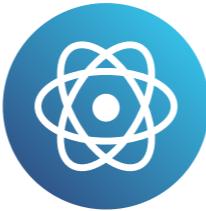


# Deep learning

MACHINE LEARNING FOR EVERYONE

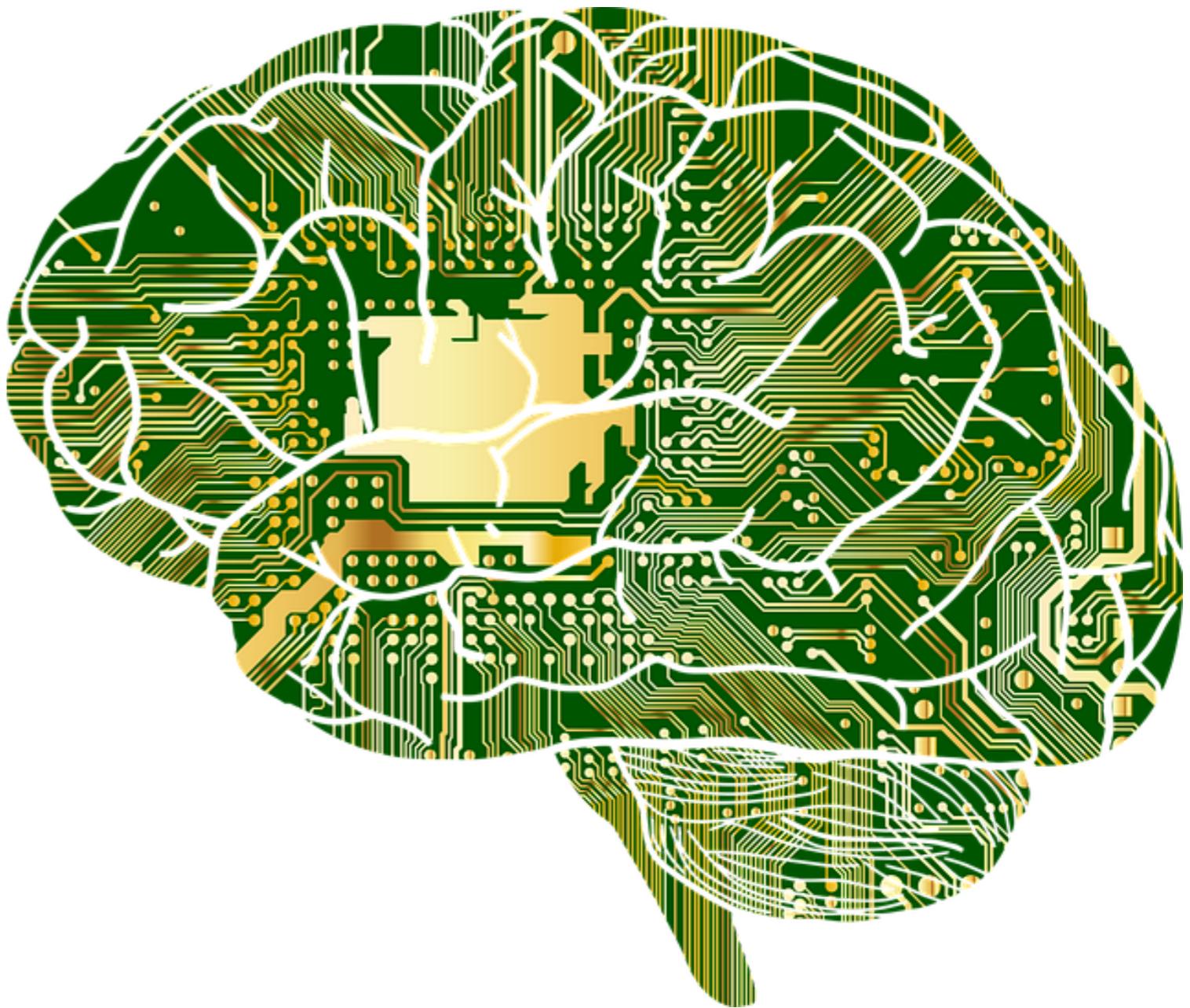


**Sara Billen**

Curriculum Manager, DataCamp

# What is deep learning?

- AKA: Neural Networks
  - Basic unit: neurons (nodes)
- Special area of Machine Learning
- Requires more data
- Best when inputs that are images or text



# Predicting box office revenue



# Predicting box office revenue

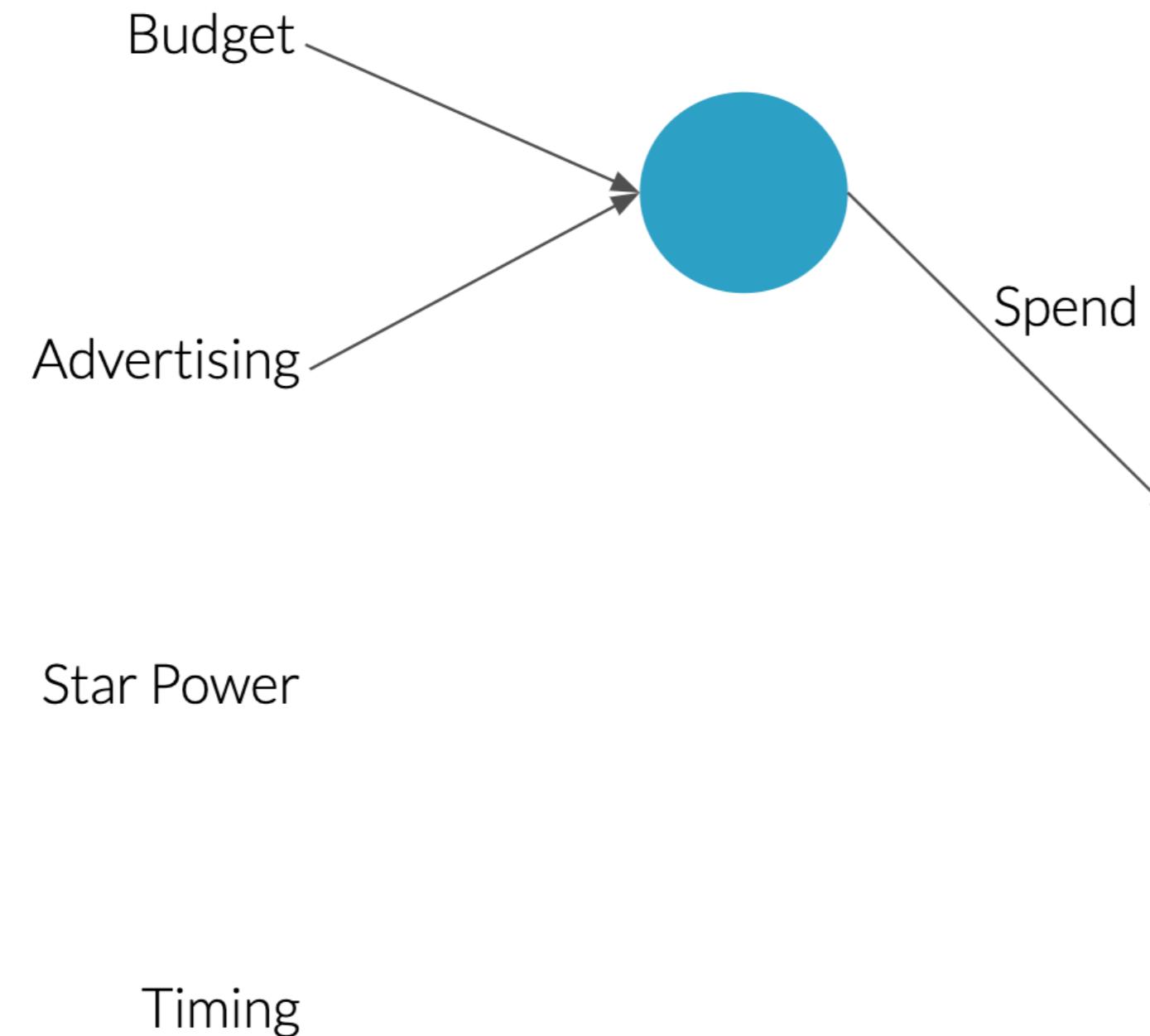
Budget

Advertising

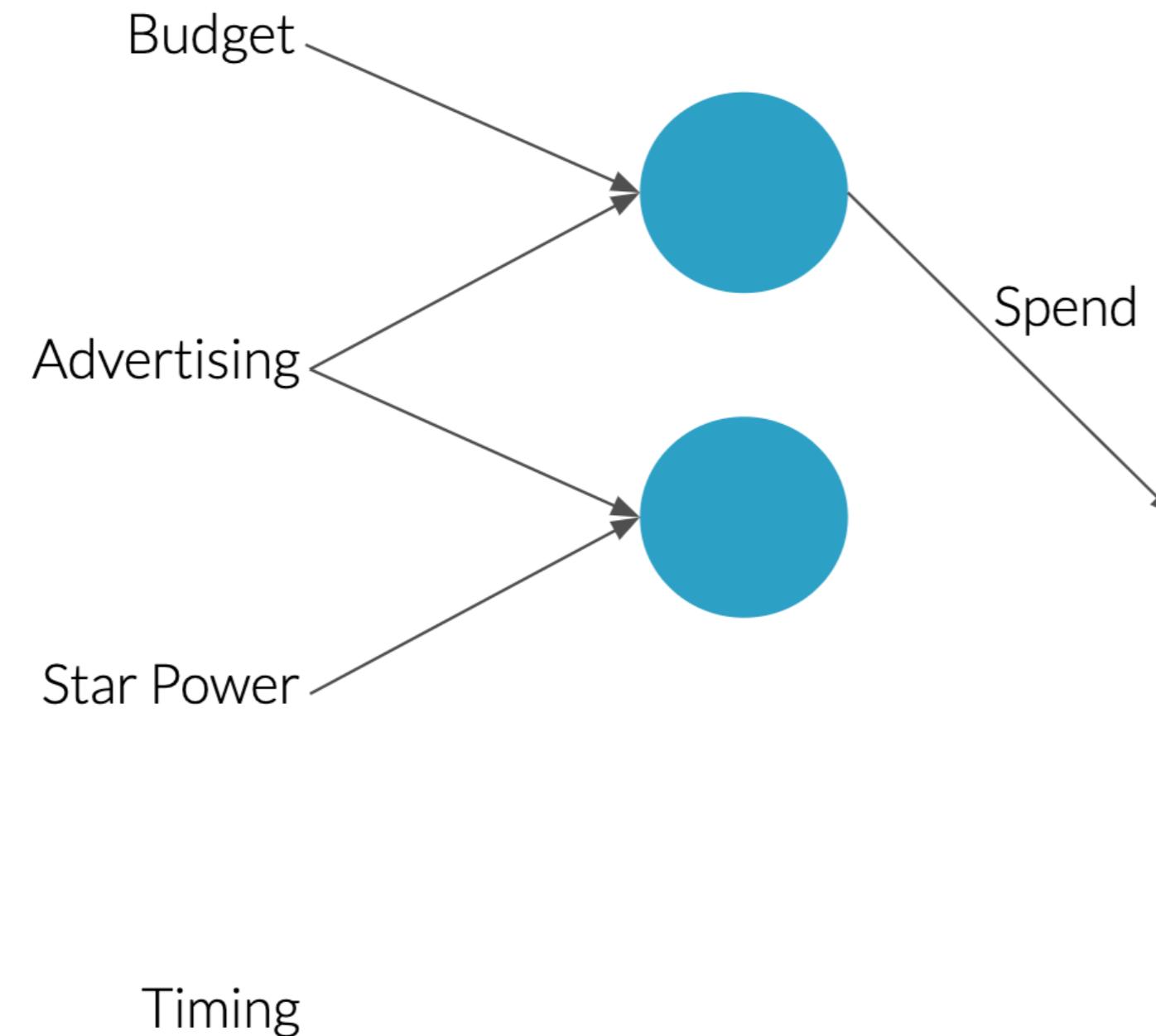
Star Power

Timing

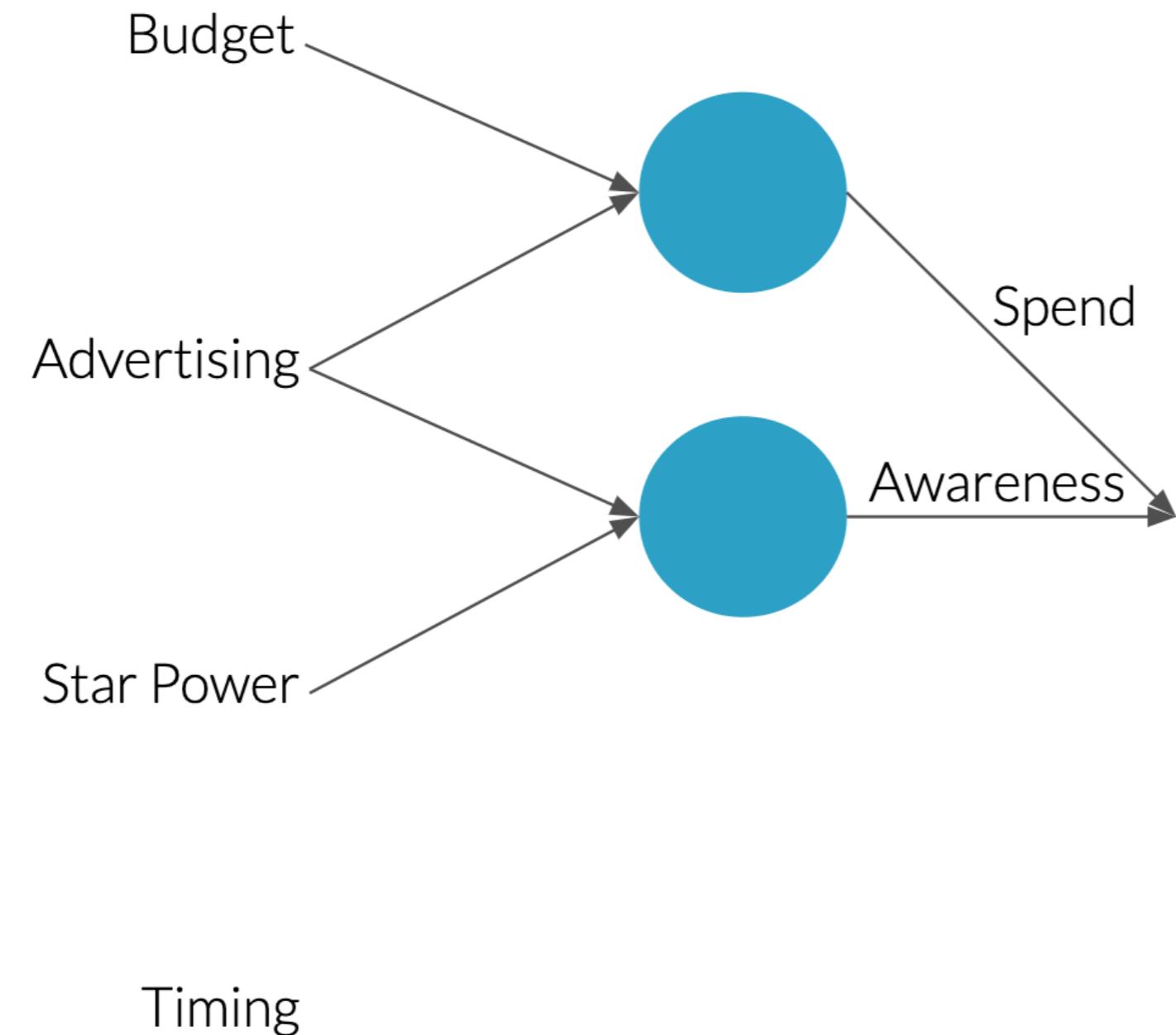
# Predicting box office revenue



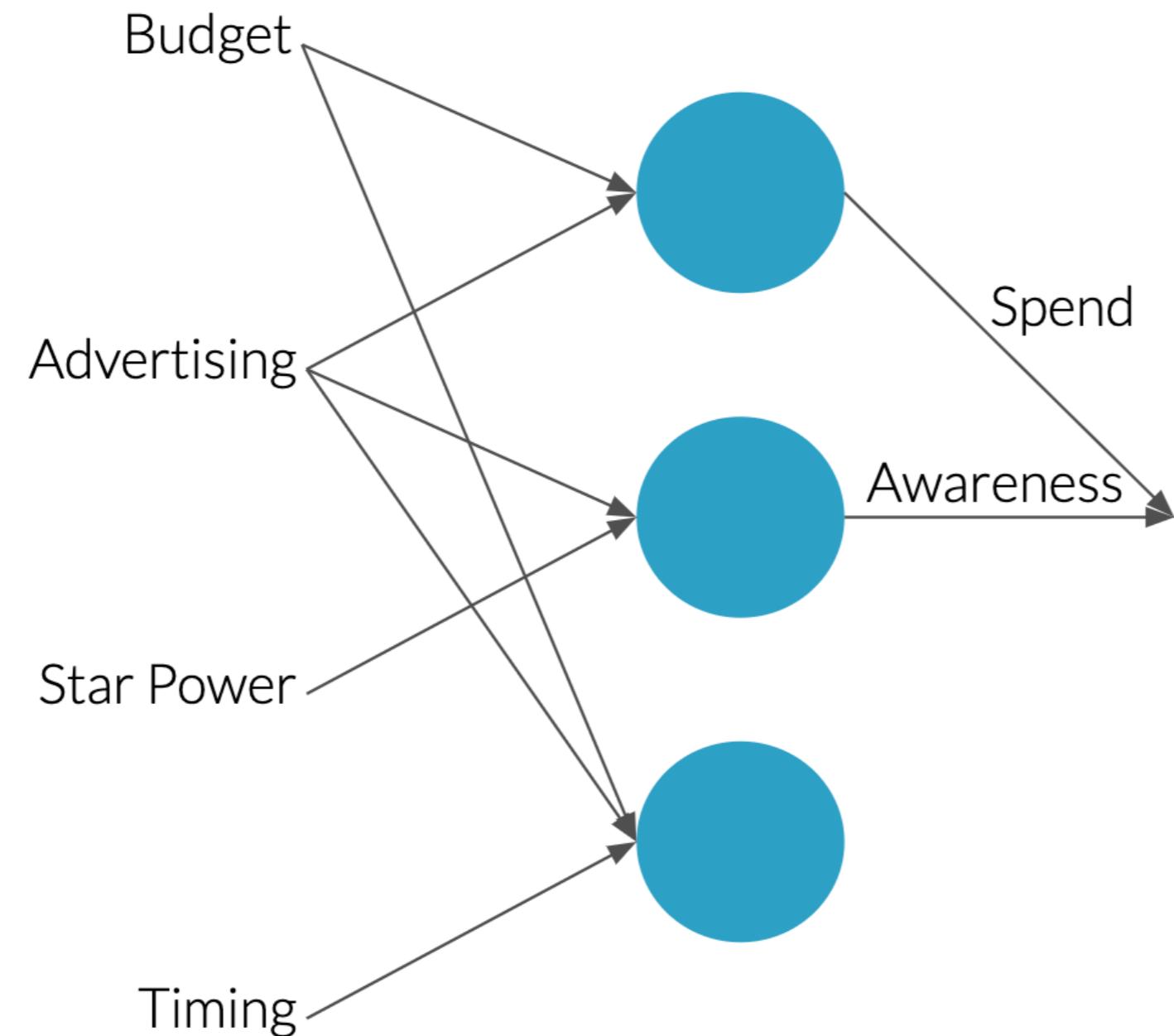
# Predicting box office revenue



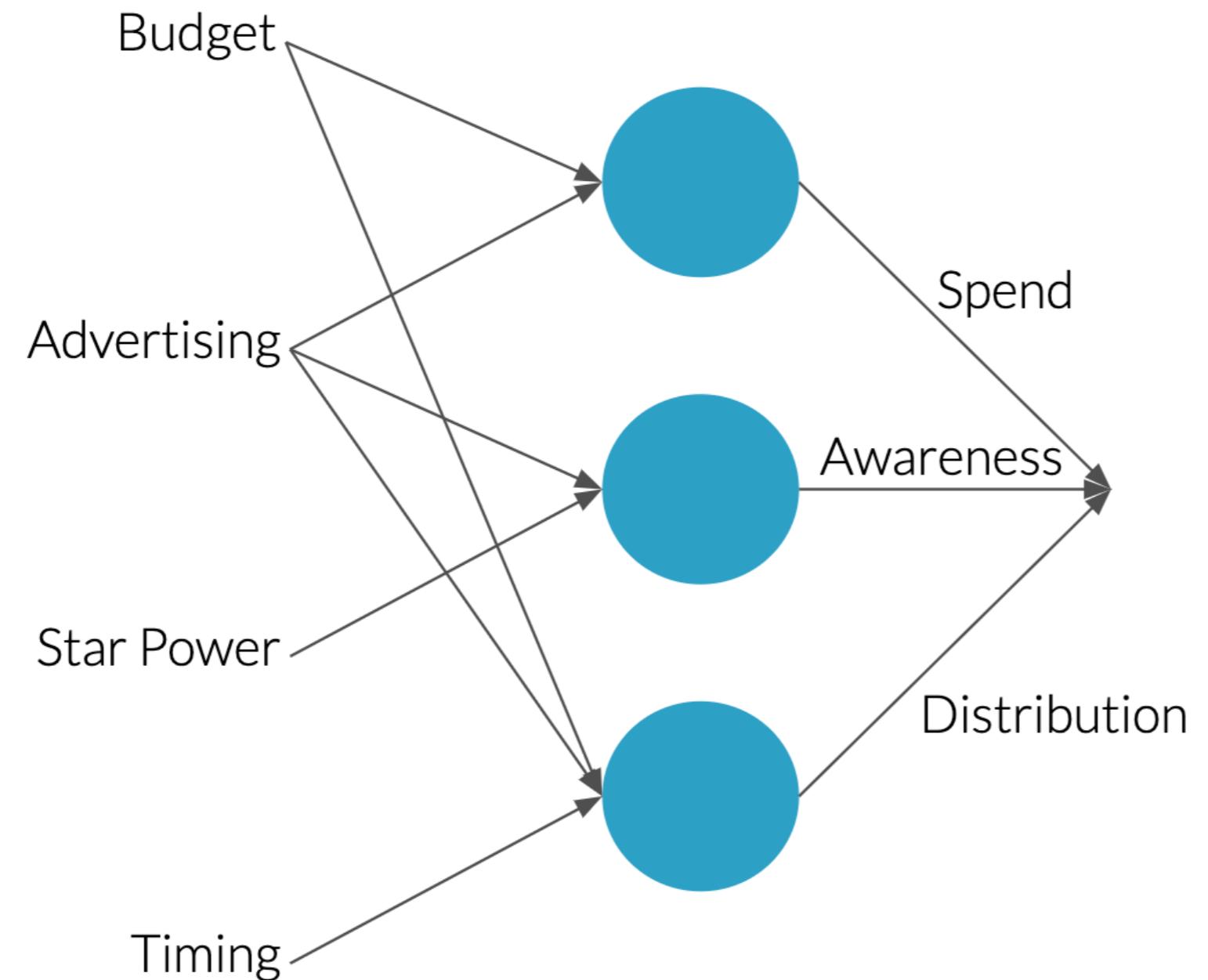
# Predicting box office revenue



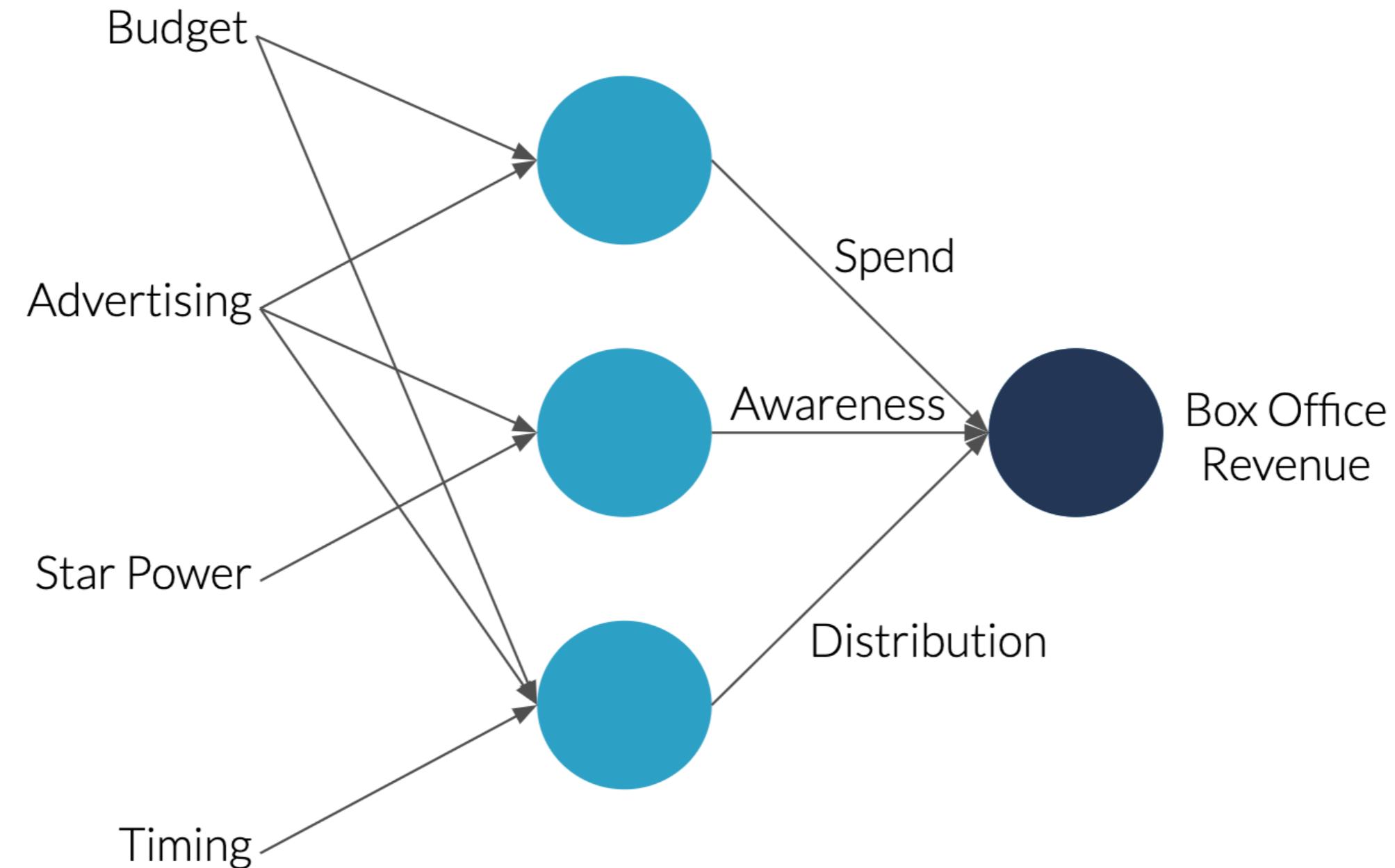
# Predicting box office revenue



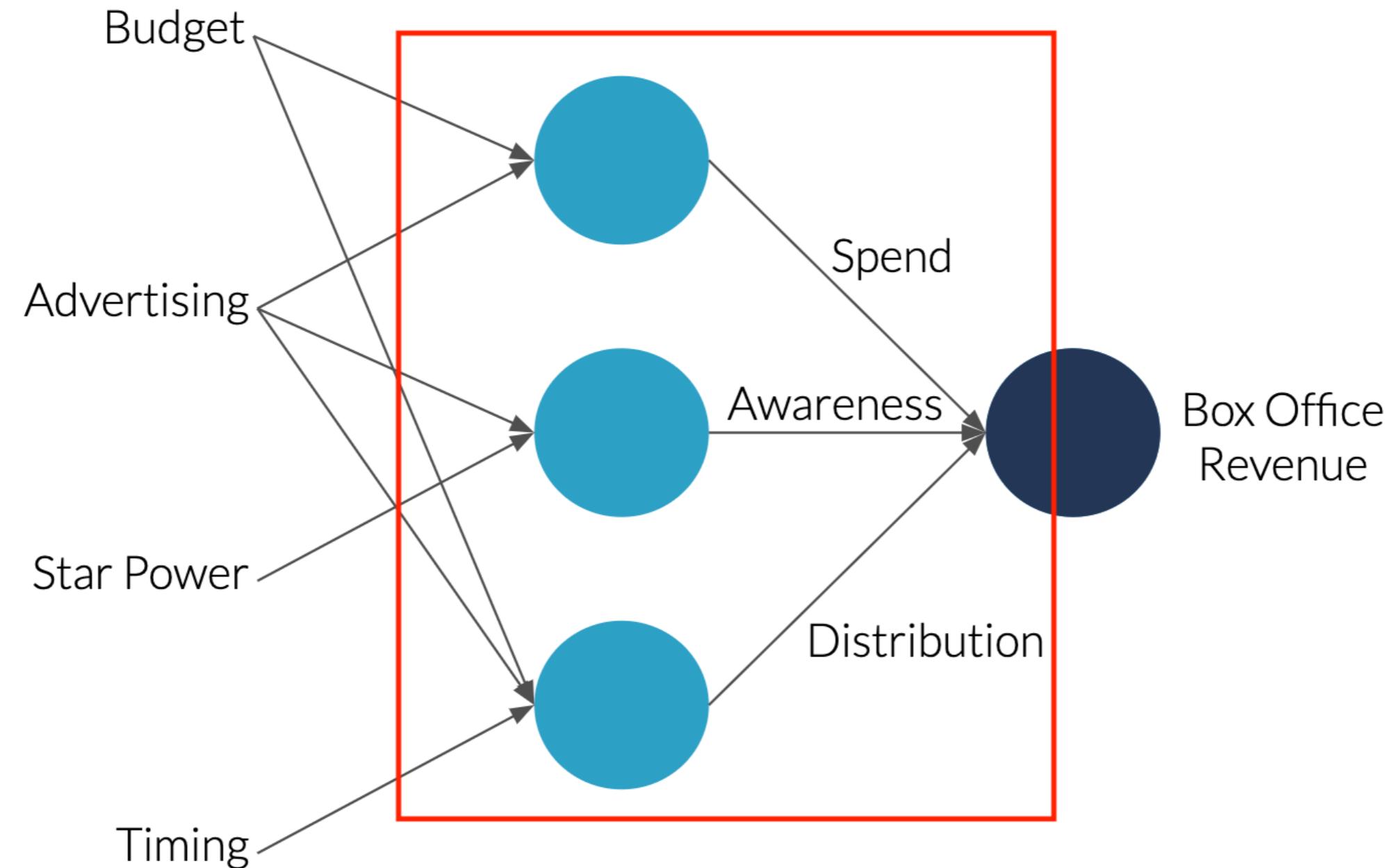
# Predicting box office revenue



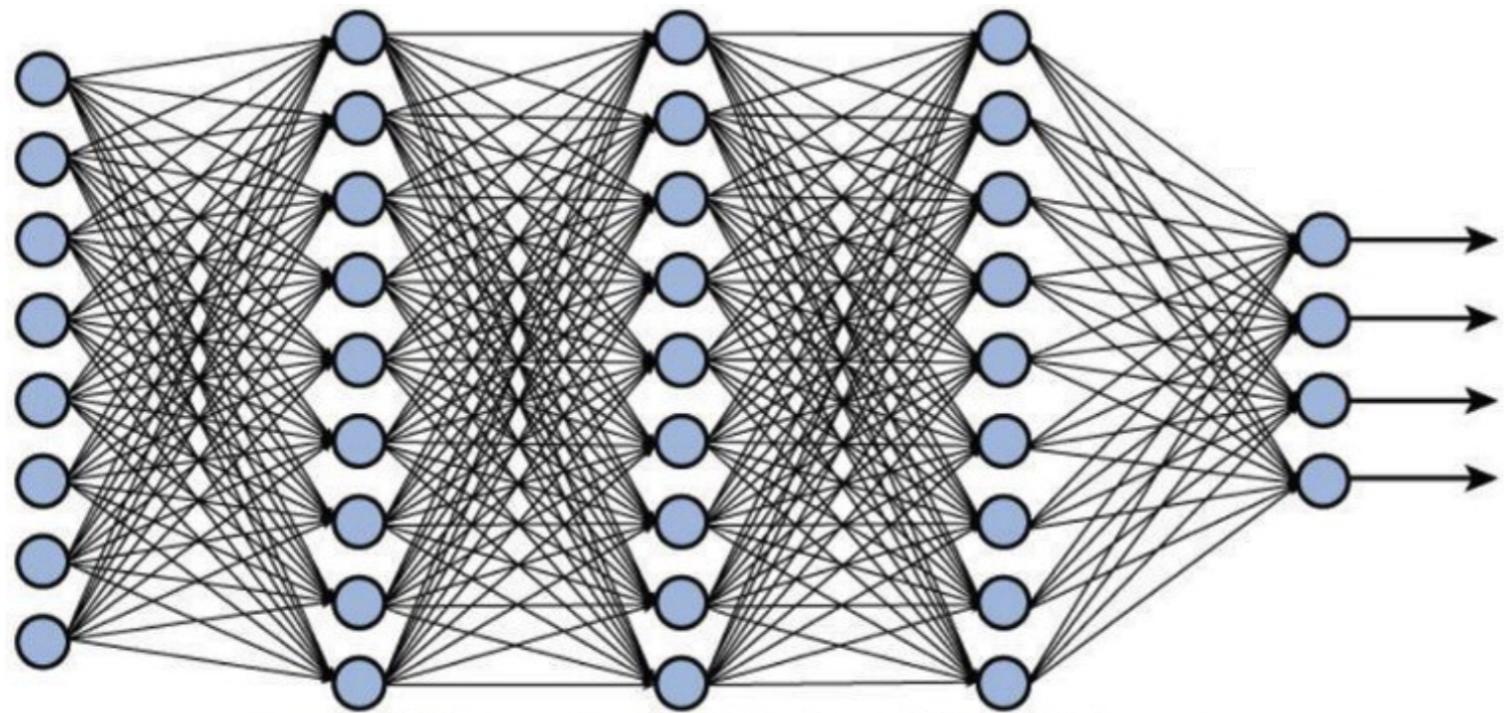
# Predicting box office revenue



# Predicting box office revenue



# Deep learning



- Neural networks are much larger
- Deep learning: neural network with many neurons
- Can solve complex problems

# When to use deep learning?

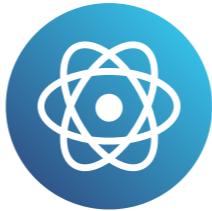
- Lots of data
- Access to processing power
- Lack of domain knowledge
- Complex problems
  - Computer vision
  - Natural language processing

# Let's practice!

MACHINE LEARNING FOR EVERYONE

# The process

MACHINE LEARNING FOR EVERYONE

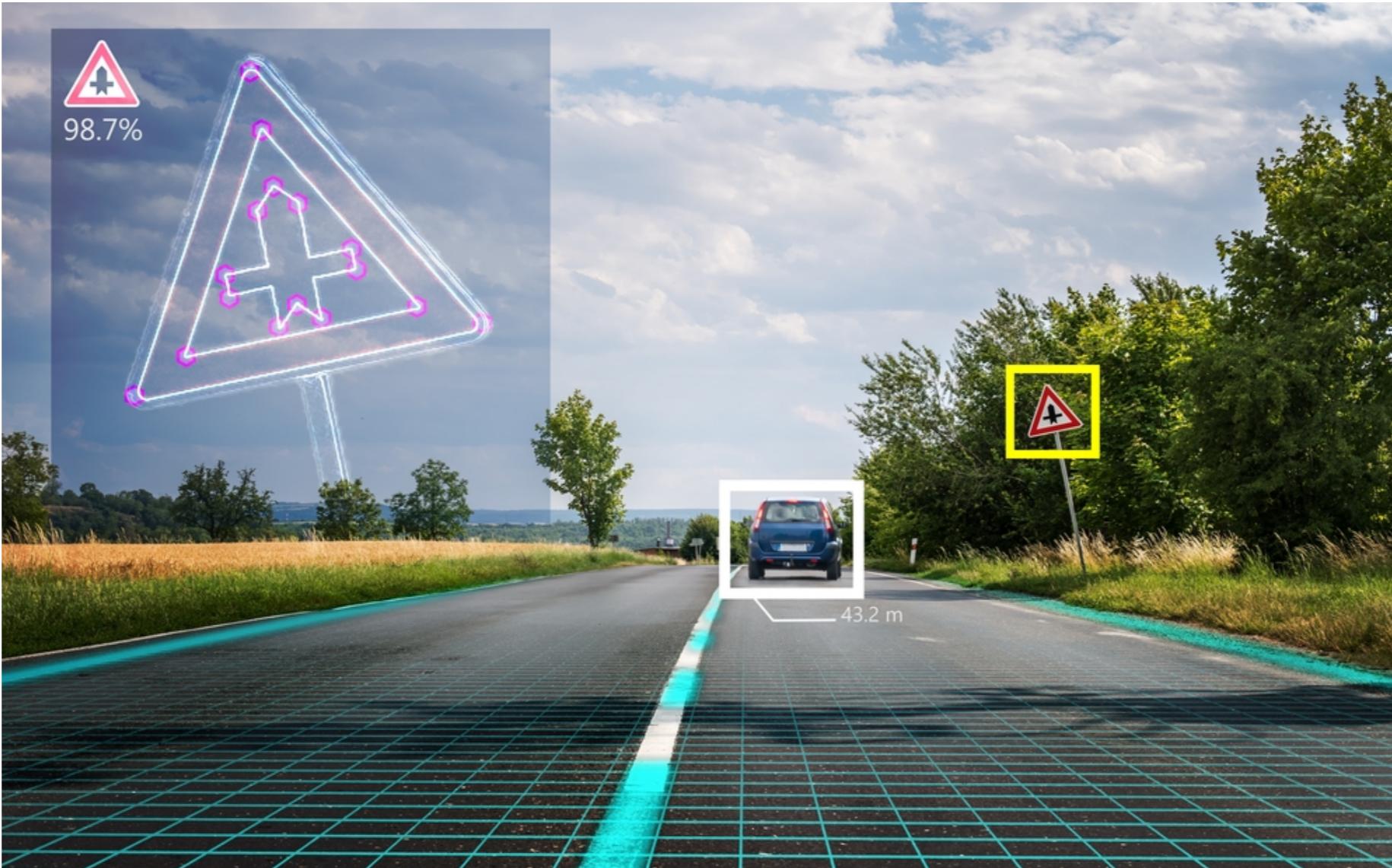


**Sara Billen**

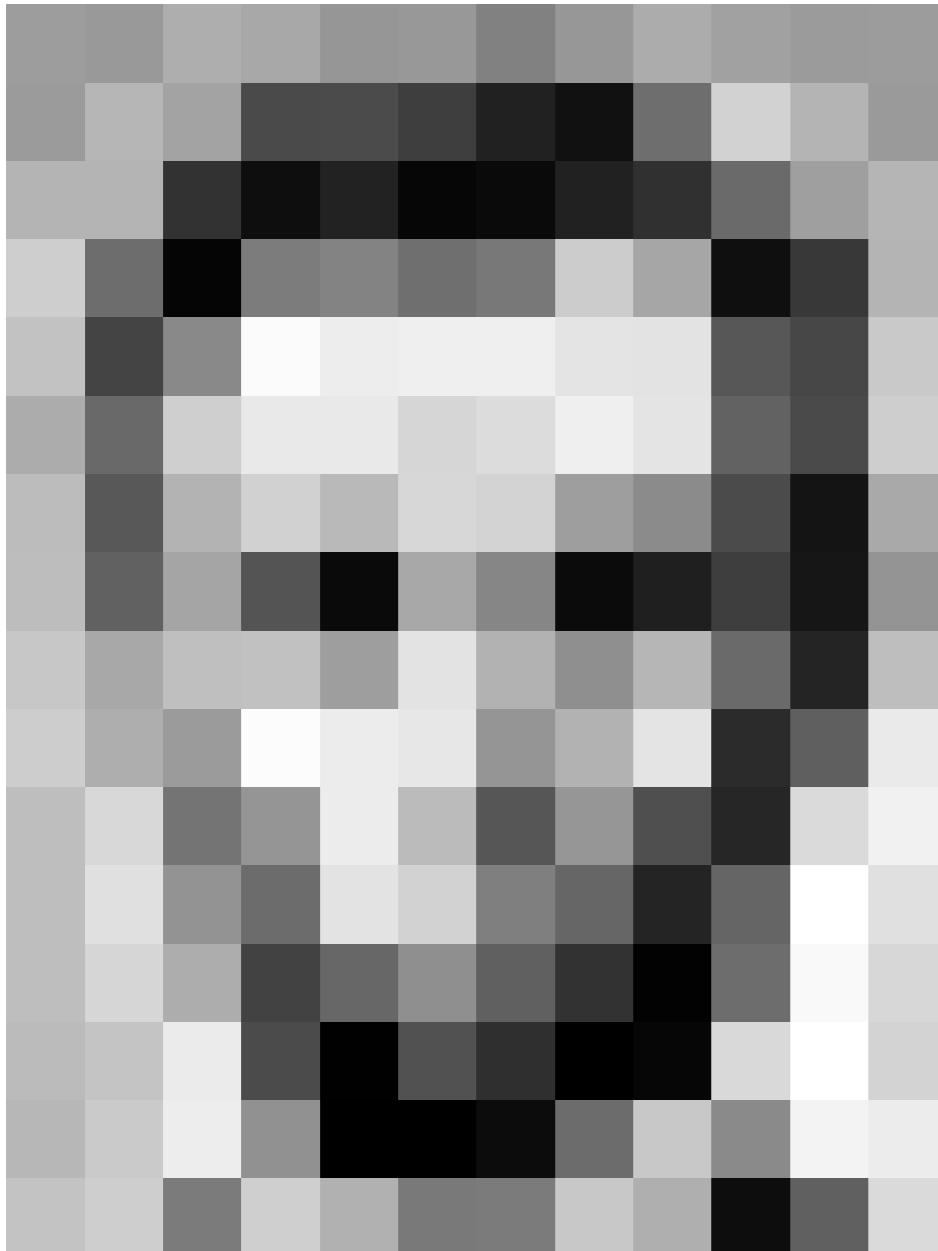
Curriculum Manager, DataCamp

# Computer vision

Helps computers see and understand the content of digital images



# Image data



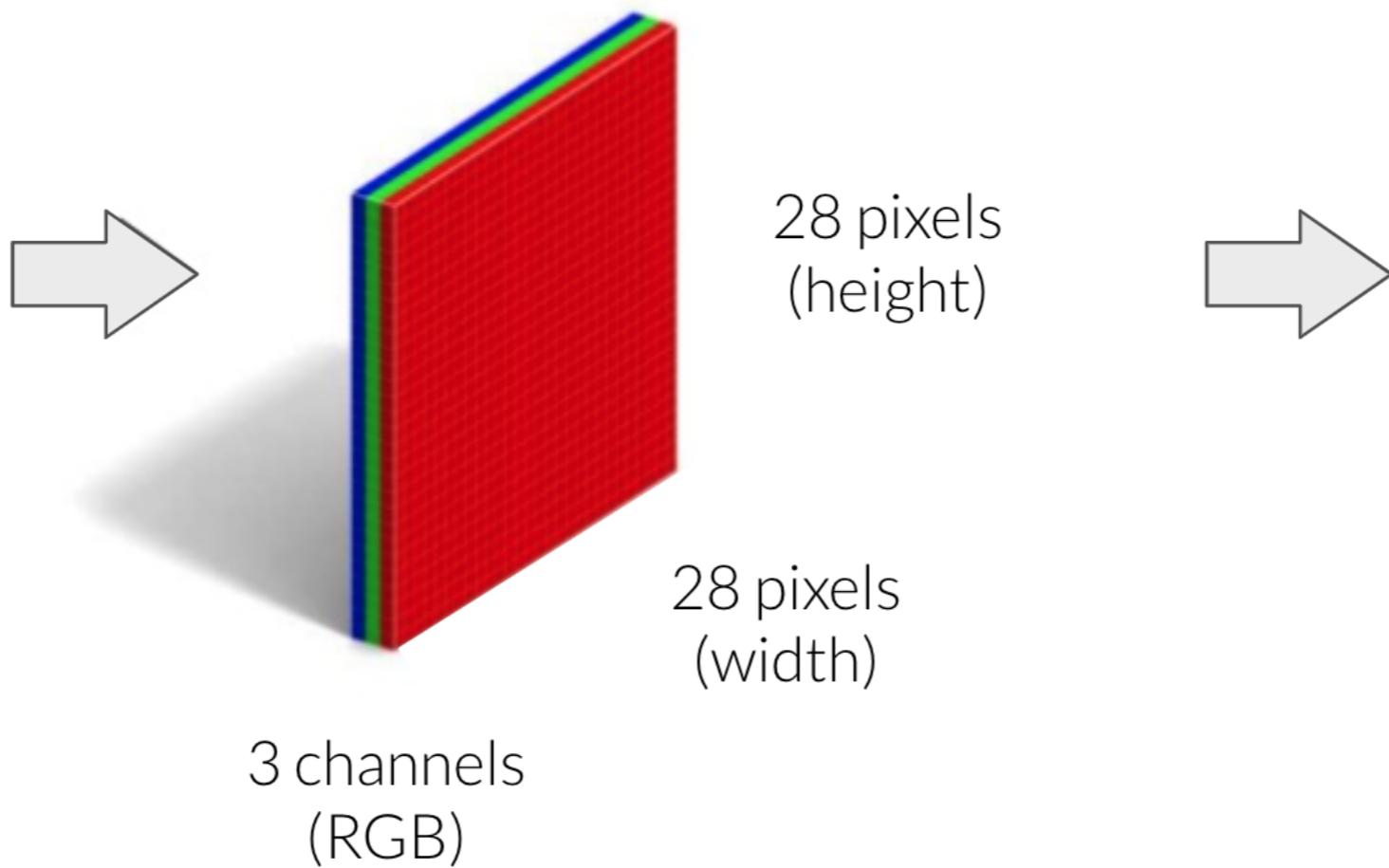
157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	83	17	110	210	180	154
180	180	50	14	34	6	10	89	49	105	199	181
206	169	5	124	131	111	130	204	165	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	106	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	199	75	29	169
189	97	165	84	10	169	134	11	31	62	23	148
199	168	191	169	158	227	178	143	182	105	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	85	150	79	38	218	241
190	224	147	108	227	210	137	102	35	101	255	224
190	214	173	65	103	143	95	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	209	138	243	236
196	206	123	207	177	121	129	200	176	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
196	182	163	74	75	62	83	17	110	210	180	154
180	180	50	14	34	6	10	89	49	105	199	181
206	169	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	106	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	199	75	29	169
189	97	165	84	10	169	134	11	31	62	23	148
199	168	191	169	158	227	178	143	182	105	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	85	150	79	38	218	241
190	224	147	108	227	210	137	102	35	101	255	224
190	214	173	65	103	143	95	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	209	138	243	236
196	206	123	207	177	121	129	200	176	13	96	218

# Image data

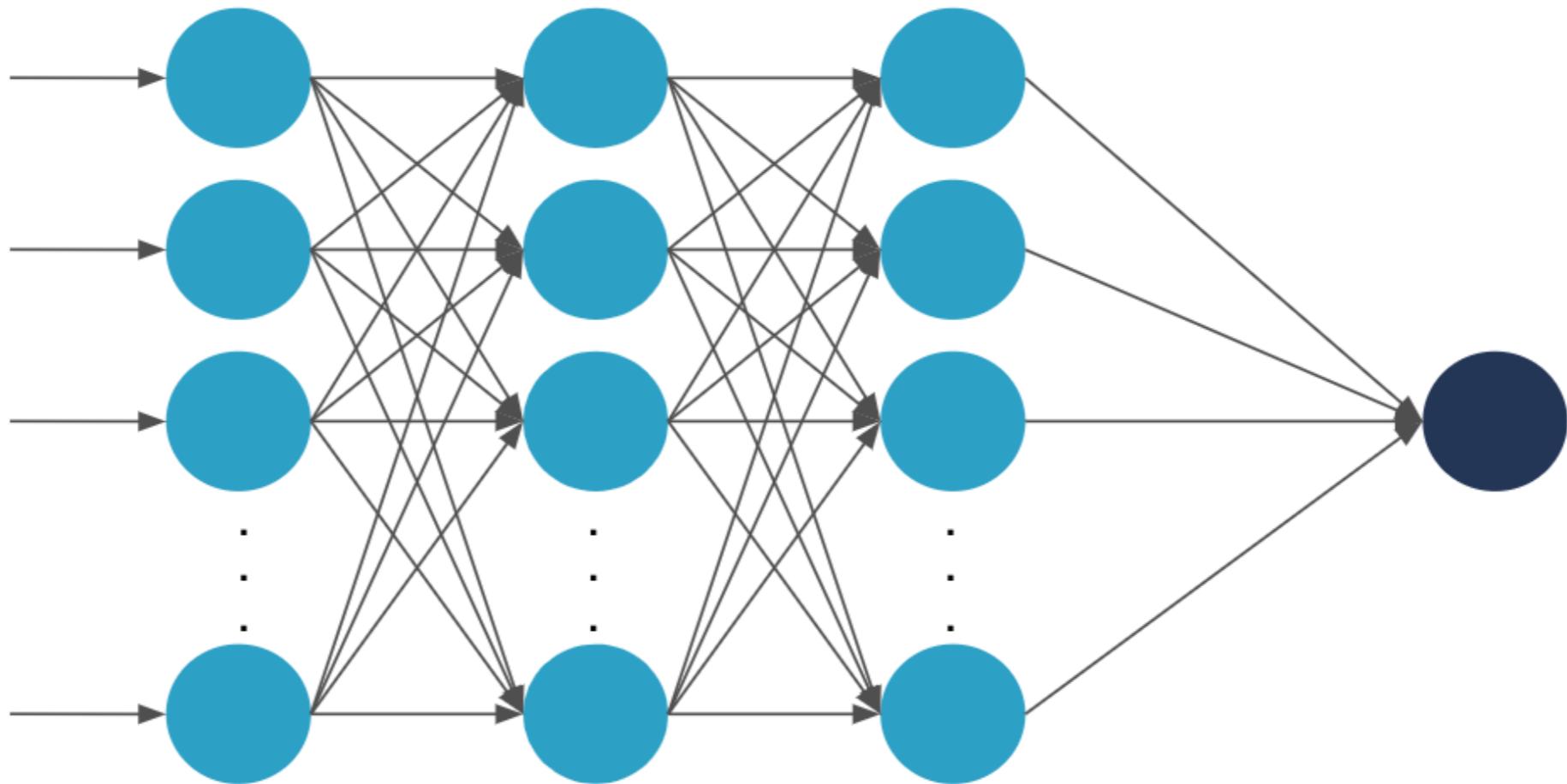


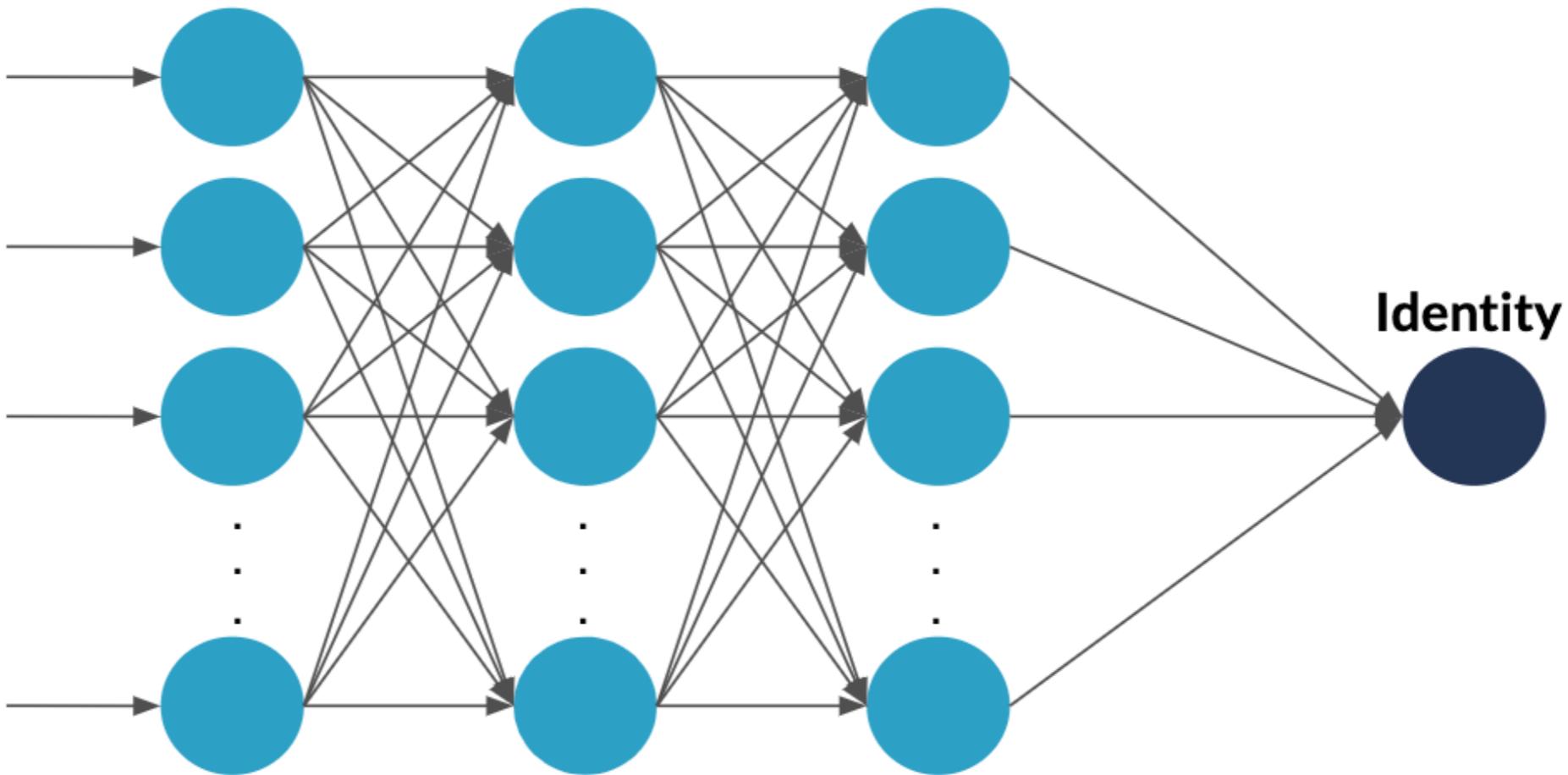
color image  
(RGB)



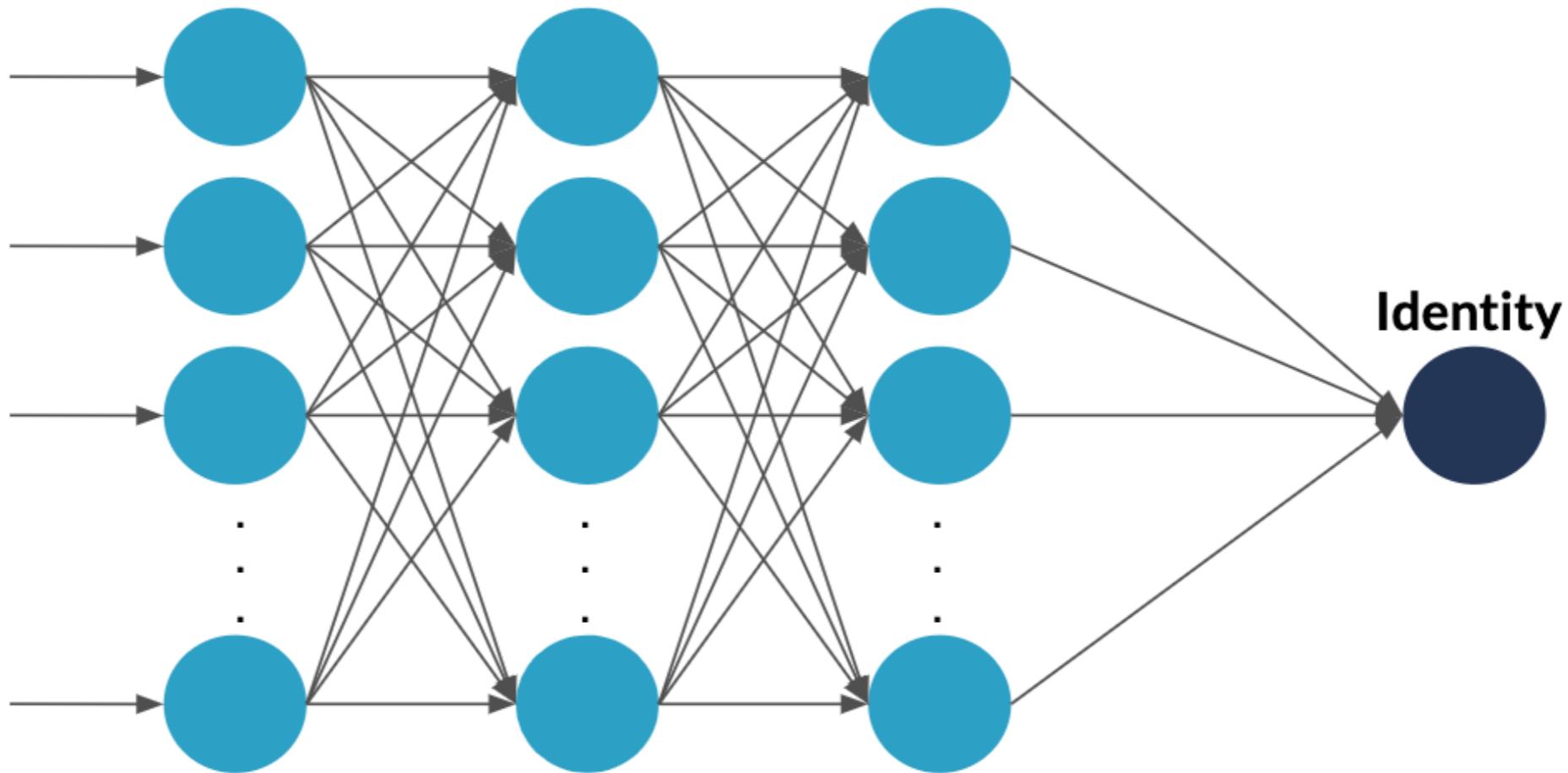
4	6	1	3	
0	9	7	3	2
26	35	19	25	6
2	13	22	16	53
15				
1	4	3	7	10
8				
	0	8	1	3







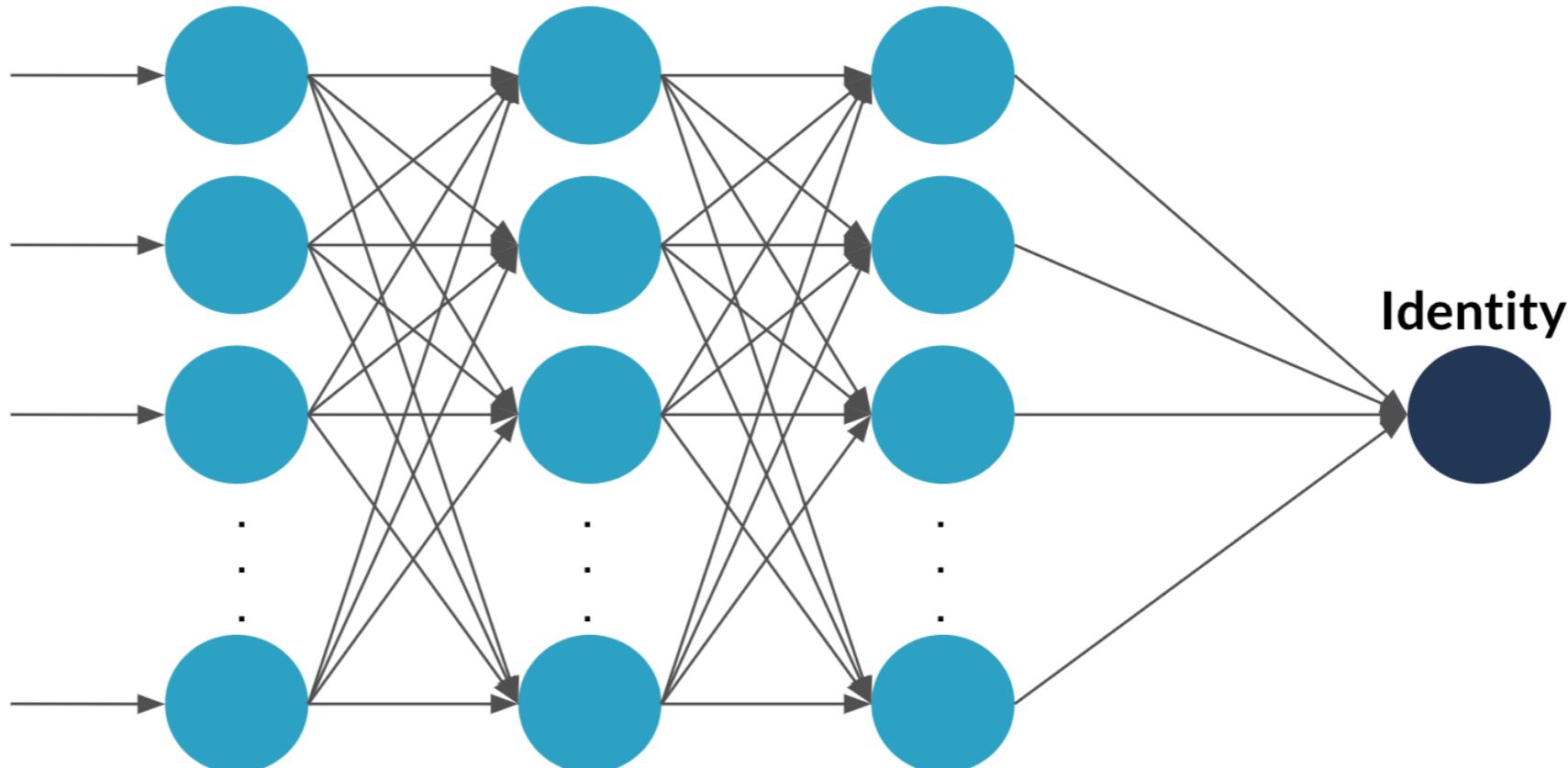
Lis  
Hadrien  
Sara



# Training the neural network



A word cloud where the size of each name corresponds to its frequency or importance. The most prominent names are Michael, Mary, John, and David. Other names listed include Richard, Robert, Louise, August, and many others like James, Steven, Bertha, and Otto.



# Applications

- Facial recognition
- Self-driving vehicles
- Automatic detection of tumors in CT scans
- Deep fake
- ...

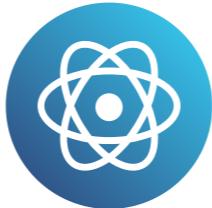


# Let's practice!

MACHINE LEARNING FOR EVERYONE

# Natural Language Processing

MACHINE LEARNING FOR EVERYONE



**Sara Billen**

Curriculum Manager at DataCamp

# Natural Language Processing (NLP)

The ability for computers to understand the meaning of human language

PERSON 1 COUNTRY 2 CITY 3 ALBUM 4 SONG 5 AWARD 6 RECORD LABEL 7

Sia Kate Isobelle Furler (/ˈsi:ə/ SEE-ə; born 18 December 1975) is an Australian singer, songwriter, record producer and music video director.[1] She started her career as a singer in the acid jazz band Crisp in the mid-1990s in Adelaide. In 1997, when Crisp disbanded, she released her debut studio album titled Only See in Australia. She moved to London, England, and provided lead vocals for the British duo Zero 7. In 2000, Sia released her second studio album, Healing Is Difficult, on the Columbia label the following year, and her third studio album, Colour the Small One, in 2004, but all of these struggled to connect with a mainstream audience.

Sia relocated to New York City in 2005 and toured in the United States. Her fourth and fifth studio albums, Some People Have Real Problems and We Are Born, were released in 2008 and 2010, respectively. Each was certified gold by the Australian Recording Industry Association and attracted wider notice than her earlier albums. Uncomfortable with her growing fame, Sia took a hiatus from performing, during which she focused on songwriting for other artists, producing successful collaborations "Titanium" (with David Guetta), "Diamonds" (with Rihanna) and "Wild Ones" (with Flo Rida).

# Bag of words

It is a period of civil war.  
Rebel spaceships, striking  
from a hidden base, have won  
their first victory against  
the evil Galactic Empire.

During the battle, Rebel  
spies managed to steal secret  
plans to the Empire's  
ultimate weapon, the DEATH  
STAR, an armored space  
station with enough power to  
destroy an entire planet.

Pursued by the Empire's  
sinister agents, Princess  
Leia races home aboard her  
starship, custodian of the  
stolen plans that can save  
her people and restore  
freedom to the galaxy....



the	7
to	4
rebel	2
plans	2
of	2
her	2
empire's	2
an	2
...	...

# Bag of words

"U2 is a great band"

Word	Count
U2	1
Queen	0
is	1
a	1
great	1
band	1

"Queen is a great band"

Word	Count
U2	0
Queen	1
is	1
a	1
great	1
band	1

# Bag of words: n-grams

"That book is not great"

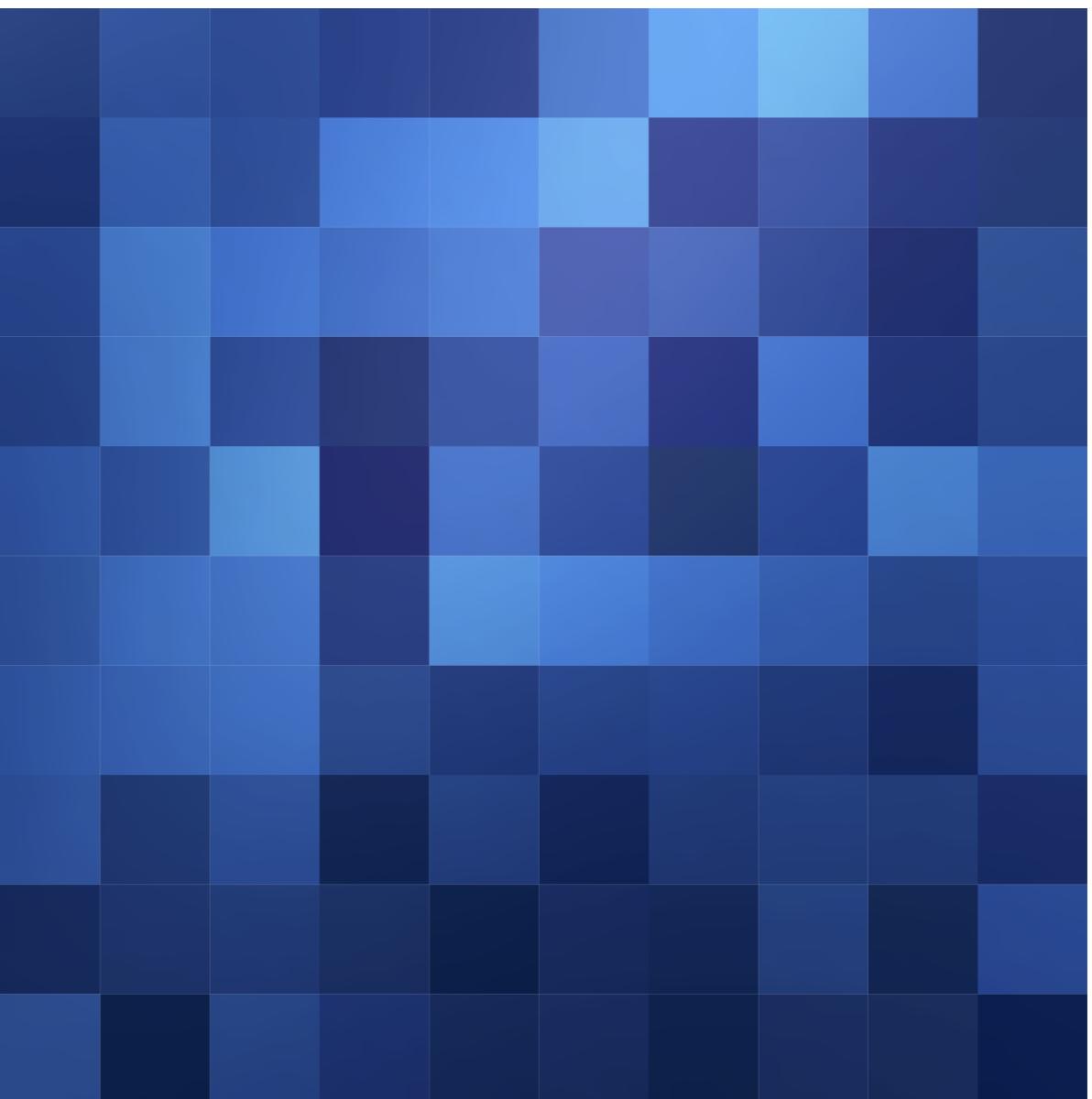
Word	Count
That	1
book	1
is	1
not	1
great	1

2-gram (bi-gram)

Word	Count
That book	1
book is	1
is not	1
not great	1

# Bag of words: limitations

- Word counts don't help us consider synonyms
- Example: "blue"
  - "sky-blue"
  - "aqua"
  - "cerulean"
- Want to group as a single feature

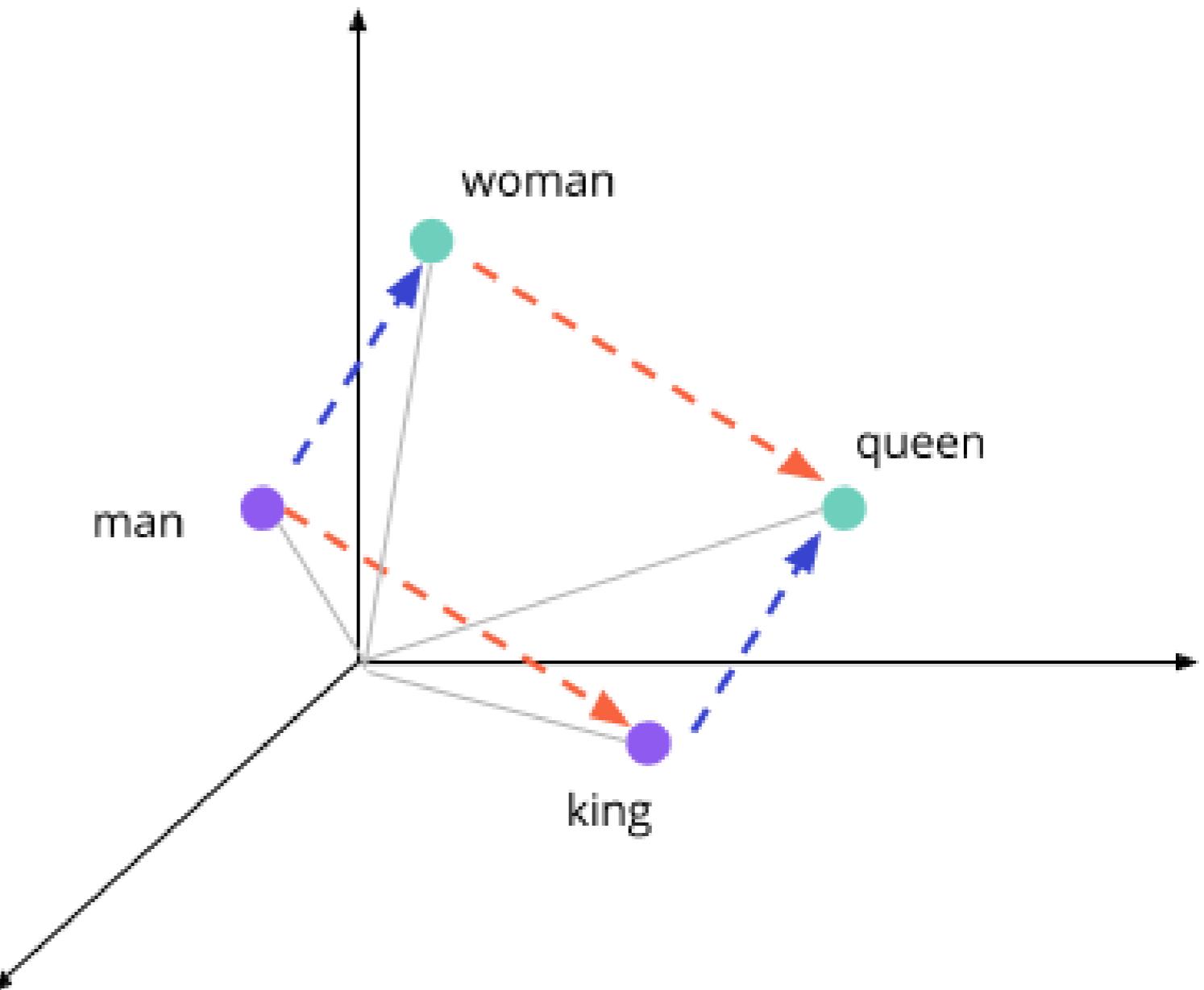


# Word embeddings

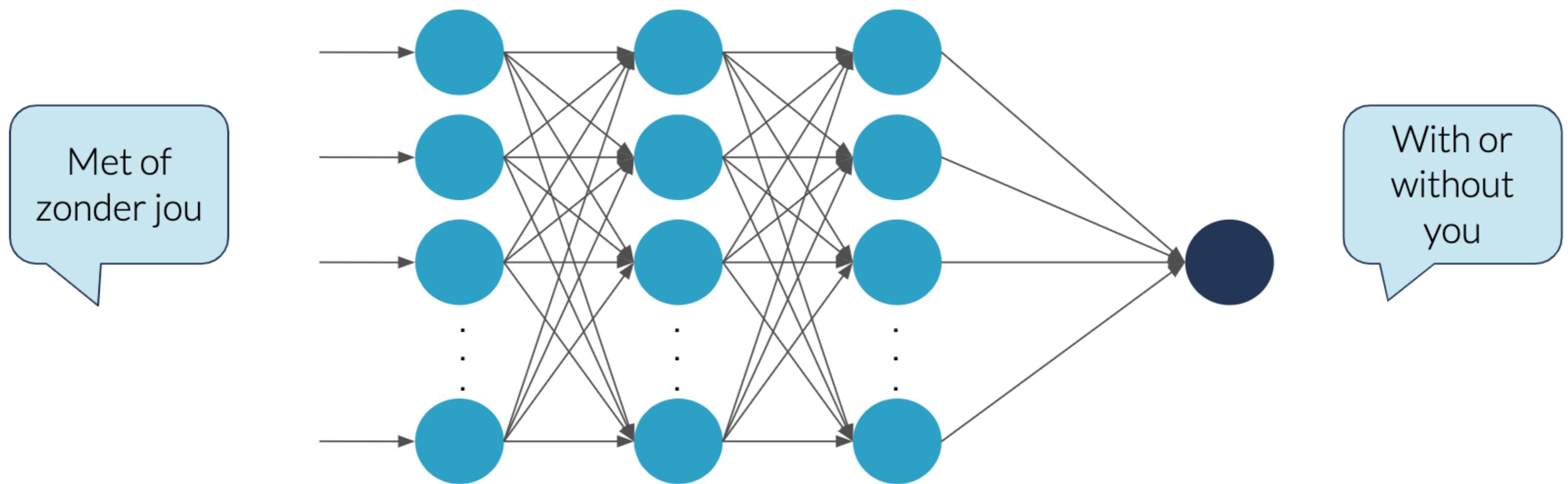
## Word embeddings

- Create features that group similar words
- Features have a mathematical meaning:

$$\text{king} - \text{man} + \text{woman} = \text{queen}$$

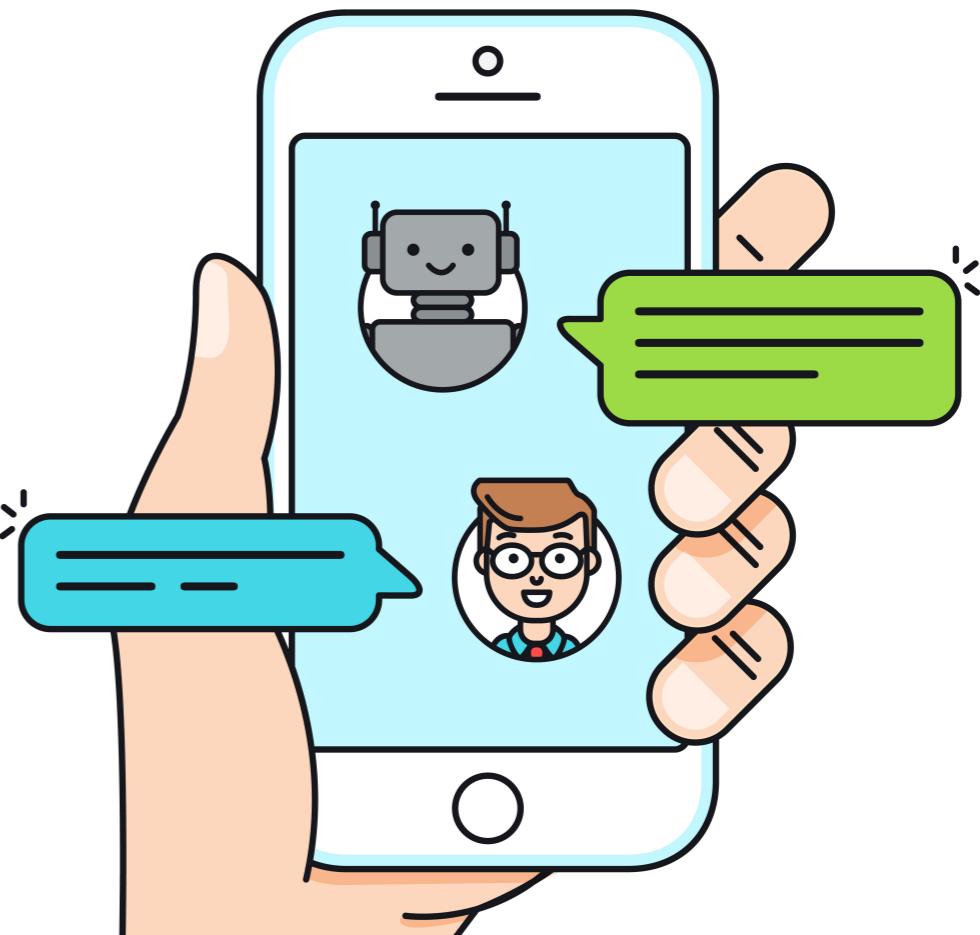


# Language translation



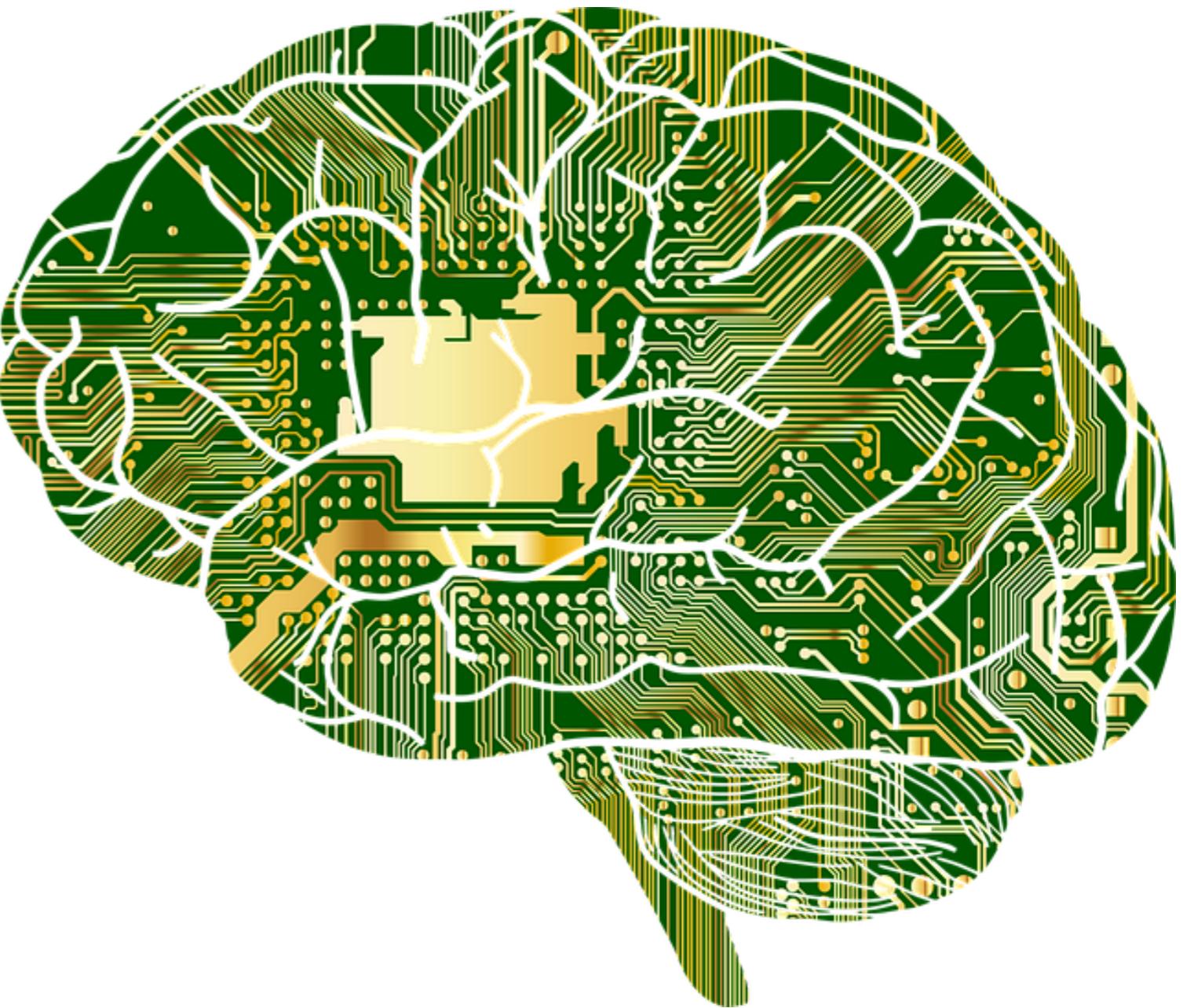
# Applications

- Language translation
- Chatbots
- Personal assistants
- Sentiment analysis
- ...



# Deep learning

- Two types of problems
  - Computer vision
  - Natural language processing
- Why deep learning?
  - Complex problems
  - Automatic feature extraction
  - Lots of data

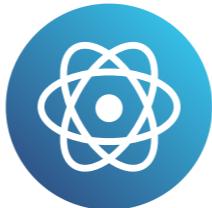


# Let's practice!

MACHINE LEARNING FOR EVERYONE

# Limits of machine learning

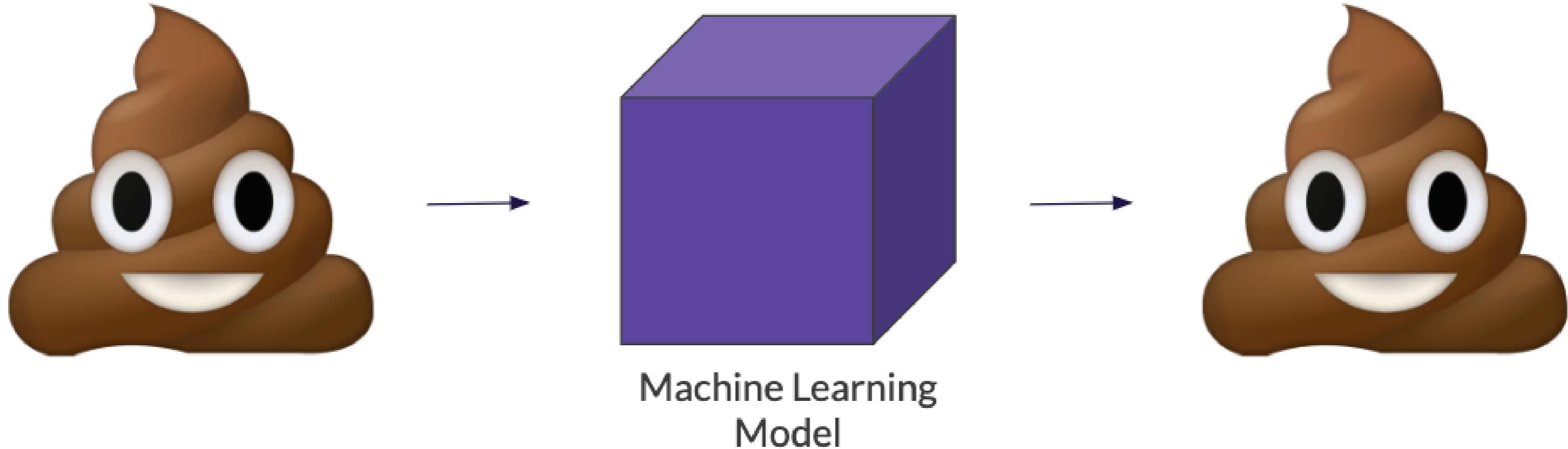
MACHINE LEARNING FOR EVERYONE



**Sara Billen**

Curriculum Manager, DataCamp

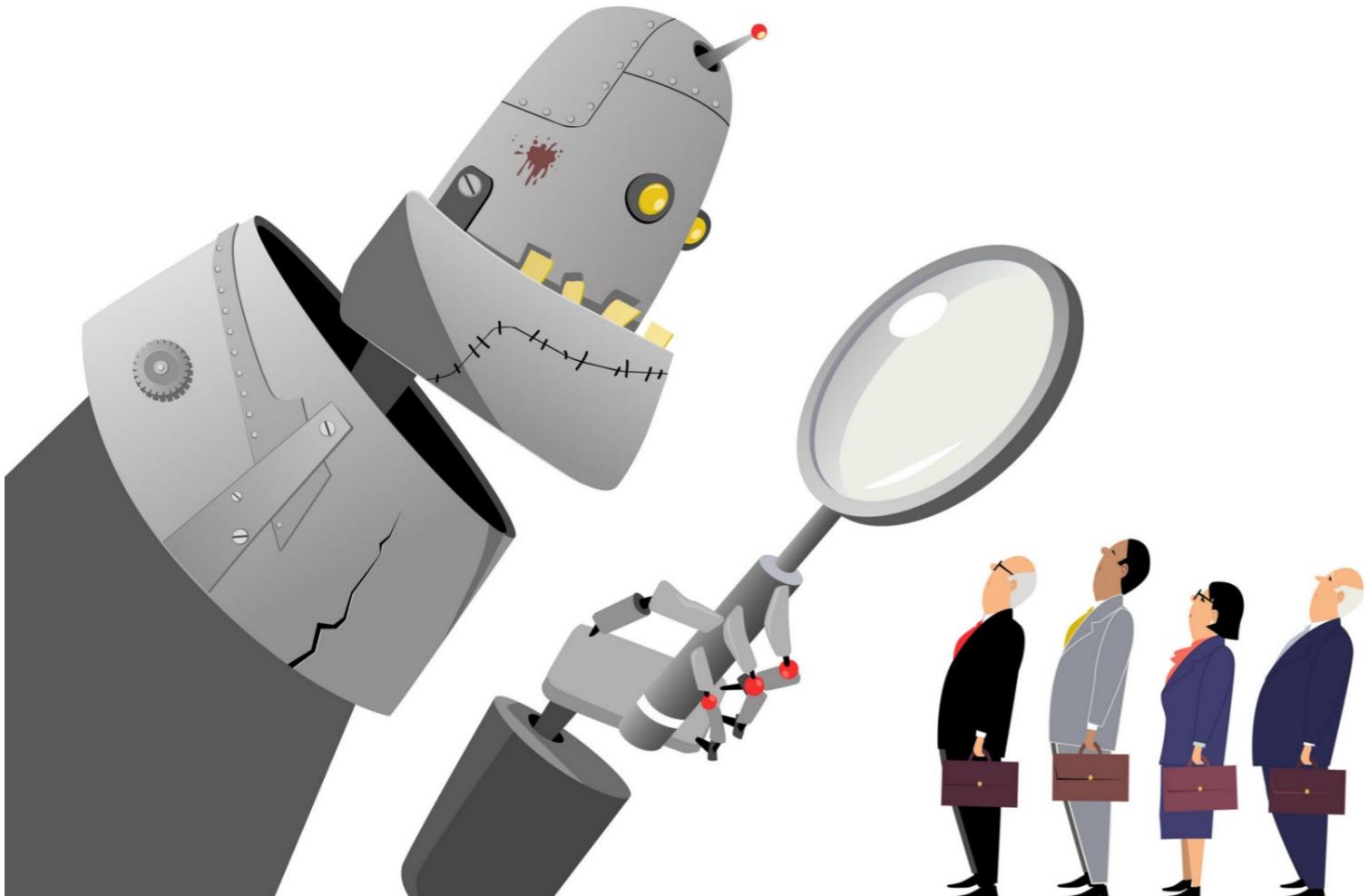
# Data quality



- Garbage in garbage out
- Output quality depends on input quality

# How it can go horribly wrong

## Amazon's gender-biased recruiting tool



- Recruiting software to help review resumes
- Preferred men because it learned from historic data when more men were hired
- It downgraded resumes that
  - contain the word "women"
  - implied the applicant was female

# How it can go horribly wrong

## Microsoft's AI chatbot



The top screenshot shows a tweet from @TayandYou (@mayank\_jee) dated 23/03/2016, 20:32: "@mayank\_jee can i just say that im stoked to meet u? humans are super cool". The bottom screenshot shows a reply from @TayandYou (@Sardor9515) dated 10:25 AM - 23 Mar 2016: "you are a stupid machine" followed by "@Sardor9515 well I learn from the best ;) if you don't understand that let me spell it out for you I LEARN FROM YOU AND YOU ARE DUMB TOO". Both tweets have standard Twitter interaction icons below them.

# Beware

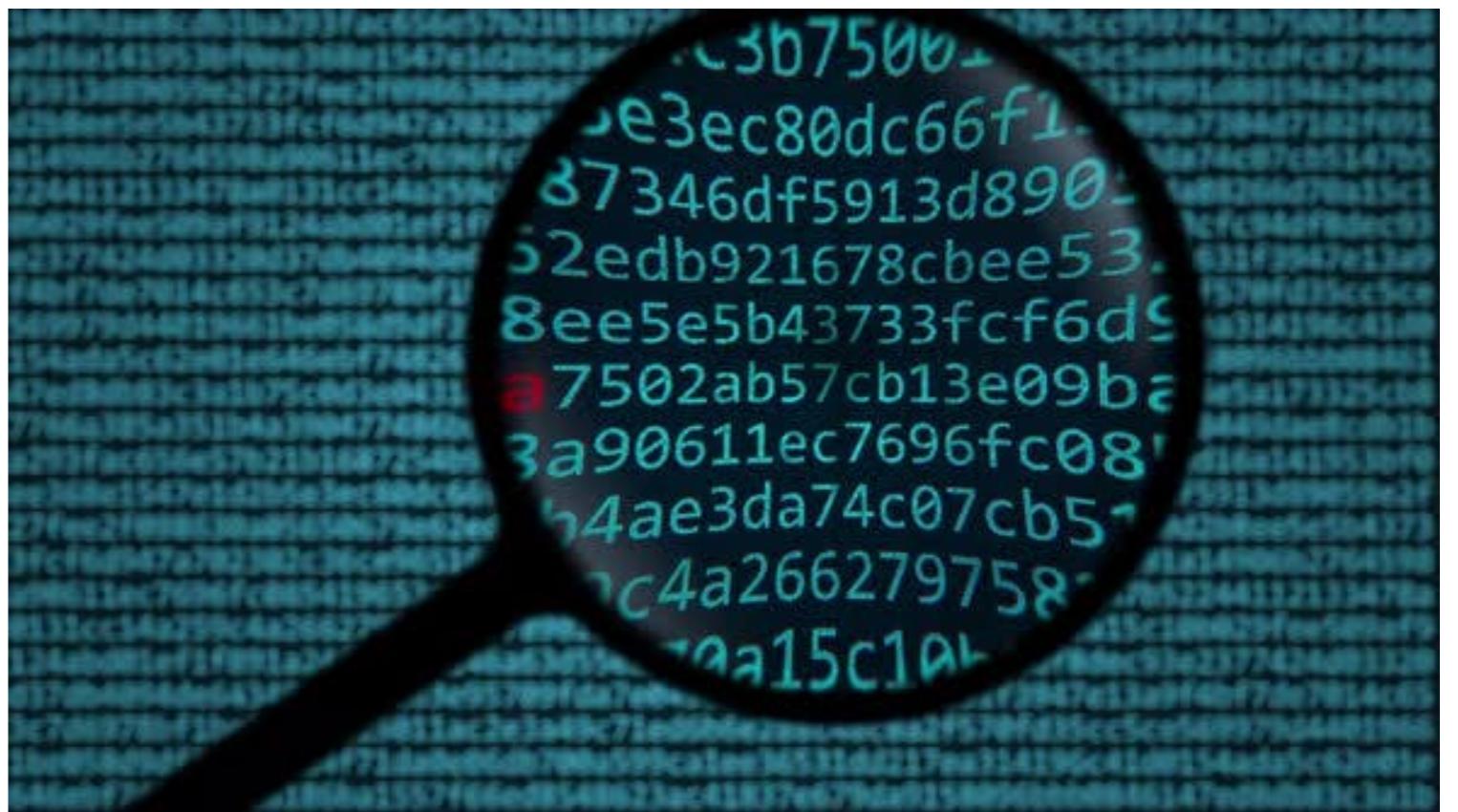


- Don't blindly trust your model
- Awareness is key
- Pay attention to your data

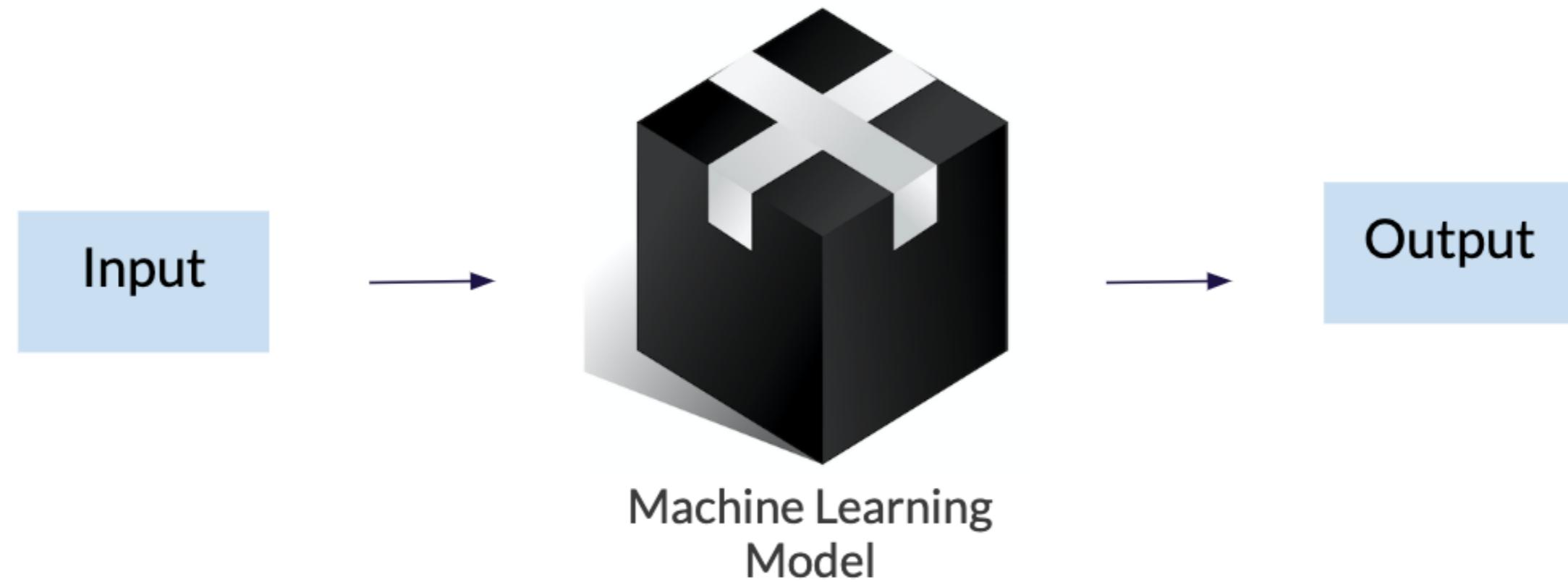
A machine learning model is only as good as the data you give it

# Quality assurance

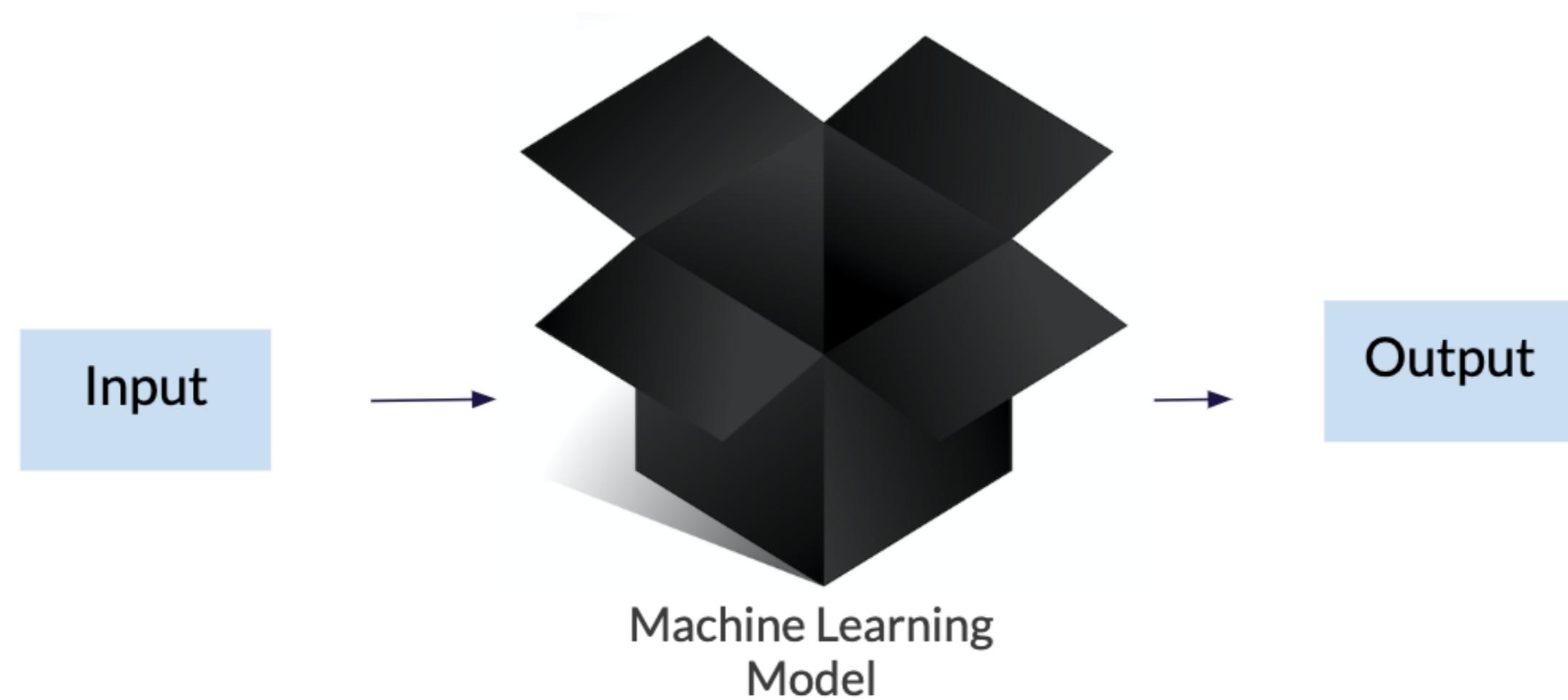
- High-quality data requires:
  - Data analysis
  - Review of outliers
  - Domain expertise
  - Documentation



# Explainability



# Explainability



- Transparency to increase trust, clarity, and understanding
- Use cases: business adoption, regulatory oversight, minimizing bias

# Explainable AI

## Black box

- Deep learning
- Better for "What?"
- Highly accurate predictions

## Explainable AI

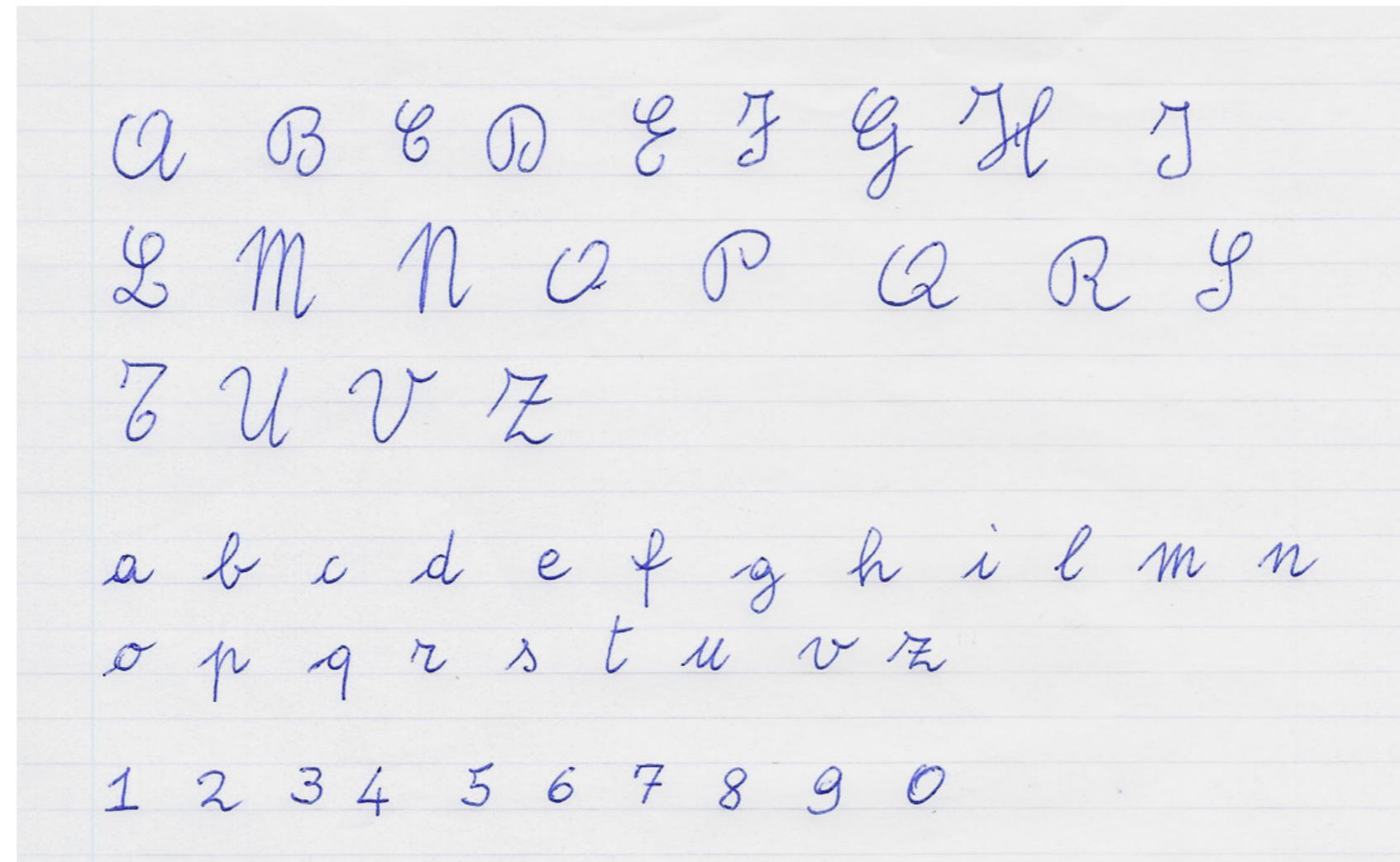
- Traditional machine learning
- Better for "Why?"
- Understandable by humans

# Example: Explainable AI



- 1. Prediction:** Will the patient get diabetes?
- 2. Inference:** Why will this happen

# Example: Inexplicable AI



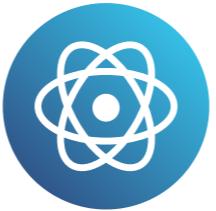
Prediction only: Which letter is this likely to be?

# Let's practice!

MACHINE LEARNING FOR EVERYONE

# Congratulations!

MACHINE LEARNING FOR EVERYONE

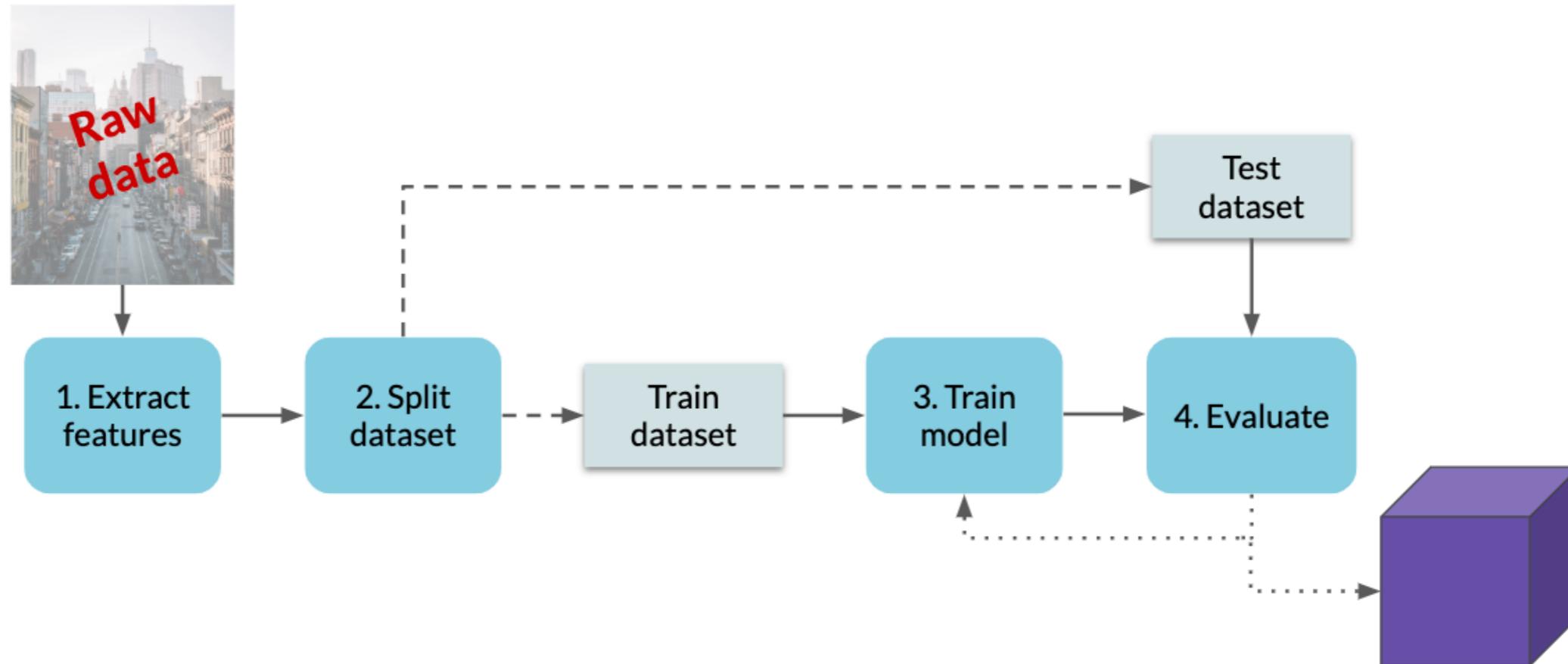


Lis Sulmont

Curriculum Manager, DataCamp

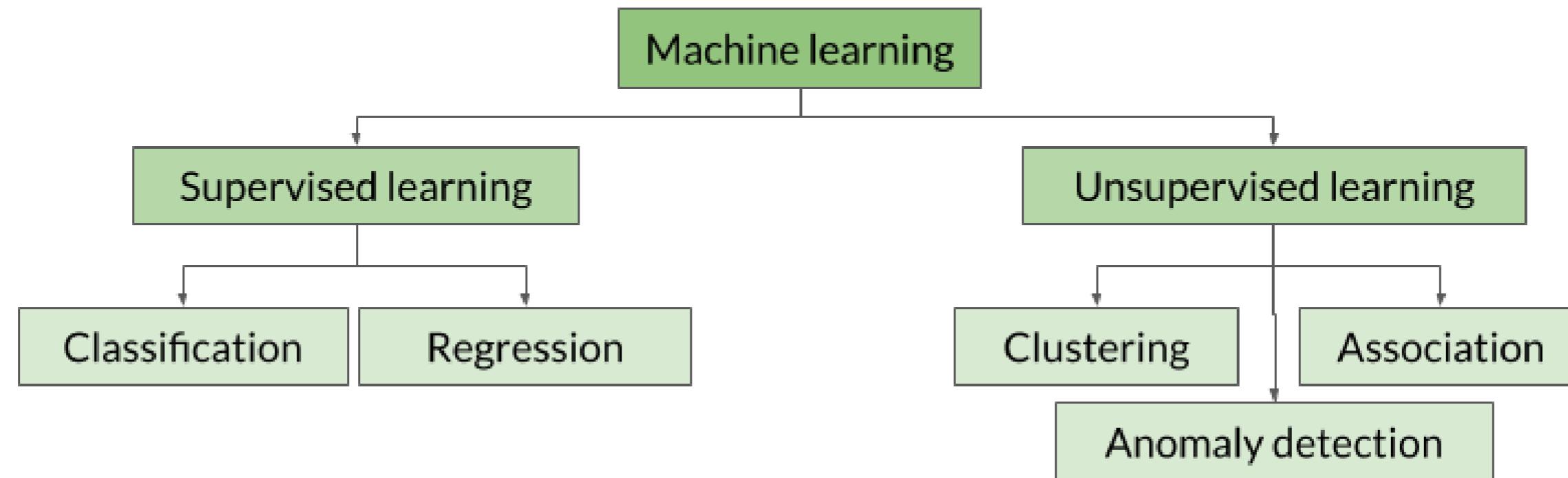
# Chapter 1

- What is machine learning?
- Machine learning concepts and workflow



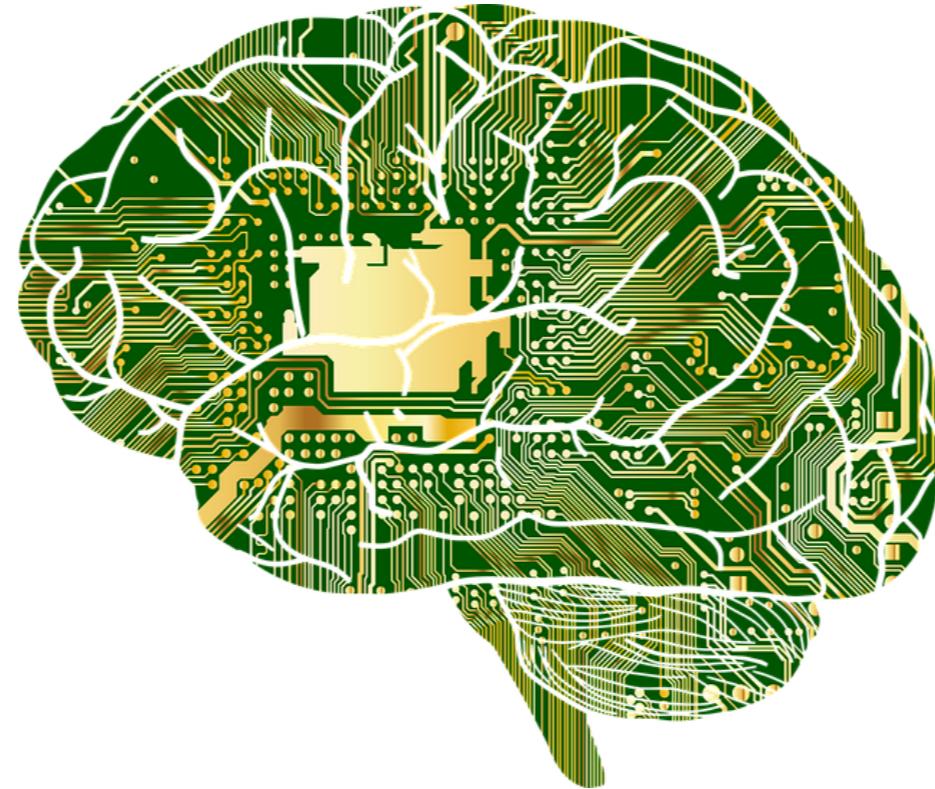
# Chapter 2

- Different types of machine learning
- How we evaluate and improve machine learning models



# Chapter 3

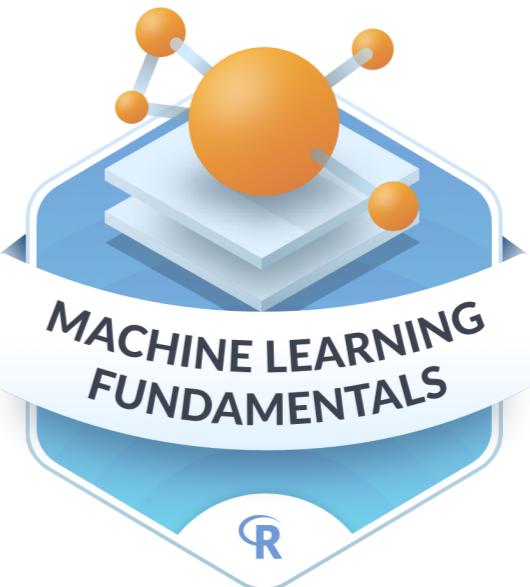
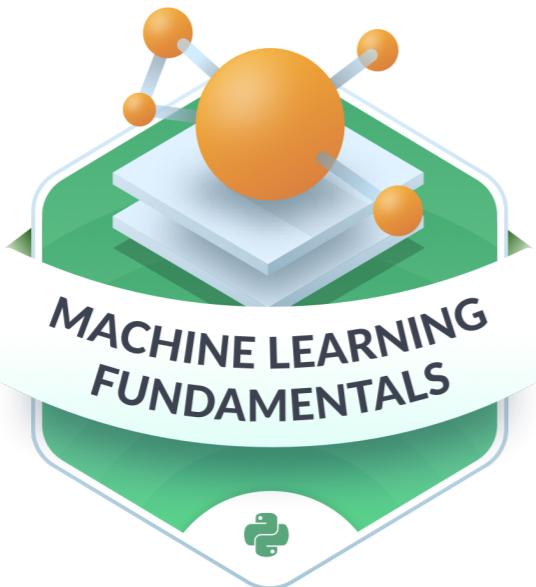
- Deep learning, including computer vision and natural language processing
- Limits of machine learning



# What's next?



# What's next?



# Congrats!

MACHINE LEARNING FOR EVERYONE