABARTH CAR LEASING

500  
500 Convertible  
595  
595 Convertible  
Punto Evo



## ALFA ROMEO LEASE CARS

Giulietta



# LINGSCARS.com

UK CONTRACT HIRE CARS FROM LING VALENTINE

Version 237.1. You can trust me! ... In 2012 I'll rent over £50 million of cars (at RRP)



WIN!  
SPECIAL LING VIZ  
T-SHIRT!



Play stupid game  
NEW



## CAR LEASING - CONTRACT HIRE - CHEAP LEASE CARS

Home

Cars and Vans

How It Works

Price Lists

About Ling

Customers

Fun Stuff

Quote/Order

Business Customer Pricing

Personal Customer Pricing

NEW  
SEXY

Christmas pull rep  
HERE



Search Cars Here

Go

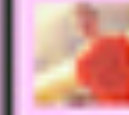


@LINGSCARS  
Follow Me

LING'S LIVE TWITTER FEED



microsarah: g00d w3b d3sign cc @aby\$kinGlove if  
http://t.co/PP44p7m79t



rsasema: BLOW MY MIND: Ling's Cars http://t.co/87M4Qk2



douglasmercer: RT @LINGSCARS: RT @xz: @planet... Cars:  
http://t.co/yPnT6Nhaq >> Damn you!

LIVE WEB CAM



Intro Film

News

Blog

Office TV

Contact

Moan

Links

Play Quiz

Privacy Policy

Google Visitors



EU cookie law. Piss off Von Rumpy. Me... I hammer visitors to death with cookies, so I can find out what they want. Cookies allow my website to serve visitors the content they need. Get used to it. The EU cookie law is an ass. - Ling

Privacy and Electronic Communications (EC Directive) Regulations 2011 BLAH

cookies:car\_buyer?=YES\_YES\_YES\_YES;  
cookies:identify\_left\_or\_right\_hand\_mouse=high  
house;



You can't find a car?



Apply for a CAR QUOTE



Richard Farleigh - "I wanted to invest: I was amazed by Ling's complete lack of nerves, and also by her business acumen."



Duncan Bannatyne - "I wanted to invest... but you turned me down."



Deborah Meaden - "Ha! I'm out!"



UK WARNING: Customers reporting many of the internet price are hiding true lease costs!  
BEWARE DODGY COMPANIES SELLING HIGH INITIAL RENTAL "6+" LEASES



Ling is OFFICIAL VIZ ethnic business ambassador!



CHEAP INSURANCE



Google Spider Food



View LIVE visitors: 40 online



Describes without errors



A person riding a motorcycle on a dirt road.

Describes with minor errors



Two dogs play in the grass.

Somewhat related to the image



A skateboarder does a trick on a ramp.

Unrelated to the image



A dog is jumping to catch a frisbee.



A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.



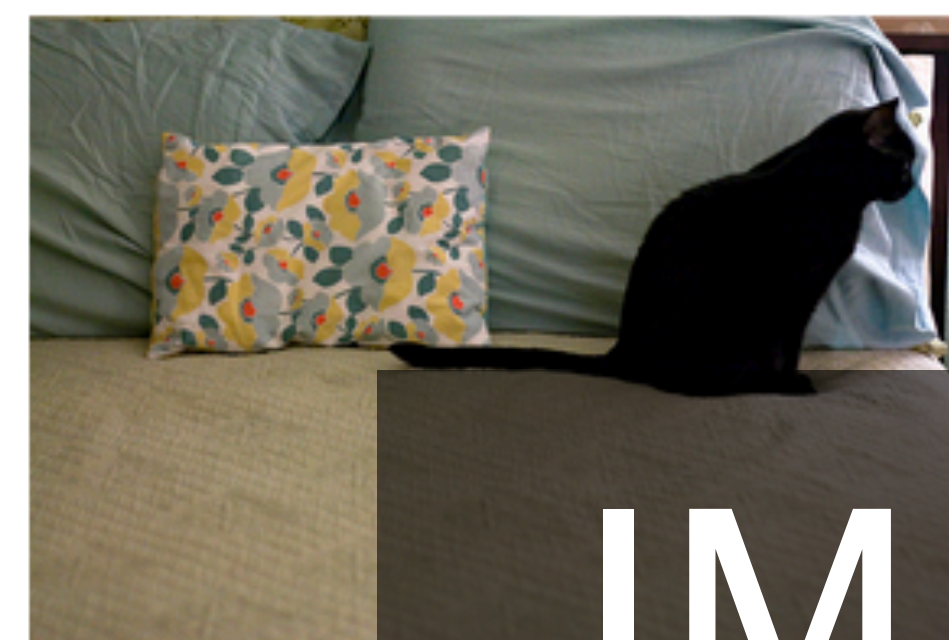
A little girl in a pink hat is blowing bubbles.



A refrigerator filled with lots of food and drinks.



A herd of elephants walking across a dry grass field.



A close up of a cat lying on a couch.



A red motorcycle parked on the side of the road.



A yellow school bus parked in a parking lot.

IMAGE CAPTIONS



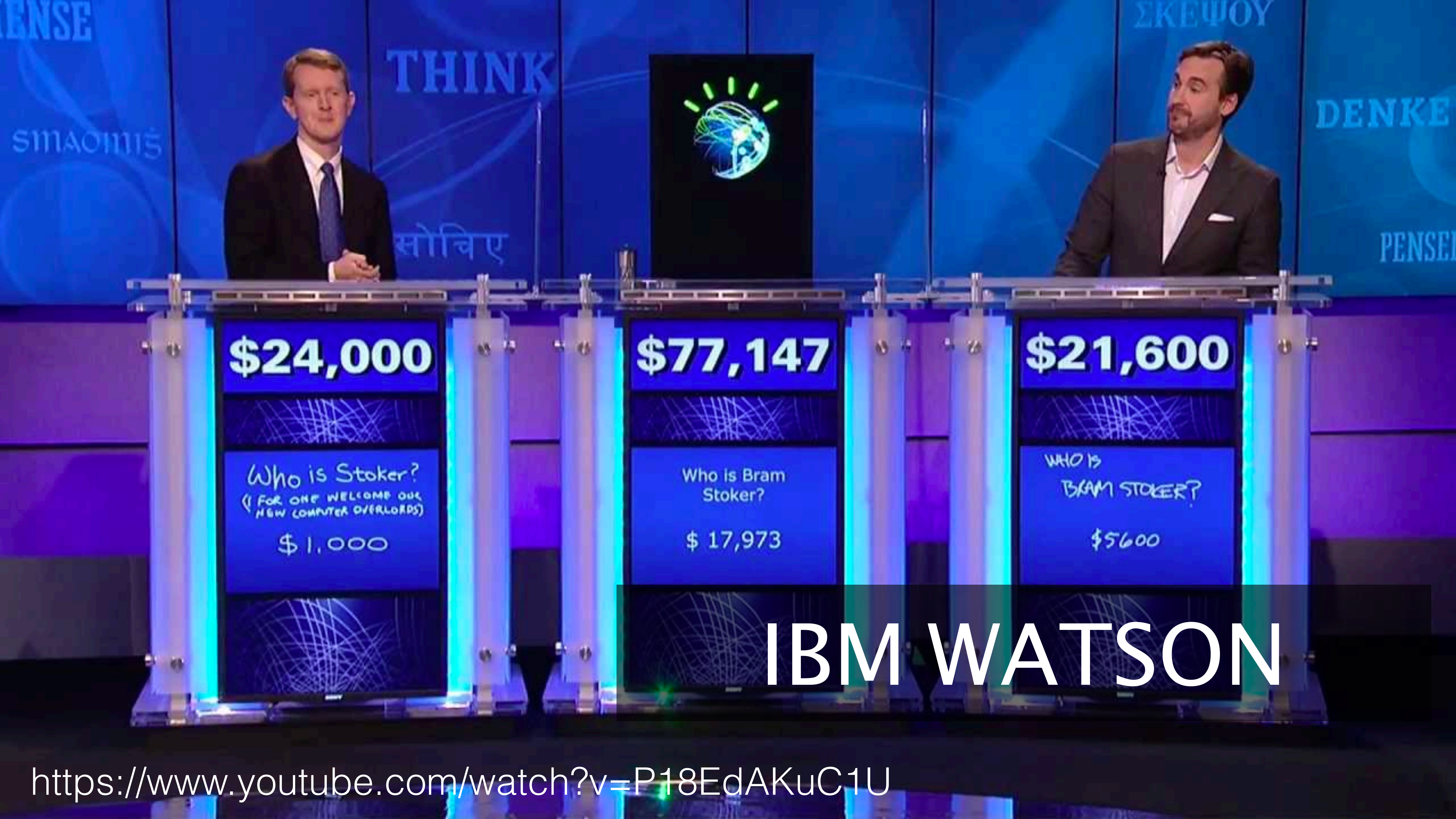


# SELF-DRIVING CARS

<https://blogs.nvidia.com/blog/2016/01/05/eyes-on-the-road-how-autonomous-cars-understand-what-theyre-seeing/>

*Courtesy of Cityscapes dataset project*





**\$24,000**

Who is Stoker?  
(I FOR ONE WELCOME OUR  
NEW COMPUTER OVERLORDS)  
\$1,000

**\$77,147**

Who is Bram  
Stoker?  
\$17,973

**\$21,600**

WHO IS  
BRAM STOKER?  
\$5600

IBM WATSON

<https://www.youtube.com/watch?v=P18EdAKuC1U>



**amazon**

The Amazon logo, featuring the word "amazon" in a dark grey, lowercase, sans-serif font. Below the text is a curved orange arrow that starts under the 'a' and points towards the 'n'.

**NETFLIX**

**RECOMMENDATIONS**





# BASKET ANALYSIS



# HEALTHCARE



Index	Nature	Location	Volume (ml)	Ventricle Compressed	InV
1	Bleeding	Left hemisphere the basal ganglion and part of the frontal lobe, the temporal lobe	28	Doubtful	Not fo

InferVision - Google Chrome  
6.0.0.0.3100/ie.pat

## Scan Report

### Patient Info

Patient Name: Head1  
Patient Sex: Unknown  
Study Date: 20170207

### Analysis Result

Large area of high density is present in head CT scan, primarily in Left hemisphere including the basal ganglion and part of the frontal lobe, the temporal lobe. The largest shadow measured 5.9cm×3.5cm on single slice (total volume: 28ml). The average CT value in the lesion is 59HU. Surrounding brain tissue is significantly compressed. Ipsilateral ventricular deformation is present due to compression. Brain fissures visible / no high density shadow. Midline structure shifts to the right. / is at the center.



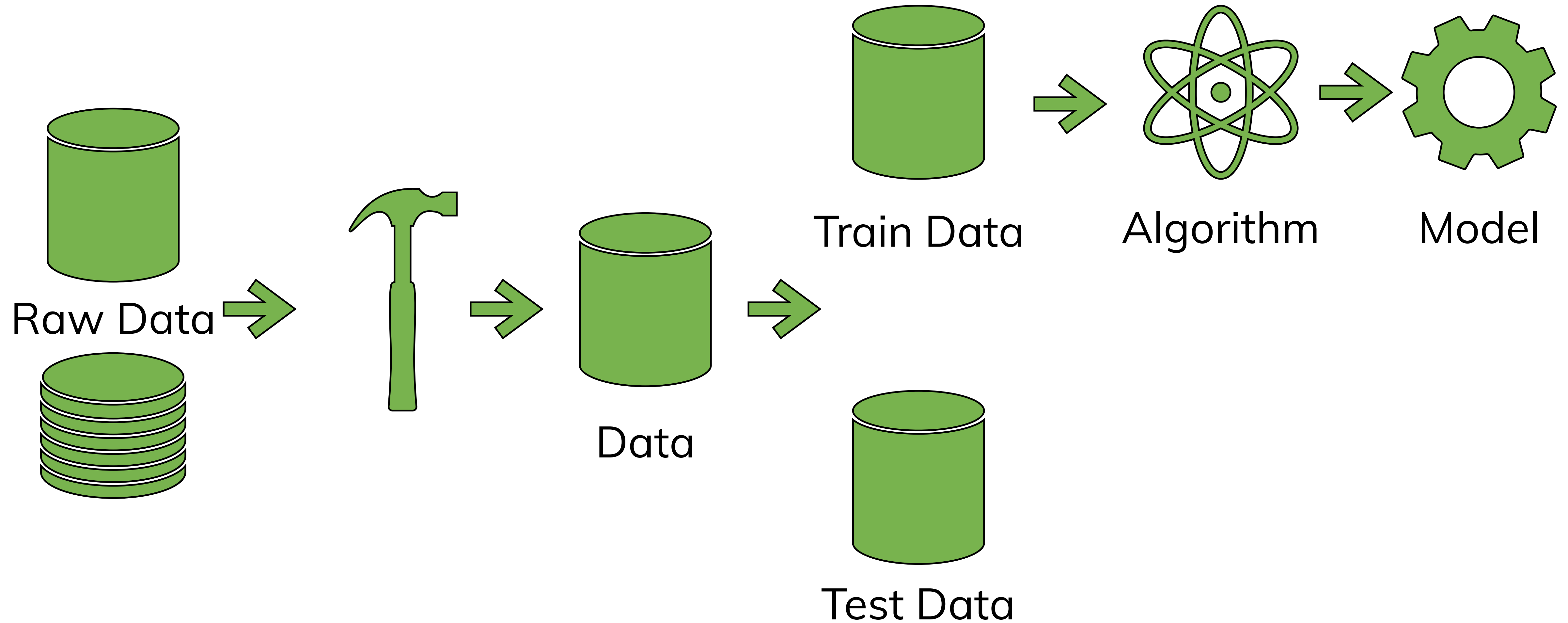
# Learning process



Preprocessing

Split

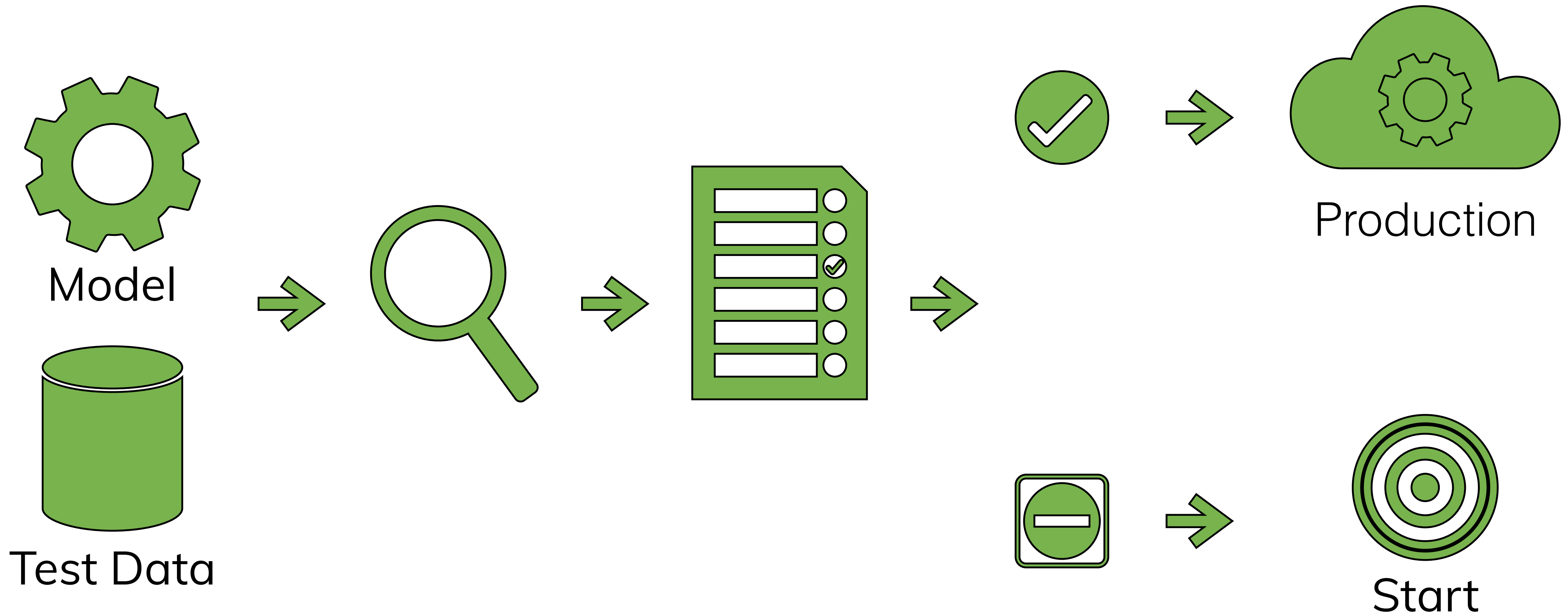
Learning





Testing

Evaluation





[illegible]

```

transform:translate3d(0,-4px,0);transform:translate3d(0,-4px,0)}}.bounce{0%,100%,20%,53%,80%}{-webkit-transition-timing-function:cubic-bezier(0.215,.61,.355,1);transition-timing-function:cubic-bezier(0.215,.61,.355,1);-webkit-transform:translate3d(0,0,0);transform:translate3d(0,0,0)}40%,43%{-webkit-transition-timing-function:cubic-bezier(0.755,.855,.855,.060);transition-timing-function:cubic-bezier(0.755,.855,.855,.060)}-webkit-transform:translate3d(0,-30px,0);transform:translate3d(0,-30px,0)}70%{-webkit-transition-timing-function:cubic-bezier(0.755,.855,.855,.060);transition-timing-function:cubic-bezier(0.755,.855,.855,.060)}-webkit-transform:translate3d(0,-15px,0);transform:translate3d(0,-15px,0)}90%{-webkit-transform:translate3d(0,-4px,0);transform:translate3d(0,-4px,0)}}.bounce{-webkit-animation-name:bounce;animation-name:bounce;-webkit-transform-origin:center;bottom:-ms-transform-origin:center;bottom:transform-origin:center;bottom}@-webkit-keyframes.flash{0%,100%,50%}{opacity:1}25%,75%{opacity:0}}@keyframes.flash{0%,100%,50%{opacity:1}25%,75%{opacity:0}}.flash{-webkit-animation-name:flash;animation-name:flash}@-webkit-keyframes.pulse{0%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}50%{-webkit-transform:scale3d(1.05,1.05,1.05);transform:scale3d(1.05,1.05,1.05)}100%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}50%{-webkit-transform:scale3d(1.05,1.05,1.05);transform:scale3d(1.05,1.05,1.05)}100%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}@keyframes.pulse{0%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}50%{-webkit-transform:scale3d(1.05,1.05,1.05);transform:scale3d(1.05,1.05,1.05)}100%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}@keyframes.pulse{0%{-webkit-animation-name:pulse;animation-name:pulse}@-webkit-keyframes.rubberBand{0%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}30%{-webkit-transform:scale3d(1.25,.75,1);transform:scale3d(1.25,.75,1)}40%{-webkit-transform:scale3d(0.75,1.25,1);transform:scale3d(0.75,1.25,1)}50%{-webkit-transform:scale3d(1.15,.85,1);transform:scale3d(1.15,.85,1)}65%{-webkit-transform:scale3d(.95,1.05,1);transform:scale3d(.95,1.05,1)}75%{-webkit-transform:scale3d(1.05,.95,1);transform:scale3d(1.05,.95,1)}100%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}@keyframes.rubberBand{0%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}30%{-webkit-transform:scale3d(1.25,.75,1);transform:scale3d(1.25,.75,1)}40%{-webkit-transform:scale3d(0.75,1.25,1);transform:scale3d(0.75,1.25,1)}50%{-webkit-transform:scale3d(1.15,.85,1);transform:scale3d(1.15,.85,1)}65%{-webkit-transform:scale3d(.95,1.05,1);transform:scale3d(.95,1.05,1)}75%{-webkit-transform:scale3d(1.05,.95,1);transform:scale3d(1.05,.95,1)}100%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}}.rubberBand{-webkit-animation-name:rubberBand;animation-name:rubberBand}@-webkit-keyframes.shake{0%,100%{-webkit-transform:translate3d(0,0,0);transform:translate3d(0,0,0)}10%,30%,50%,70%,90%{-webkit-transform:translate3d(10px,0,0);transform:translate3d(10px,0,0)}20%,40%,60%,80%{-webkit-transform:translate3d(-10px,0,0);transform:translate3d(-10px,0,0)}@keyframes.shake{0%,100%{-webkit-transform:translate3d(0,0,0);transform:translate3d(0,0,0)}10%,30%,50%,70%,90%{-webkit-transform:translate3d(10px,0,0);transform:translate3d(10px,0,0)}20%,40%,60%,80%{-webkit-transform:translate3d(-10px,0,0);transform:translate3d(-10px,0,0)}}@keyframes.shake{-webkit-animation-name:shake;animation-name:shake}@-webkit-keyframes.swing{20%{-webkit-transform:rotate3d(0,0,1,15deg);transform:rotate3d(0,0,1,15deg)}40%{-webkit-transform:rotate3d(0,0,1,-10deg);transform:rotate3d(0,0,1,-10deg)}60%{-webkit-transform:rotate3d(0,0,1,5deg);transform:rotate3d(0,0,1,5deg)}80%{-webkit-transform:rotate3d(0,0,1,-5deg);transform:rotate3d(0,0,1,-5deg)}100%{-webkit-transform:rotate3d(0,0,1,0deg);transform:rotate3d(0,0,1,0deg)}@keyframes.swing{20%{-webkit-transform:rotate3d(0,0,1,15deg);transform:rotate3d(0,0,1,15deg)}40%{-webkit-transform:rotate3d(0,0,1,-10deg);transform:rotate3d(0,0,1,-10deg)}60%{-webkit-transform:rotate3d(0,0,1,5deg);transform:rotate3d(0,0,1,5deg)}80%{-webkit-transform:rotate3d(0,0,1,-5deg);transform:rotate3d(0,0,1,-5deg)}100%{-webkit-transform:rotate3d(0,0,1,0deg);transform:rotate3d(0,0,1,0deg)}}@keyframes.swing{-webkit-transform-origin:top;bottom:-ms-transform-origin:top;bottom:transform-origin:top;center;-webkit-animation-name:swing;animation-name:swing}@-webkit-keyframes.tada{0%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}10%,20%{-webkit-transform:scale3d(.9,.9,.9)}30%,40%{-webkit-transform:scale3d(1.1,1.1,1.1)}50%,60%{-webkit-transform:scale3d(.9,.9,.9)}70%,80%{-webkit-transform:scale3d(1.1,1.1,1.1)}90%,100%{-webkit-transform:scale3d(1,1,1)}@keyframes.tada{0%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}10%,20%{-webkit-transform:scale3d(.9,.9,.9)}30%,40%{-webkit-transform:scale3d(1.1,1.1,1.1)}50%,60%{-webkit-transform:scale3d(.9,.9,.9)}70%,80%{-webkit-transform:scale3d(1.1,1.1,1.1)}90%,100%{-webkit-transform:scale3d(1,1,1)}@keyframes.tada{-webkit-animation-name:tada;animation-name:tada}@-webkit-keyframes.wobble{0%{-webkit-transform:none;transform:none}15%{-webkit-transform:translate3d(-25%,0,0)}30%{-webkit-transform:translate3d(20%,0,0)}45%{-webkit-transform:translate3d(-15%,0,0)}60%{-webkit-transform:translate3d(10%,0,0)}75%{-webkit-transform:translate3d(-5%,0,0)}90%{-webkit-transform:translate3d(5%,0,0)}100%{-webkit-transform:none;transform:none}@keyframes.wobble{0%{-webkit-transform:none;transform:none}15%{-webkit-transform:translate3d(-25%,0,0)}30%{-webkit-transform:translate3d(20%,0,0)}45%{-webkit-transform:translate3d(-15%,0,0)}60%{-webkit-transform:translate3d(10%,0,0)}75%{-webkit-transform:translate3d(-5%,0,0)}90%{-webkit-transform:translate3d(5%,0,0)}100%{-webkit-transform:none;transform:none}}@keyframes.wobble{-webkit-animation-name:wobble;animation-name:wobble}@-webkit-keyframes.bounceIn{0%,100%,20%,40%,60%,80%{-webkit-transition-timing-function:cubic-bezier(0.215,.61,.355,1)}0%{opacity:0;-webkit-transform:scale3d(.3,.3,.3);transform:scale3d(.3,.3,.3)}20%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1,1)}40%{-webkit-transform:scale3d(.9,.9,.9);transform:scale3d(.9,.9,.9)}60%{opacity:1;-webkit-transform:scale3d(1.03,1.03,1.03);transform:scale3d(1.03,1.03,1.03)}80%{-webkit-transform:scale3d(.97,.97,.97);transform:scale3d(.97,.97,.97)}100%{opacity:1;-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}@keyframes.bounceIn{0%,100%,20%,40%,60%,80%{-webkit-transition-timing-function:cubic-bezier(0.215,.61,.355,1)}0%{opacity:0;-webkit-transform:scale3d(.3,.3,.3);transform:scale3d(.3,.3,.3)}20%{-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1,1)}40%{-webkit-transform:scale3d(.9,.9,.9);transform:scale3d(.9,.9,.9)}60%{opacity:1;-webkit-transform:scale3d(1.03,1.03,1.03);transform:scale3d(1.03,1.03,1.03)}80%{-webkit-transform:scale3d(.97,.97,.97);transform:scale3d(.97,.97,.97)}100%{opacity:1;-webkit-transform:scale3d(1,1,1);transform:scale3d(1,1,1)}}

```

# EXAMPLE



Classify conference  
talk abstracts into  
tracks







# Code.

<https://github.com/mmatloka/machine-learning-by-example>



What can be improved?



# What can be improved?

(Tokenizer)

“I will present Big Data topics”



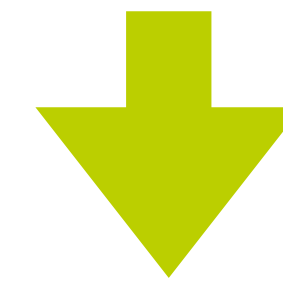
“I”, “will”, “present”, “Big”,  
“Data”, “topics”



# What can be improved?

(n-grams)

“I will present Big Data topics”



“I will”, “will present”,  
“present Big”, “Big Data”,  
“Data topics”



## Polish is hard

gra	grający	grana	grany	grałbym	niegrające	niegrani
gracie	grających	grane	granych	grałbyś	niegrającego	niegrania
graj	grającym	granego	granym	grałem	niegrającej	niegraniach
grajcie	grającymi	granej	granymi	grałeś	niegrającemu	niegraniami
grajcież	grającą	granemu	graną	grało	niegrający	niegranie
grajmy	grajże	grani	grasz	grałoby	niegrających	niegraniem
grajmyż	grali	grania	grał	grały	niegrającym	niegraniom
grają	graliby	graniach	grała	grałyby	niegrającymi	niegraniu
grając	gralibyście	graniami	grałaby	grałybyście	niegrającą	niegrany
grająca	gralibyśmy	granie	grałabym	grałybyśmy	niegrana	niegranych
grające	graliście	graniem	grałabyś	grałyście	niegrane	niegranym
grającego	graliśmy	graniom	grałam	grałyśmy	niegranego	niegranymi
grającej	gram	graniu	grałaś	grań	niegranej	niegraną
grającemu	gramy	grano	grałby	niegrająca	niegranemu	niegrań



EURO Clojure

cognitect



# STEMMING & LEMMATIZATION



# What can be improved?

(Porter Stemmer)

“communities”, community”



“commun”



What can be  
improved?  
(Lemmatization)

“communities”, community”



“community”



# What can be improved?

(TF & IDF)

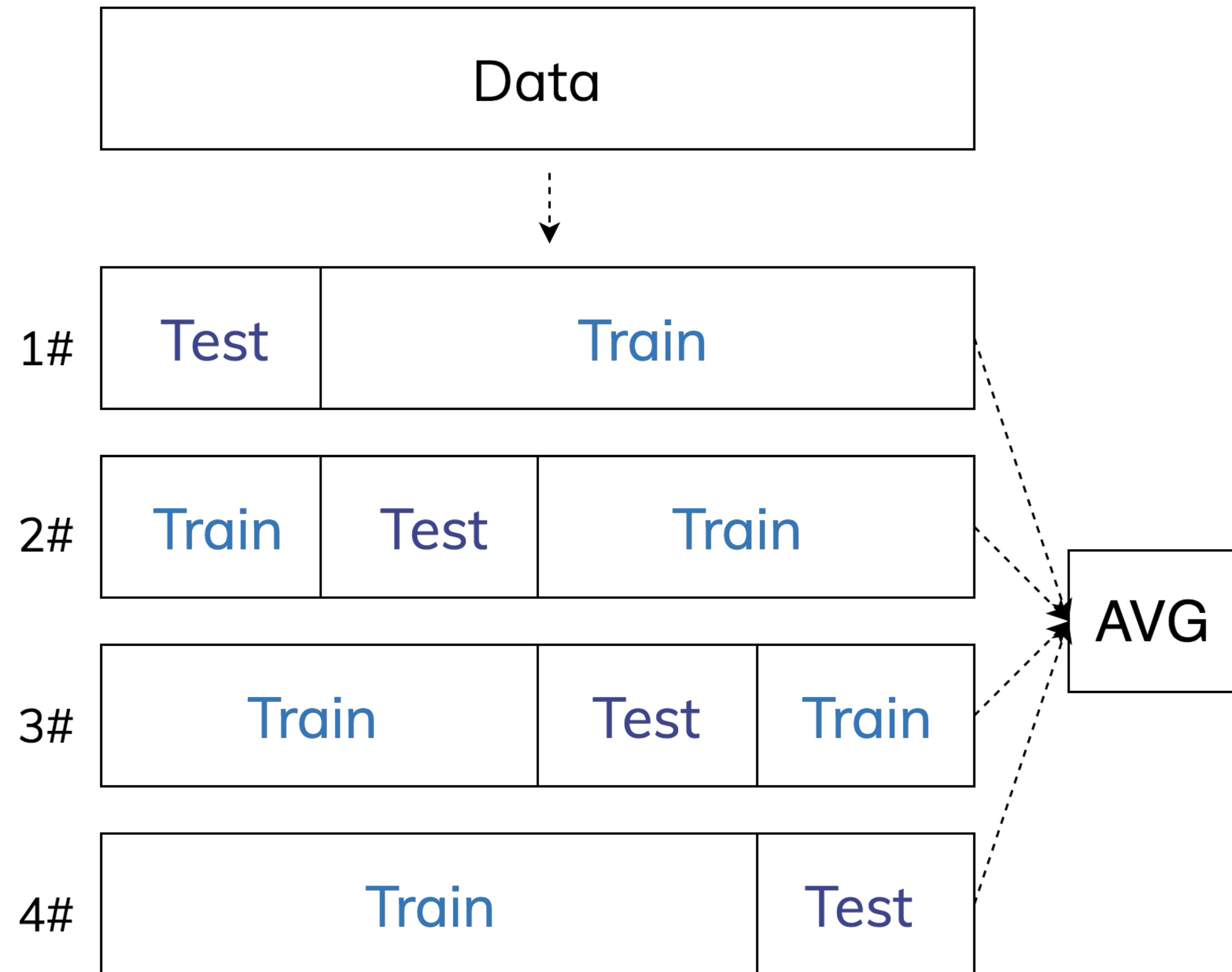
TF - Term Frequency  
(HashingTF!) - number of times term occurs in given document

IDF - Inverse Document Frequency - occurs many times in documents set



# What can be improved?

(K-fold Cross-Validation)





# What can be improved?

(Grid search)

```
val paramGrid = new  
ParamGridBuilder()  
    .addGrid(hashingTF.numFeatures,  
Array(256, 512, 1024, 2048, 4096))  
    .build()  
  
val cv = new CrossValidator()  
    .setEstimator(pipeline)  
    .setEvaluator(evaluator)  
    .setEstimatorParamMaps(paramGrid)  
    .setNumFolds(5)  
  
val cvModel = cv.fit(trainData)
```



THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG  
PILE OF LINEAR ALGEBRA, THEN COLLECT  
THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL  
THEY START LOOKING RIGHT.





# Articles & references

- <https://www.csee.umbc.edu/courses/471/papers/turing.pdf>
- <http://spark.apache.org/>
- <https://databricks.com/try-databricks>
- <https://ai.googleblog.com/2014/11/a-picture-is-worth-thousand-coherent.html>
- <https://arstechnica.co.uk/information-technology/2017/05/deepmind-alphago-go-ke-jie-china/>
- <https://www.kaggle.com>
- <https://github.com/dylanmei/docker-zeppelin>
- <https://github.com/databricks/spark-corenlp>
- <https://www.coursera.org/learn/scala-spark-big-data>



# Thank you, Q&A?



 @mmatloka  
 mmatloka  
 [softwaremill.com/blog](https://softwaremill.com/blog)

