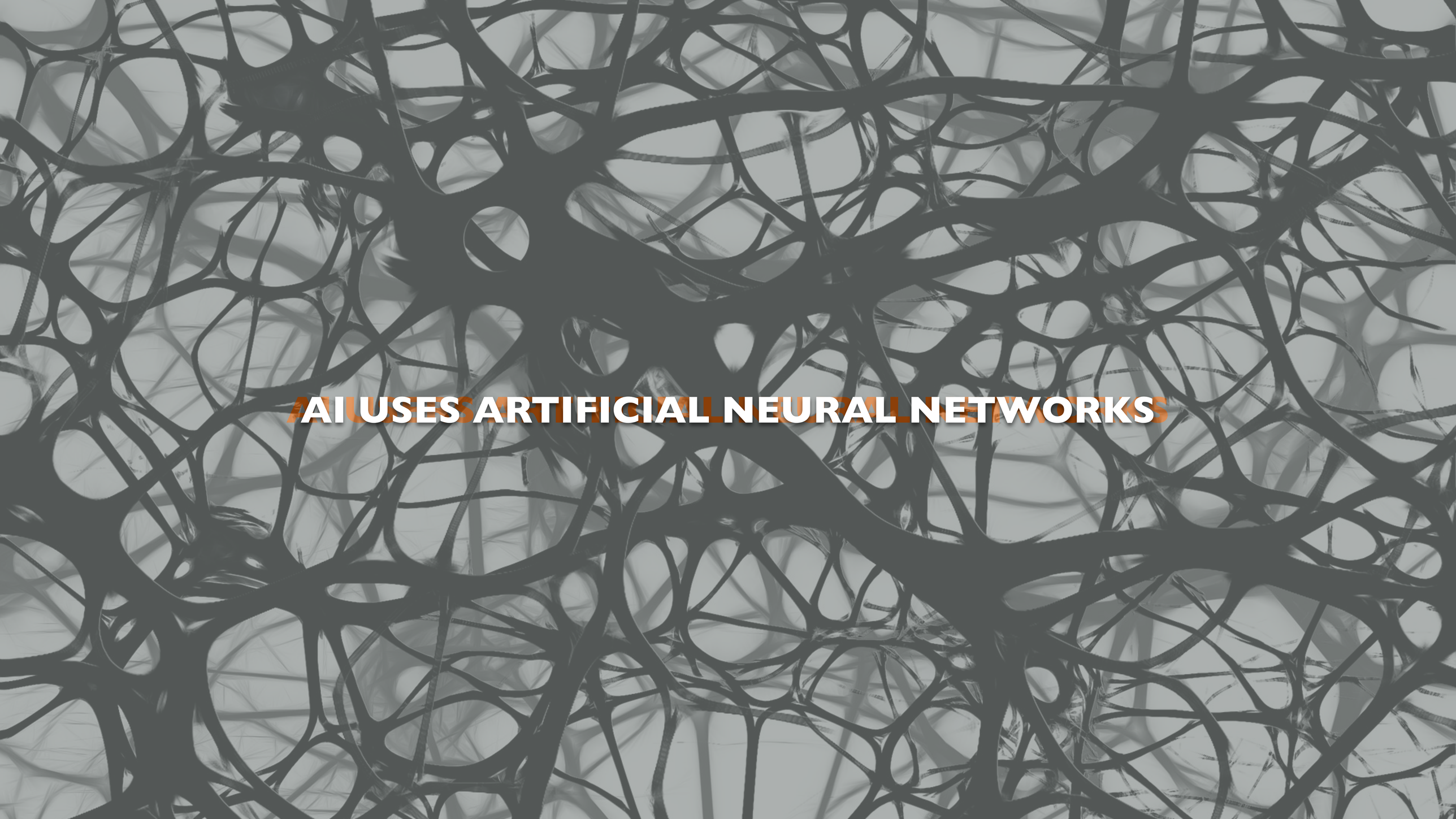
DATA

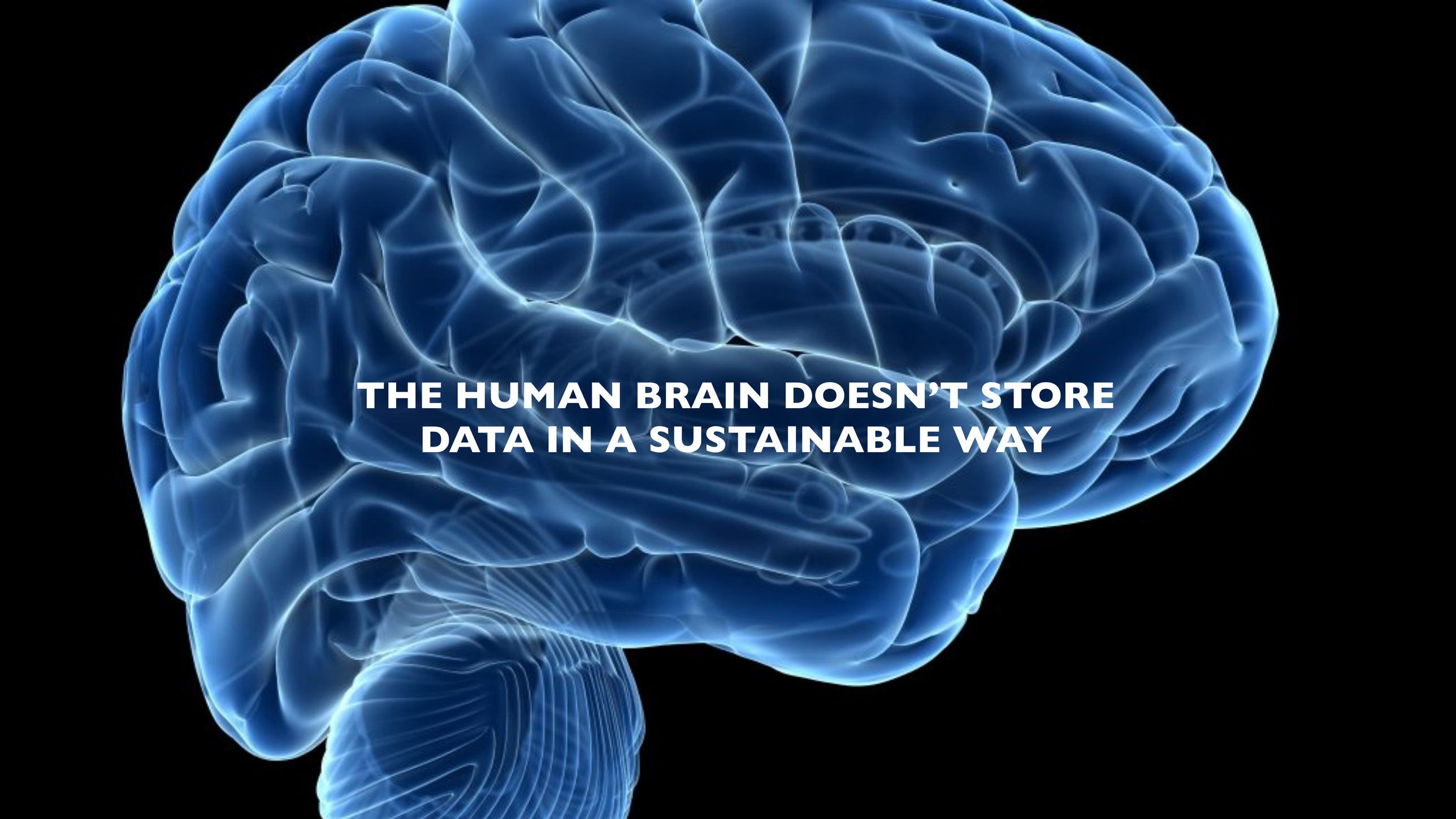
-

Data are the objectively observable **registration of ‘facts’** on a **particular medium,** such that these data can be **exchanged** and **stored** for a longer period of time.


NATURE STORES DATA IN DNA


ANIMALS / HUMAN BEINGS STORE INFORMATION
IN NERVES + BRAIN

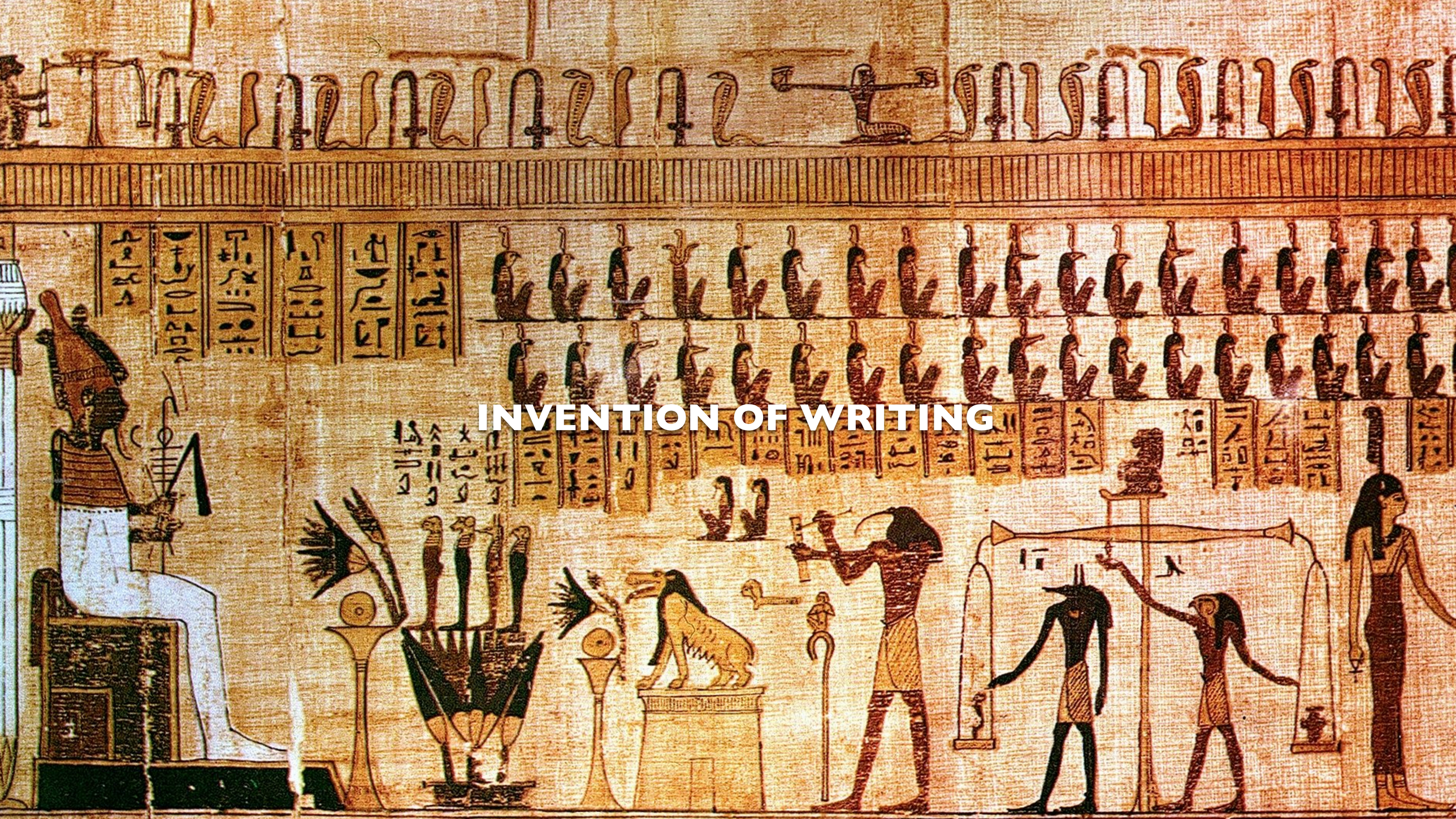

AI USES ARTIFICIAL NEURAL NETWORKS


THE HUMAN BRAIN DOESN’T STORE
DATA IN A SUSTAINABLE WAY


When you die,
everything you
learned is gone



PAINTINGS + PETROGLYPHS



INVENTION OF WRITING


PRINTING PRESS

TEXT

NUMBERS

LISTS

ILLUSTRATIONS

CAPTURED IMAGES

MOVING IMAGES

SOUND

90%

90% OF ALL DATA THAT HUMANITY HAS PRODUCED
DATES BACK TO THE LAST **10 YEARS**

SOCIAL MEDIA USAGE IN 2018/MINUTE

Twitter users sent 473,400 tweets
Snapchat users shared 2 million photos
Instagram users posted 49,380 pictures
LinkedIn gained 120 new users

GOOGLE SEARCHES

Google processes more than 40,000 searches every second, or 3.5 billion searches a day.

FACEBOOK USERS

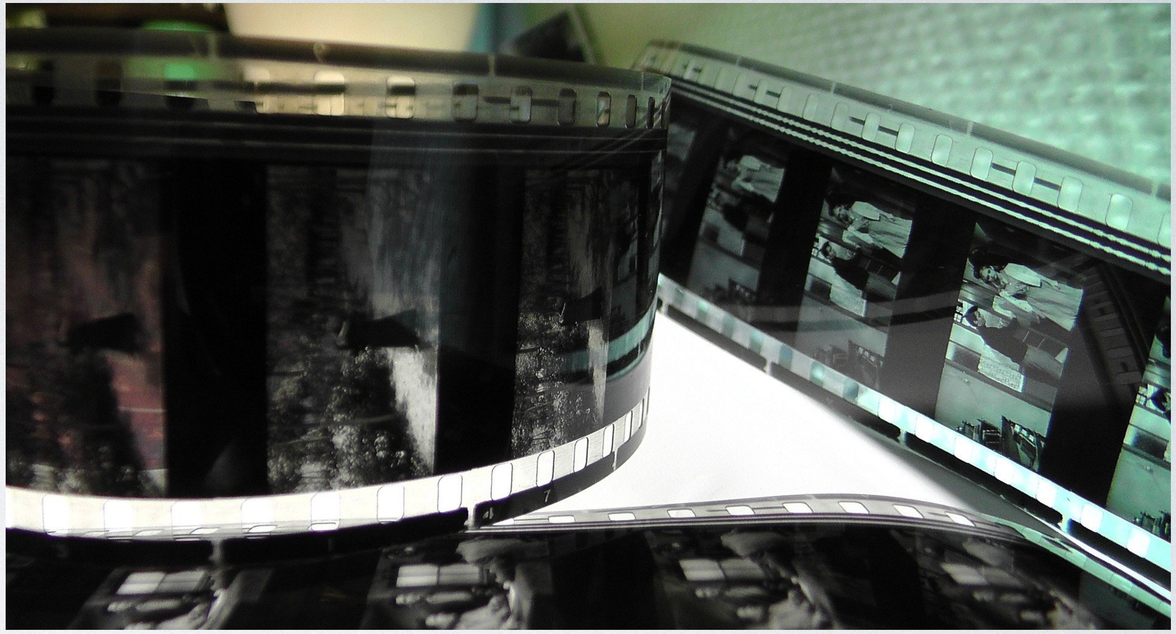
1.5 billion people are active on
Facebook every day.
That's one-fifth of the world's population.

SMARTPHONES

Two-thirds of the world's population now
own a mobile phone.

The **average household** annually creates enough data to **fill 65 iPhones of 32 gigabytes**. That means that we can all fill up so many iPads that we can make a stack that covers two-thirds the distance to the moon (Source: EMC)






2/55
HELIOS

ANALOG
MAGNETIC
OPTICAL
DIGITAL





MOST DATA IS DIGITAL

Digital describes electronic technology that **generates, stores, and processes data in terms of two states: positive (1) and non-positive (0)**

DIGITUS = FINGER

COUNTING ON YOUR FINGERS



When you digitize data, you convert it into numbers



A DIGITAL SYSTEM DOES NOT UNDERSTAND THE INFORMATION IT STORES

Some examples

EXAMPLE 1

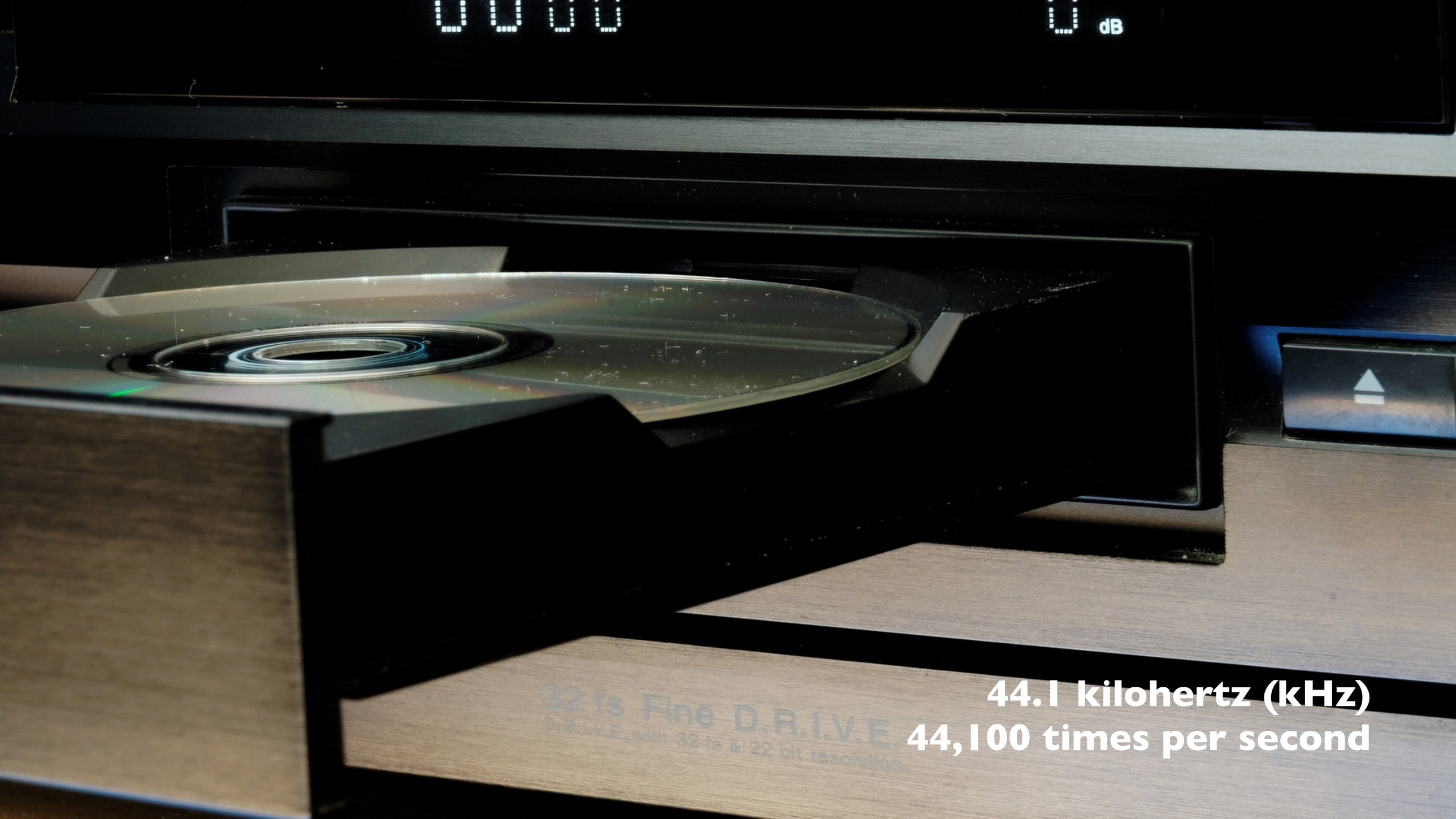
AUDIO + SOUND




VOLUME
PITCH

Digitizing sound means

1. *measuring* **volume** and **pitch**
2. Agree **how often you take measurements** (sampling rate)



44.1 kilohertz (kHz)
44,100 times per second

The result of digitization is a **series of numbers**

	PITCH	VOLUME
Frame 1	10200	120
Frame 2	10199	119,3
Frame 3	10198	119
…	…	…

For every second of (mono) digital music, a computer saves 44100 * 2 different numbers



A computer
doesn't know decimal numbers,
only **binary numbers**



Every decimal number must be **converted** to binary

Decimal numbers use 10 different characters to represent numbers

Binary numbers use 2 different characters to represent numbers

10 different characters:
0,1,2,3,4,5,6,7,8,9

2 different characters:
0,1

Position of digit in number (from right to left)	...	2	1	0
power or exponent "n".	...	Power 2	Power 1	Power 0
factor decimal	...	10^2	10^1	10^0
factor hexadecimal	...	16^2	16^1	16^0
factor binary	...	2^2	2^1	2^0

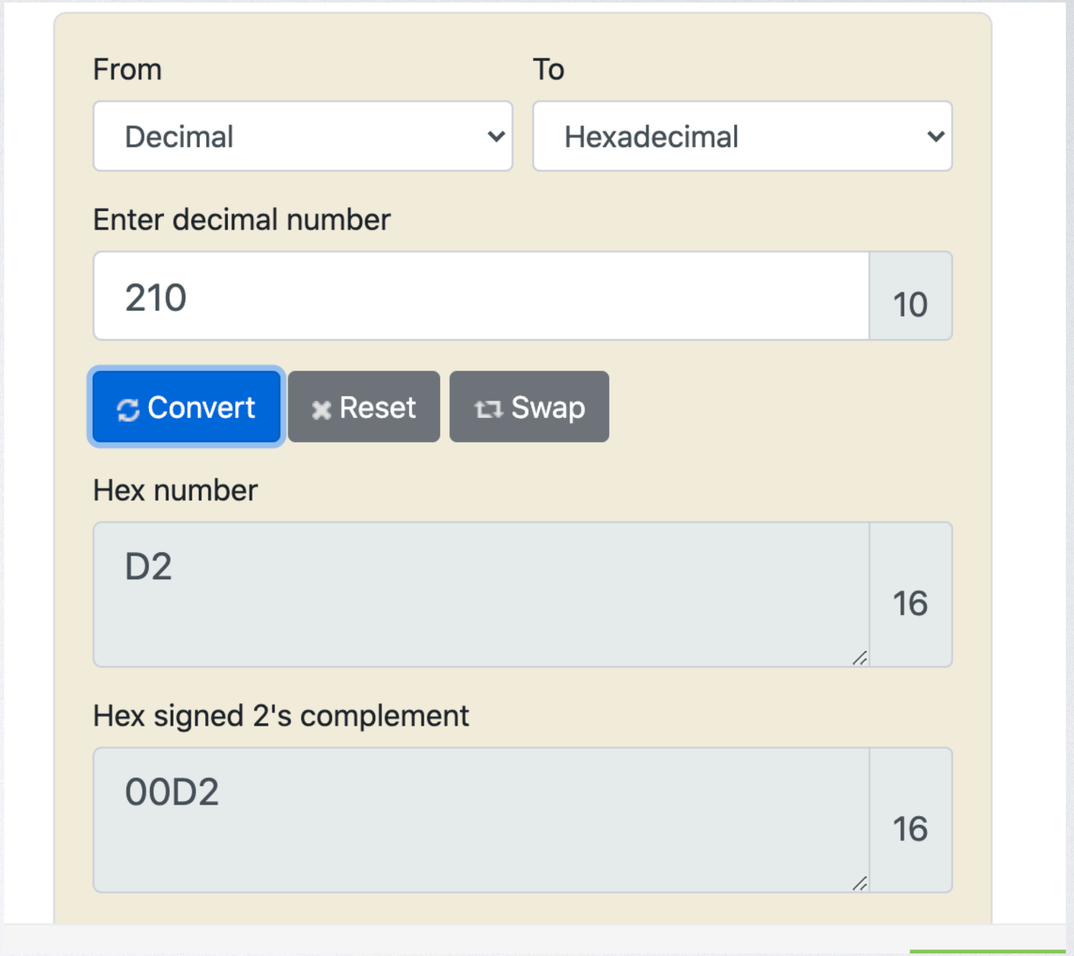
HEXADECIMAL

16 different characters:

0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F

A COOL TOOL

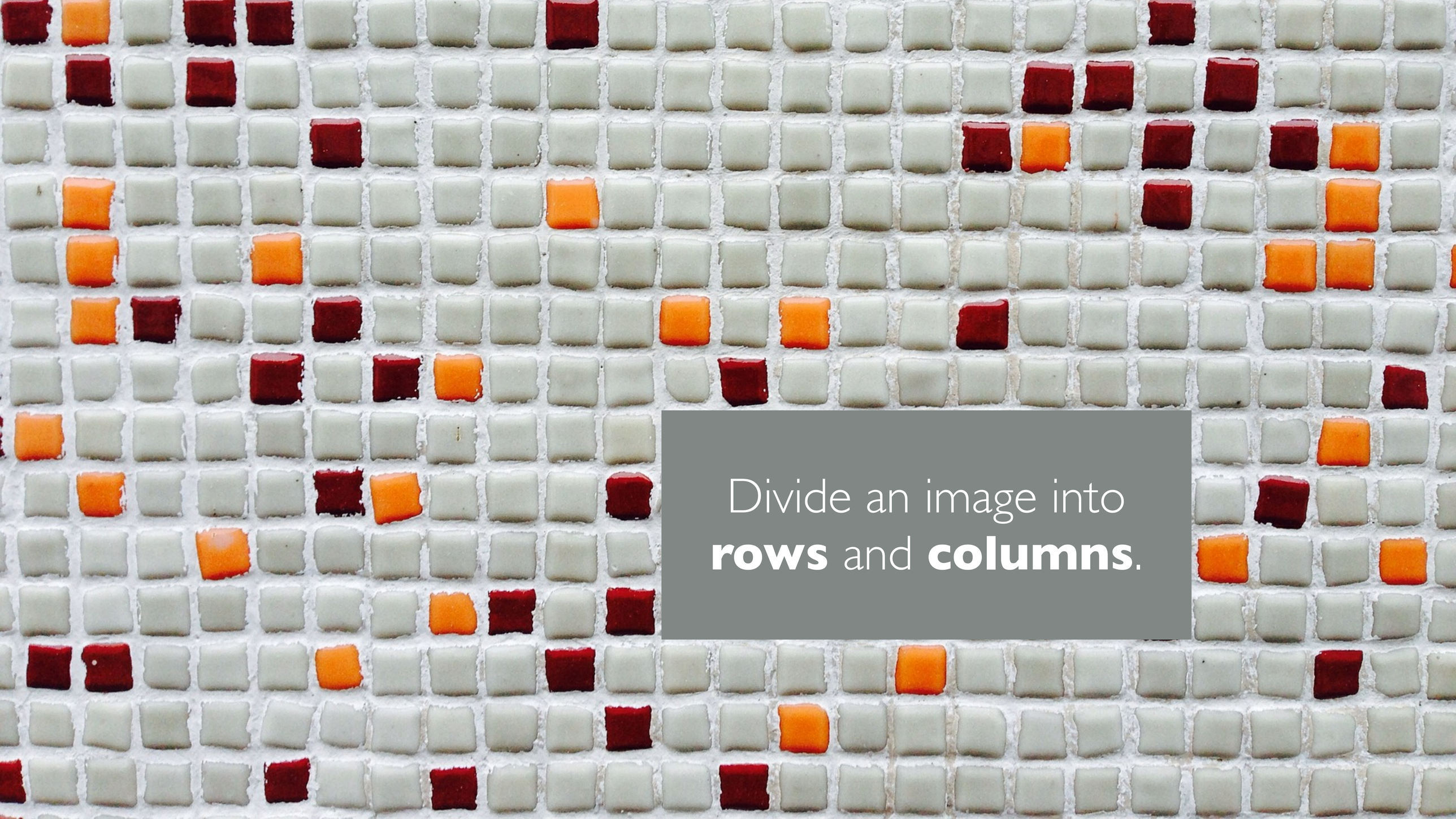
https://www.rapidtables.com/convert/number/decimal-to-hex.html

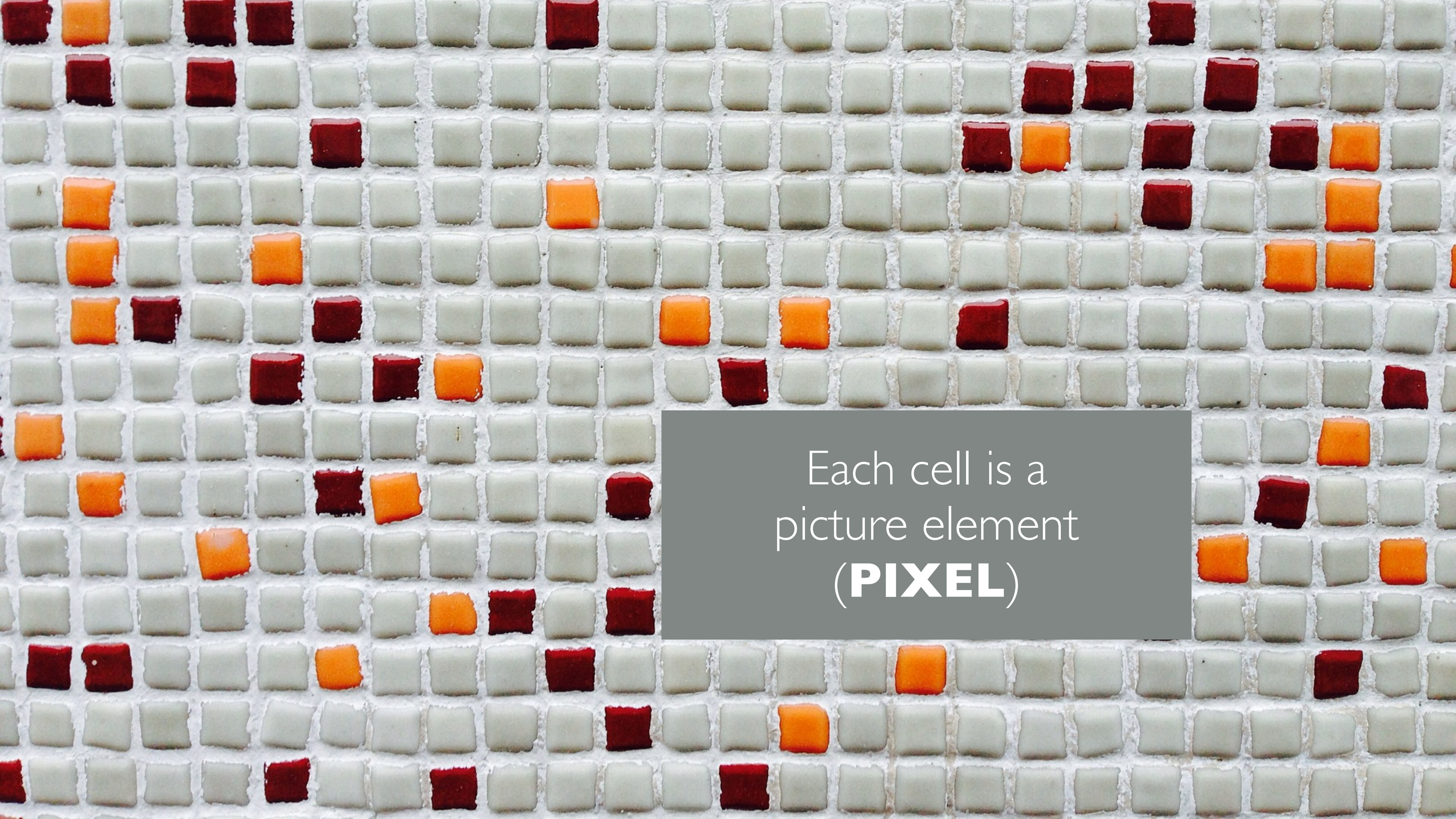


EXAMPLE 2

IMAGES




Divide an image into
rows and **columns**.


Each cell is a
picture element
(**PIXEL**)

Digitizing images means

1. *measuring the* ***color value*** *of each cell*
2. Determine how many rows and columns you will use (**number of pixels**)



On a screen each color is composed of a combination of **red**, **green**, **blue**

Digitizing images means

1. *measuring the **RGB values** of each cell*
2. determine how much color variations of red, green and blue you want to measure (**Color depth**)
3. Determine how many rows and columns you will use (**number of pixels**)

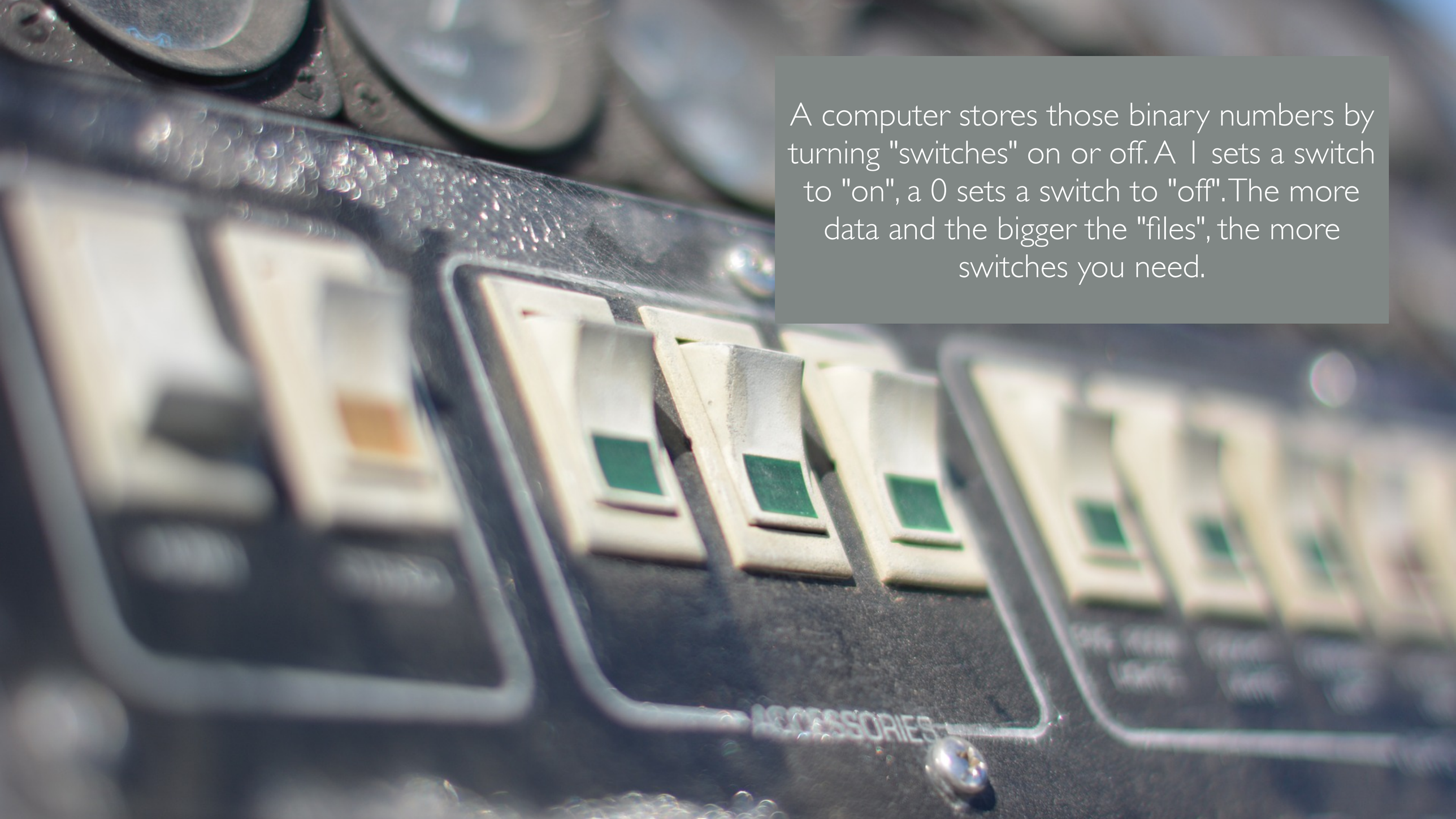
For every pixel in an image you need to store 5 different numbers

X-position	Y-position	R	G	B
1	1	255	0	0
2	1	0	255	0
3	1	0	0	255
…	…	…	…	…

more measurements?

More “numbers” to “remember”

File size gets bigger



A computer stores those binary numbers by turning "switches" on or off. A 1 sets a switch to "on", a 0 sets a switch to "off". The more data and the bigger the "files", the more switches you need.

EXAMPLE 3

TEXT DATA



HOW CAN YOU CONVERT TEXT INTO NUMBERS?



Each key on a keyboard is assigned a specific number.

Letter	ASCII Code	Binary	Letter	ASCII Code	Binary
a	097	01100001	A	065	01000001
b	098	01100010	B	066	01000010
c	099	01100011	C	067	01000011
d	100	01100100	D	068	01000100
e	101	01100101	E	069	01000101
f	102	01100110	F	070	01000110
g	103	01100111	G	071	01000111
h	104	01101000	H	072	01001000
i	105	01101001	I	073	01001001
j	106	01101010	J	074	01001010
k	107	01101011	K	075	01001011
l	108	01101100	L	076	01001100
m	109	01101101	M	077	01001101
n	110	01101110	N	078	01001110
o	111	01101111	O	079	01001111
p	112	01110000	P	080	01010000
q	113	01110001	Q	081	01010001
r	114	01110010	R	082	01010010

Digitizing text data means

1. ***Map*** *any "symbol/character" to a specific number*
2. Make ***international agreements***

AMERICAN STANDARD CODE FOR INFORMATION INTERCHANGE

ASCII

SOME PROBLEMS WITH ASCII

lack of support for non-western symbols and characters

UTF-8 COMES TO THE RESCUE

UTF-8

If you enter a number on your keyboard,
ASCII / UTF-8 causes strange "problems"

48	30	060	0	**0**
49	31	061	1	**1**
50	32	062	2	**2**
51	33	063	3	**3**
52	34	064	4	**4**
53	35	065	5	**5**
54	36	066	6	**6**
55	37	067	7	**7**
56	38	070	8	**8**
57	39	071	9	**9**

0 = 48 in the ascii table

48 decimal = 110000 in binary

This means: you need 6 “memory slots” to store a 0

Saving text data as ASCII / UTF-8 is not always smart.
After all, it often produces larger files,
especially if you only want to keep numerical data.

For this reason, text data is often no longer saved as ASCII / UTF-8, but encoded in such a way that the information takes up less space.

When you encode data in this way, you speak of **BINARY DATA**.
If you keep pure ASCII / UTF-8, we speak of **TEXT DATA**.

EXAMPLES OF BINARY ENCODED FILES

DOCX, PPTX, XLS, JPG, MP4, PDF…

Binary encoded is a misleading name because all digital data is "binary". Binary encoded in this case means that the data has not been saved as ASCII or UTF-8. In our lessons we will mainly work with ASCII / UTF-8 encoded files.

BINARY ENCODED FILES

TEXT DATA
ASCII/UTF-8 ENCODED FILES



A DIGITAL SYSTEM DOES NOT UNDERSTAND THE INFORMATION IT STORES


Biggest problem for
automating
data analysis