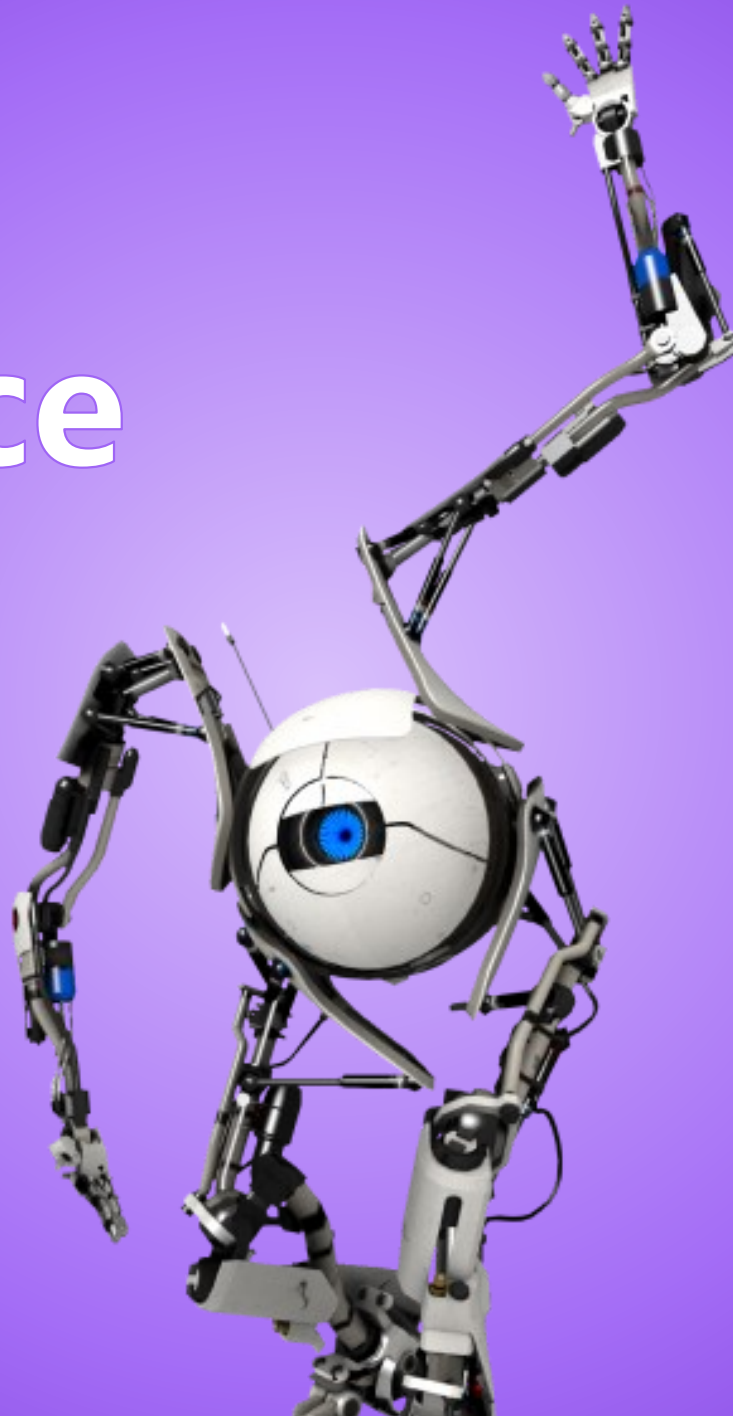


artificial
intelligence
tools

vs.
privacy



Steve Muller

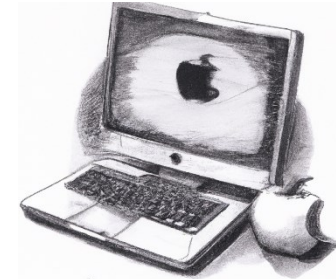
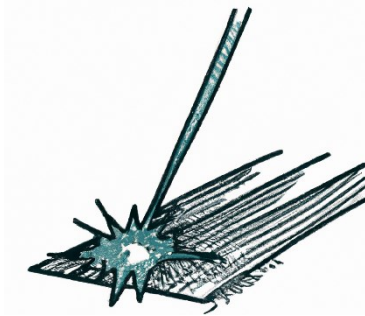
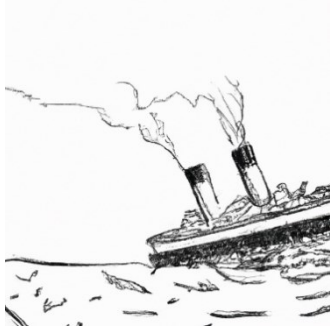
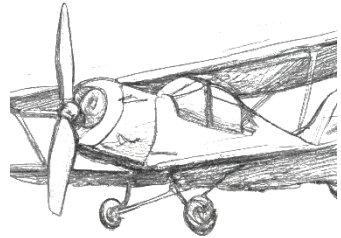
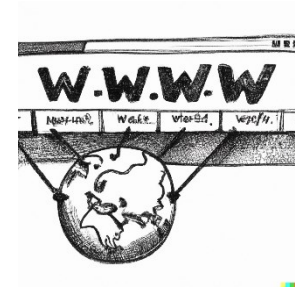
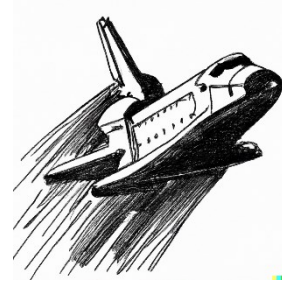
2022



ChatGPT

1956

term "AI"



1900

1920

1940

1960

1970

1980

1990

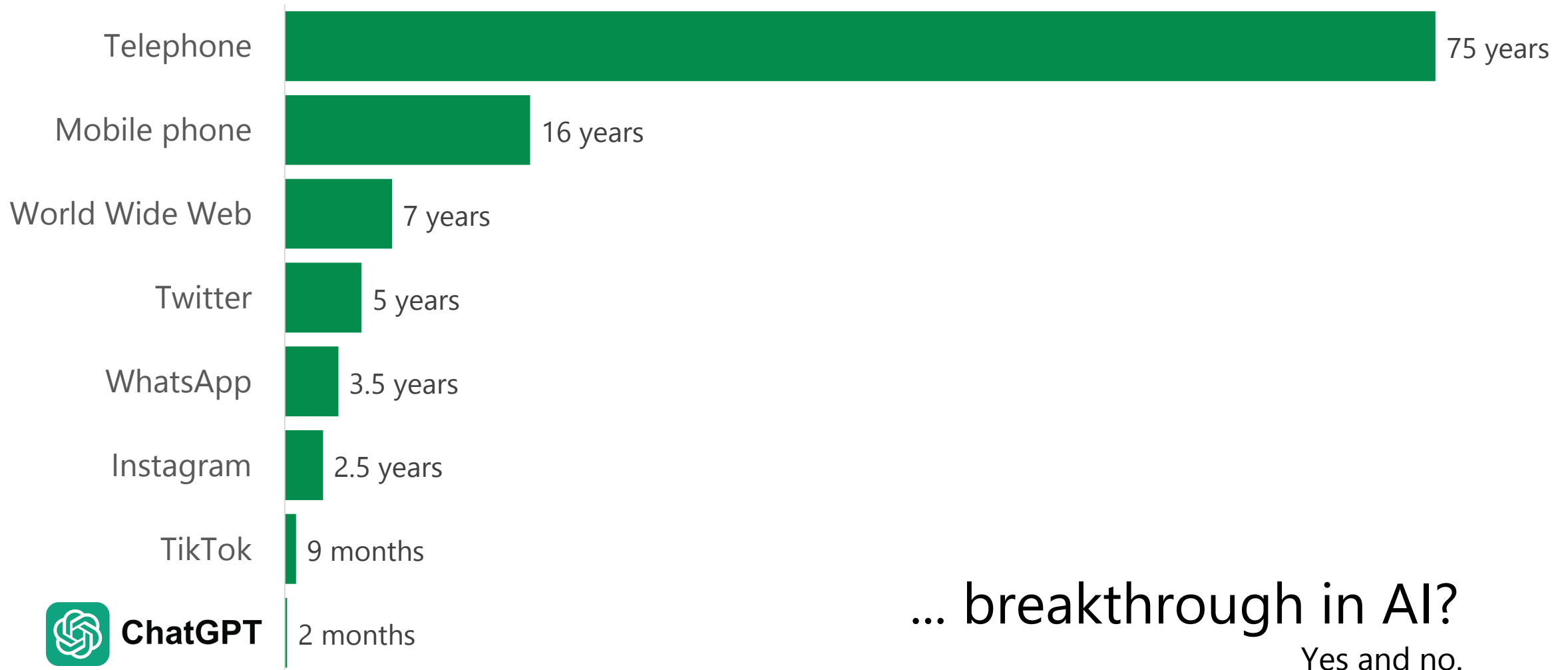
2000

2010

"AI winter"

"AI winter"

Time to reach 100 million users



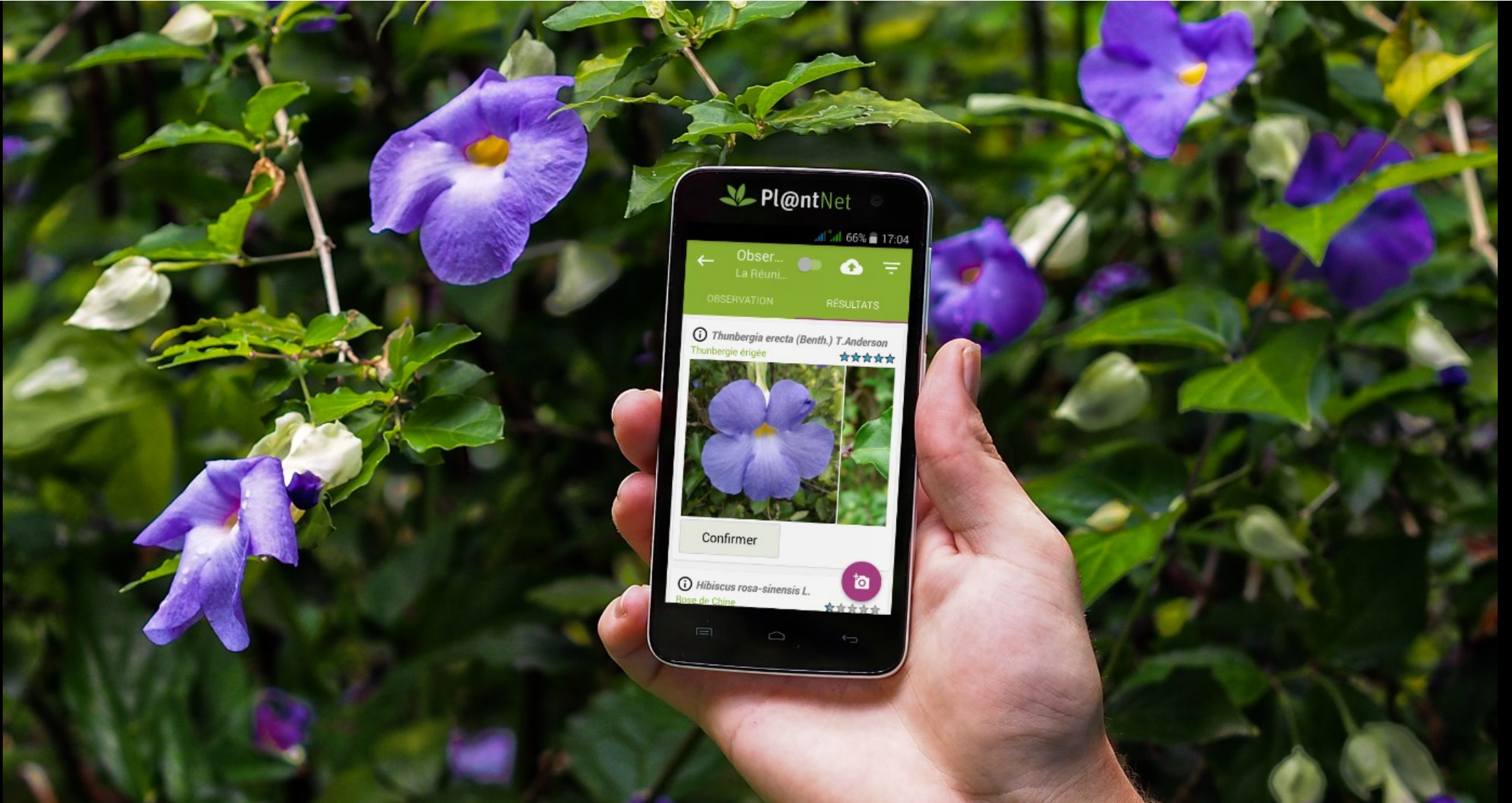
1

recommendation systems



2

image recognition





voice recognition

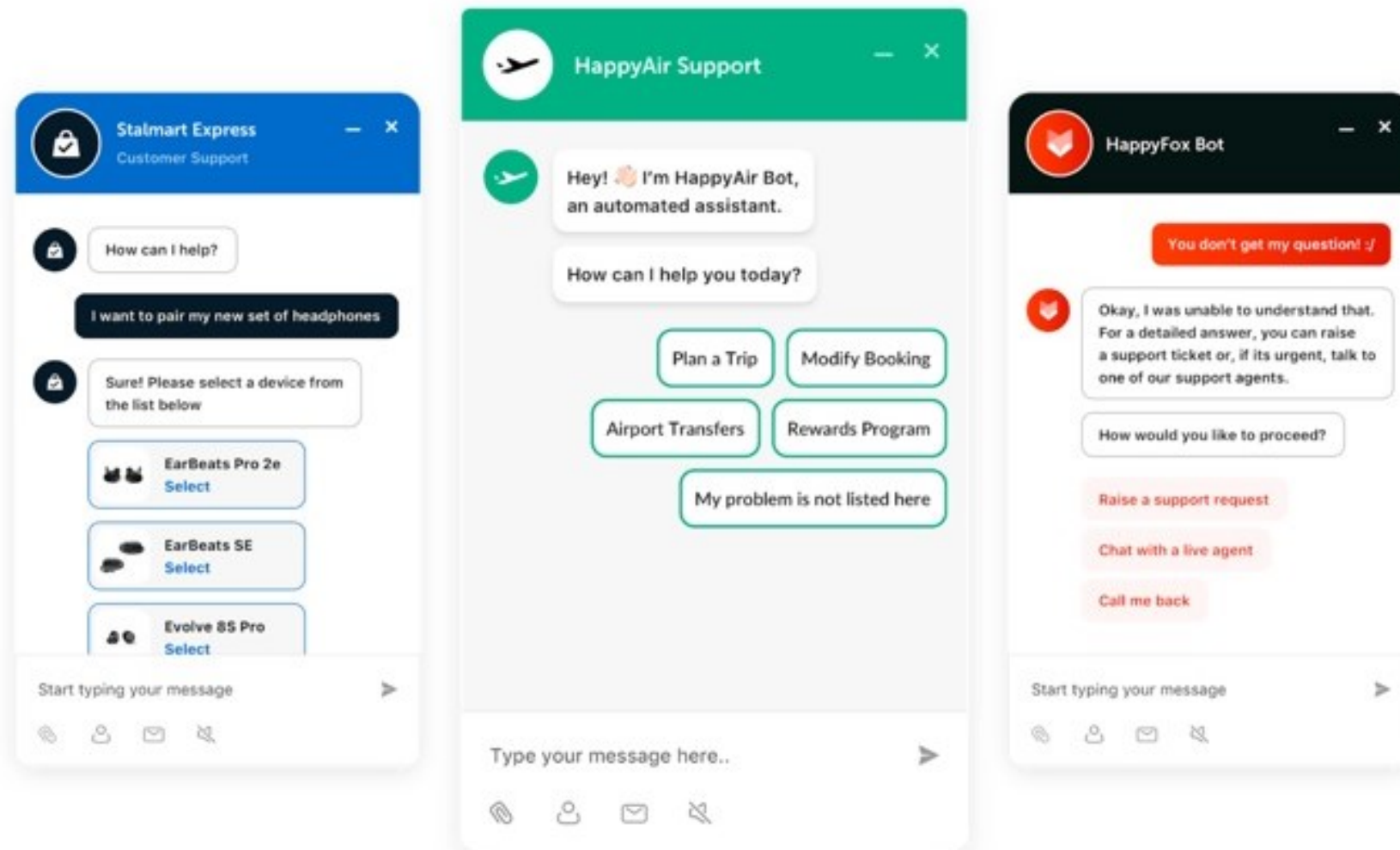


Siri



4

text generation





self-driving cars



... so what has brought us ChatGPT?

WHAT IS REALLY NEW?

WHY THE HYPE?

- ◆ 1 step closer to true intelligence



- ◆ **democratization** of AI

what is

artificial intelligence?

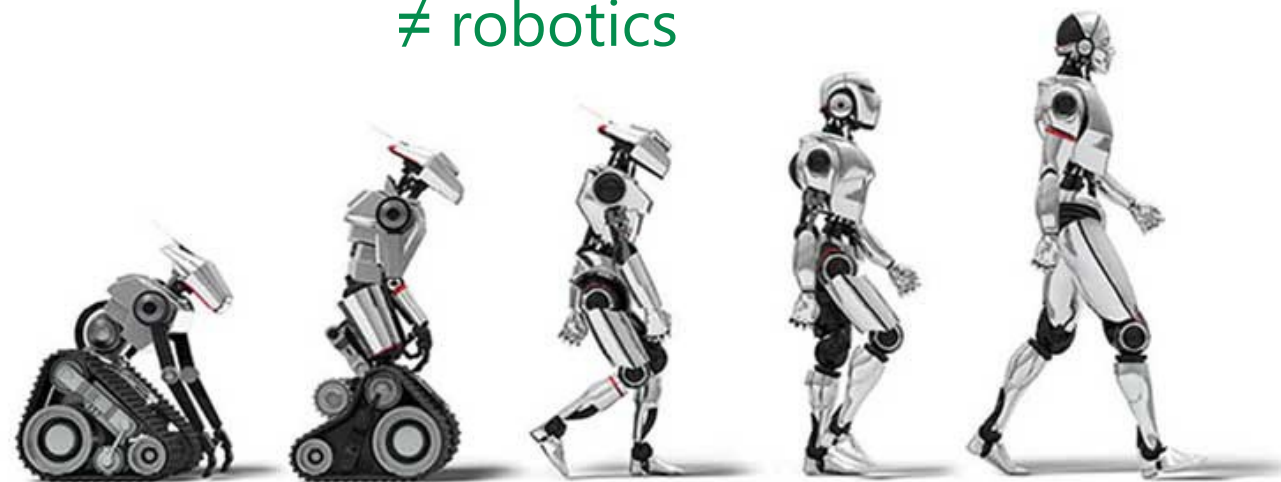
perception
synthetization
deduction

} of information by a machine

≠ thinking

≠ robotics

≠ conscience



artificial intelligence



artificial **general** intelligence

“machine that can accomplish **any intellectual task** that human beings or animals can perform”

narrow artificial intelligence

“goal-oriented version of AI designed to better perform **a single task**”



— DISCIPLINE —

Artificial intelligence

perception, synthetization, deduction
of information by a machine

— TECHNIQUES —

Machine learning

prediction based on *known* facts

Data mining

discovery of *new* facts

imitation of behaviour

discovery of
patterns / similarities

DISCIPLINE

Artificial intelligence

perception, synthetization, deduction
of information by a machine

TECHNIQUES

Machine learning

prediction based on *known* facts

Data mining

discovery of *new* facts

ALGORITHMS

clustering

regression

support vector machine

artificial neural networks

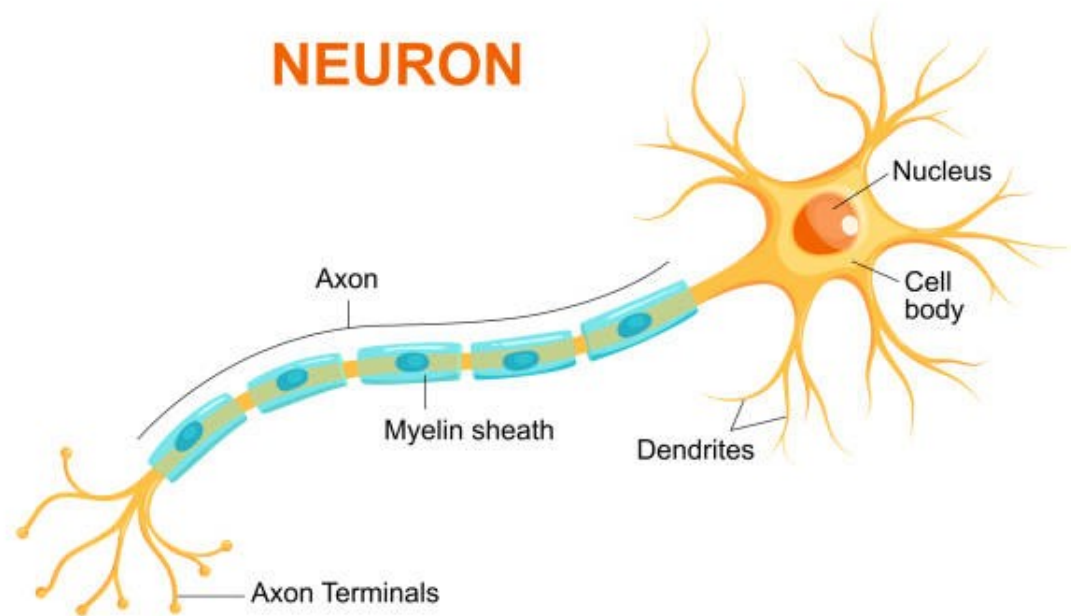
decision tree

naive Bayes

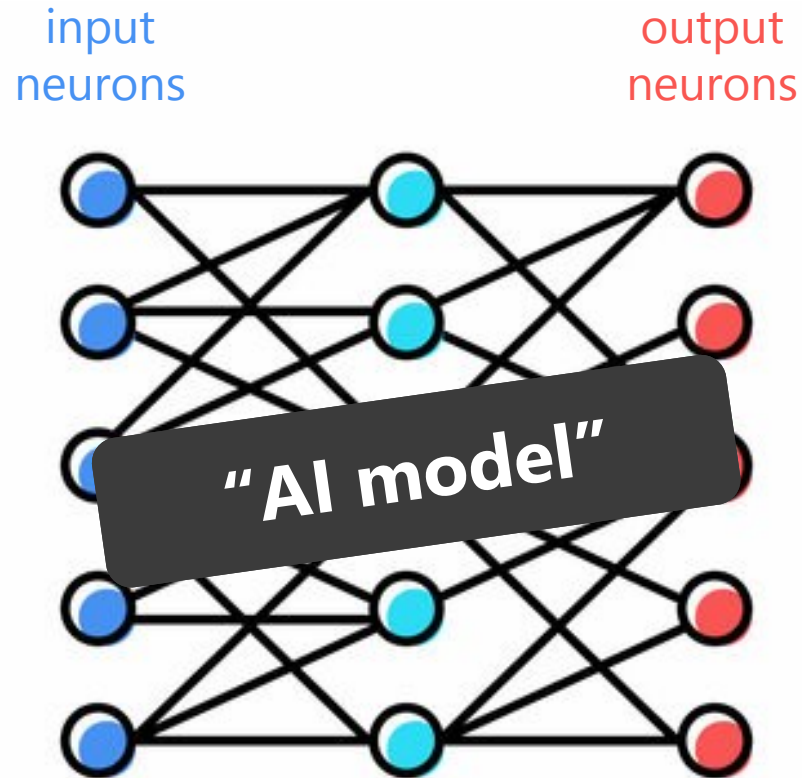


-artificial neural networks

NEURON

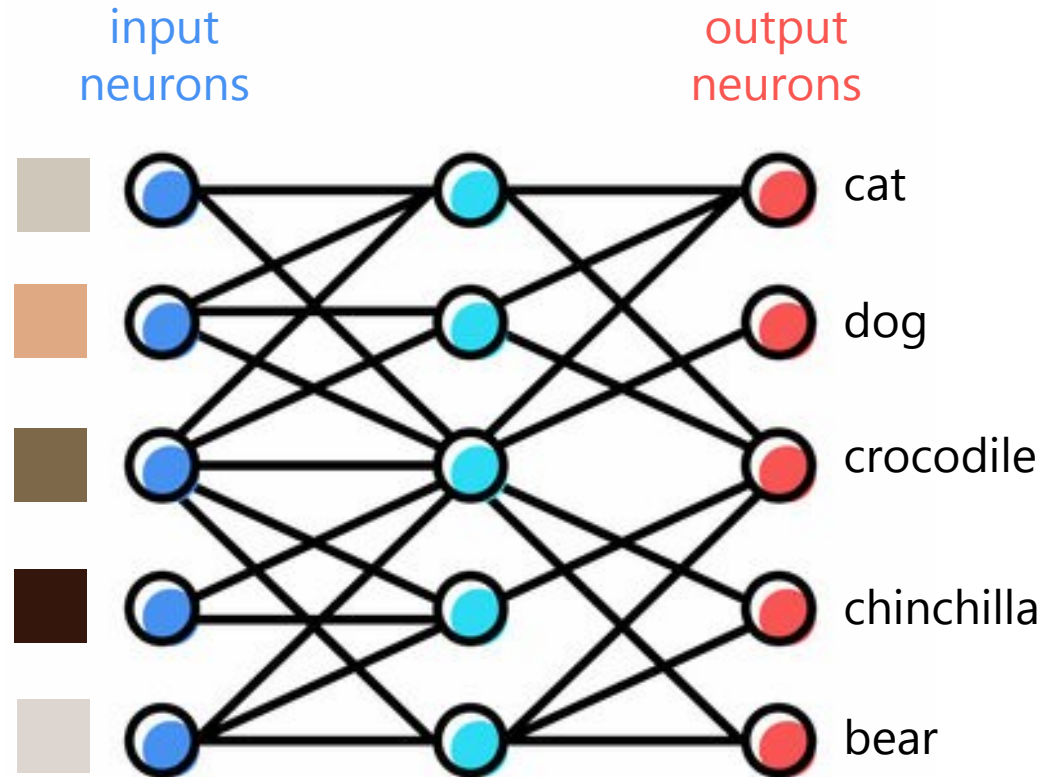
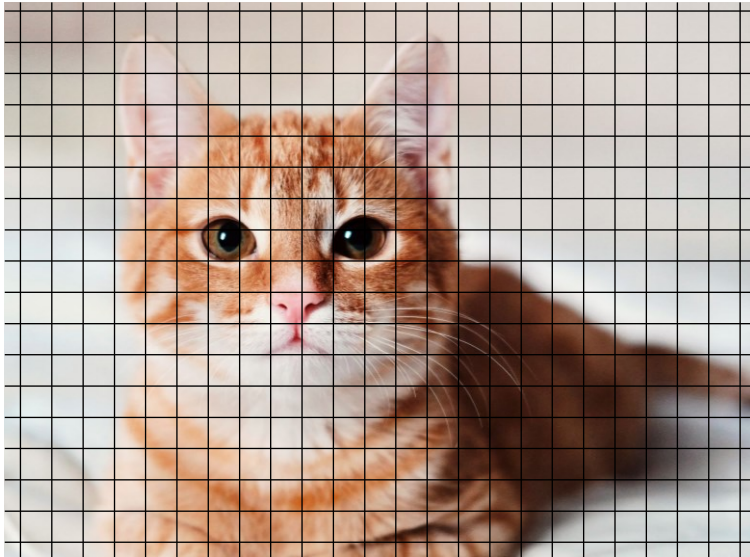


Artificial neural network



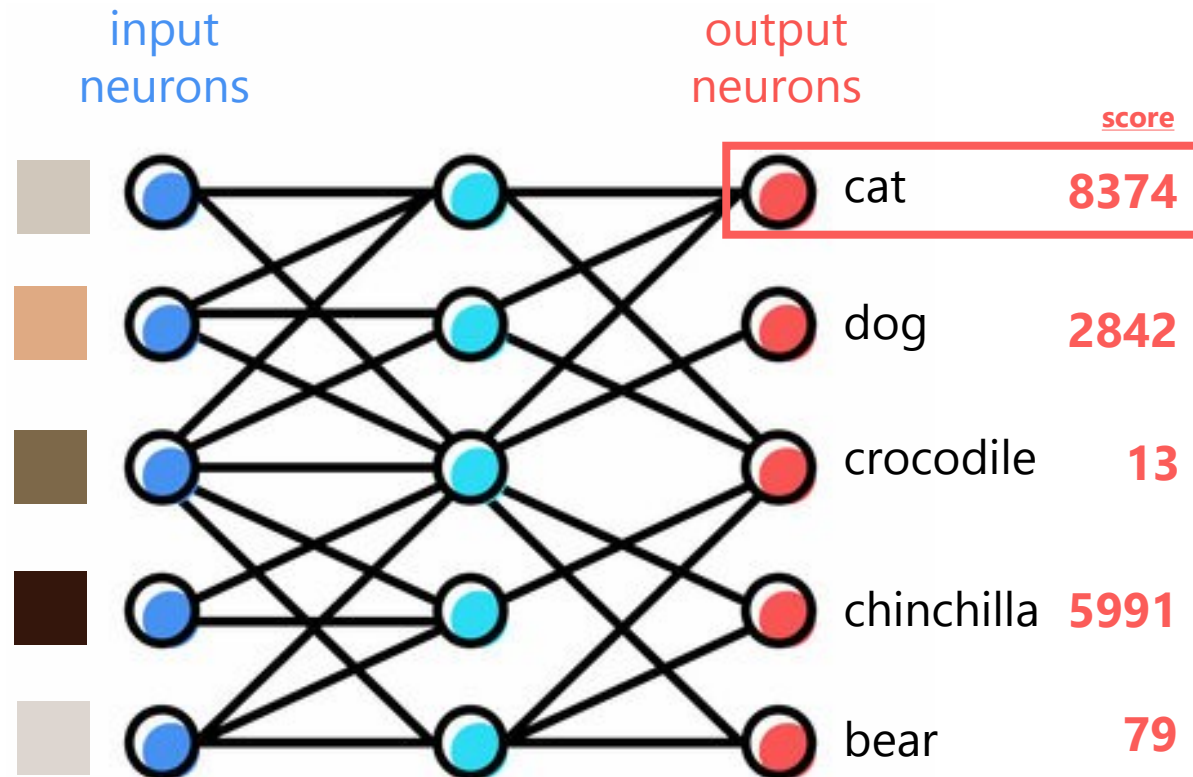
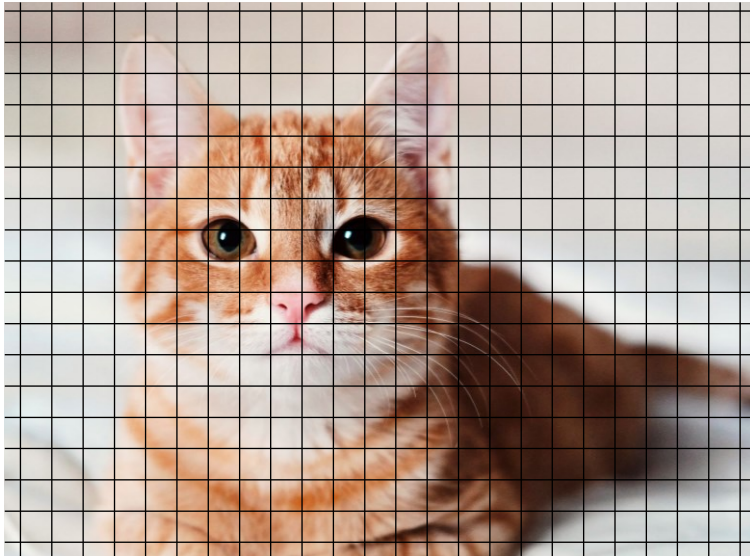
"see something" → "return/do something"

Artificial neural network



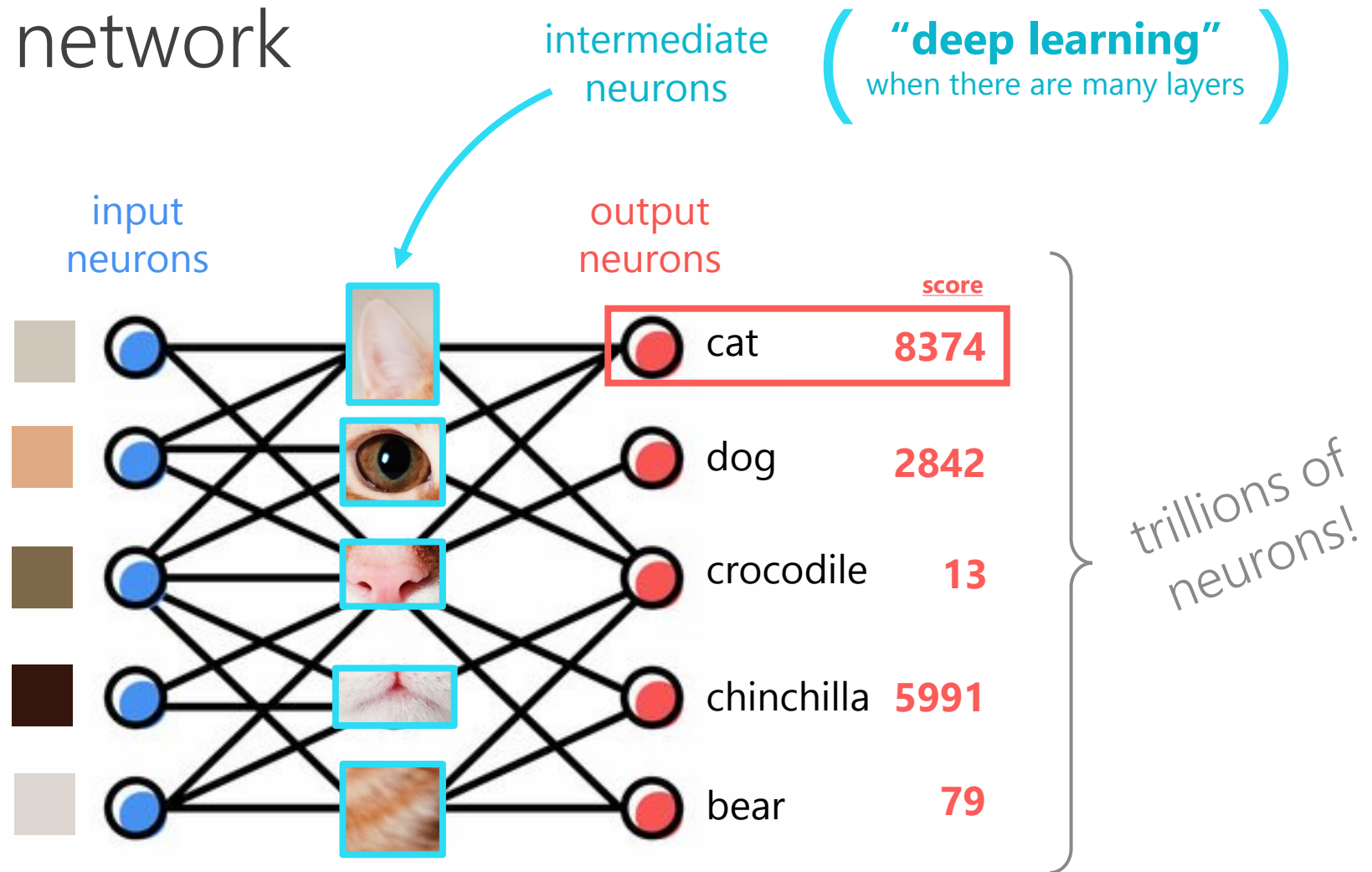
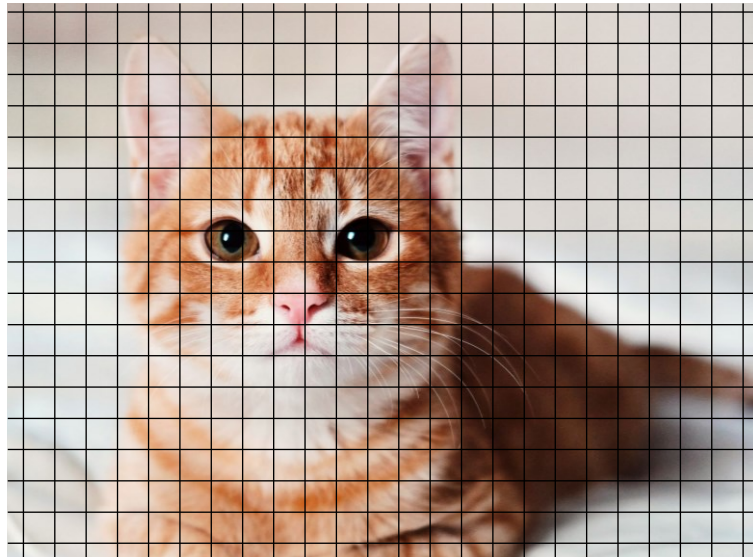
“see something → return/do something”

Artificial neural network



“see something → return/do something”

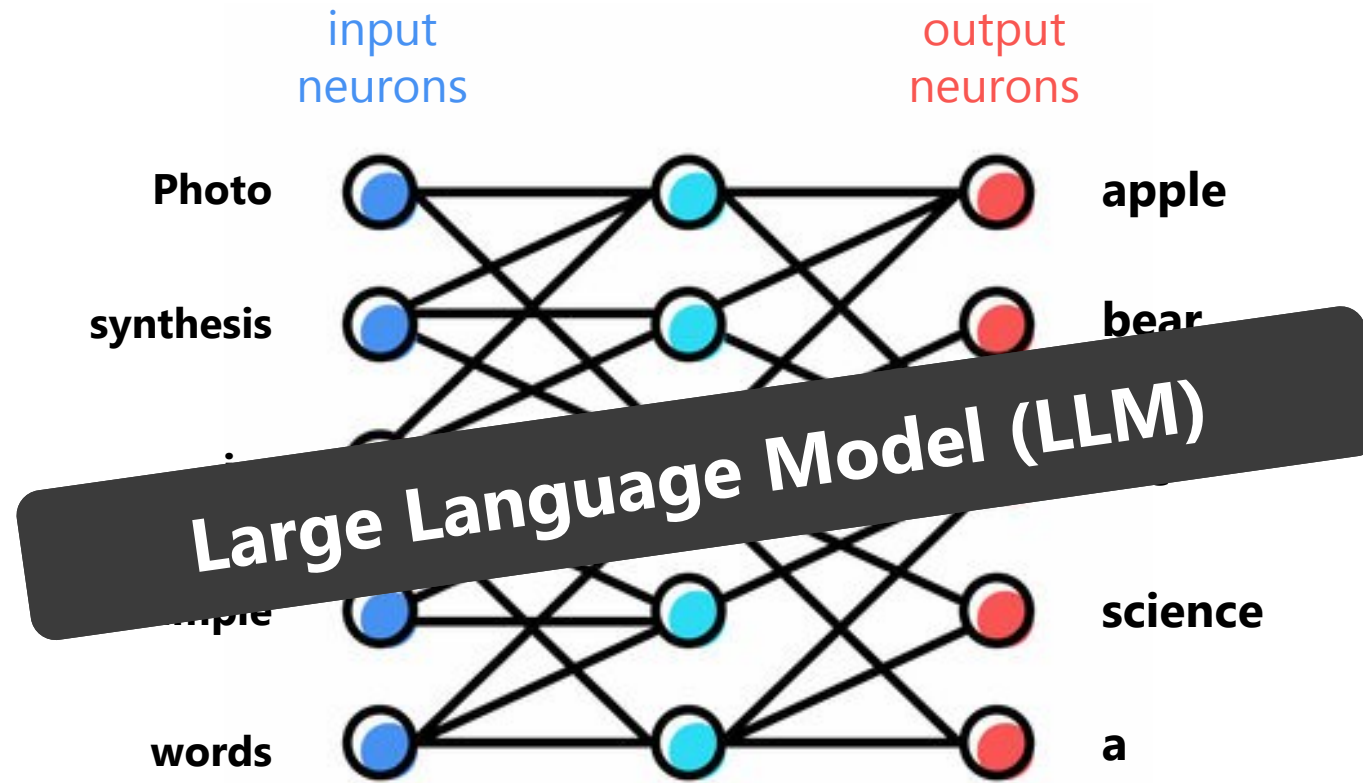
Artificial neural network



“see something” → “return/do something”

Artificial neural network (text generation)

"Photosynthesis,
in simple words,
is ..."



conversation up to now

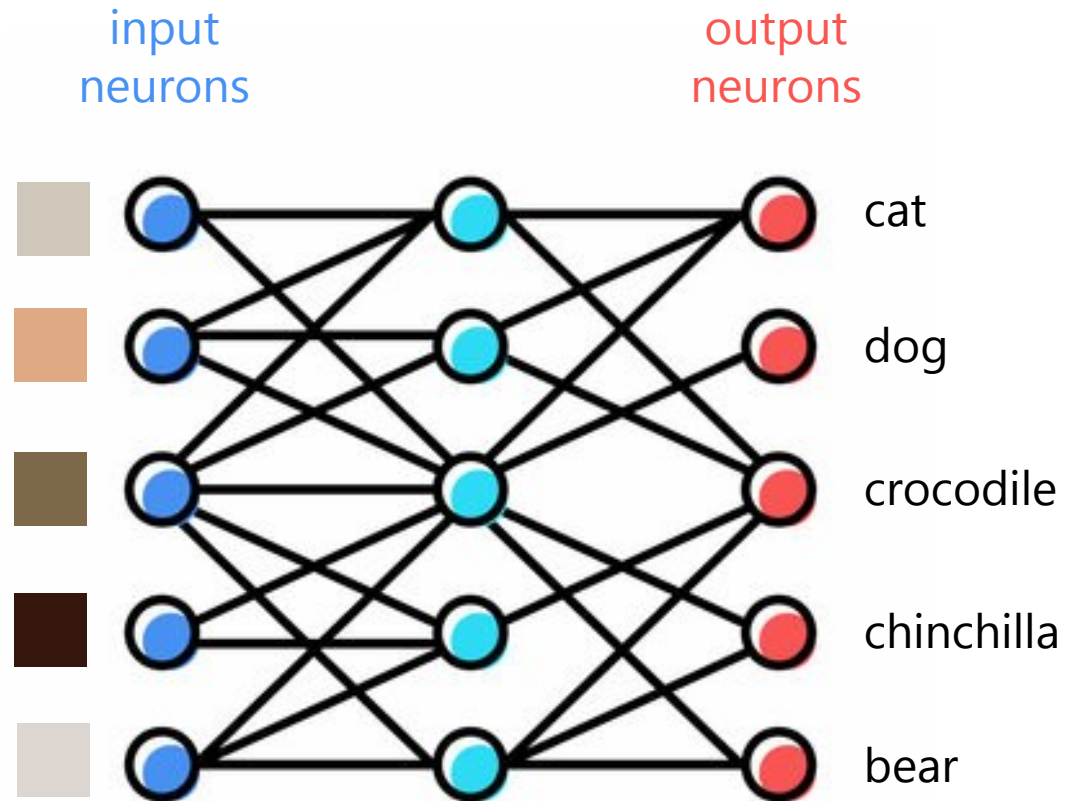
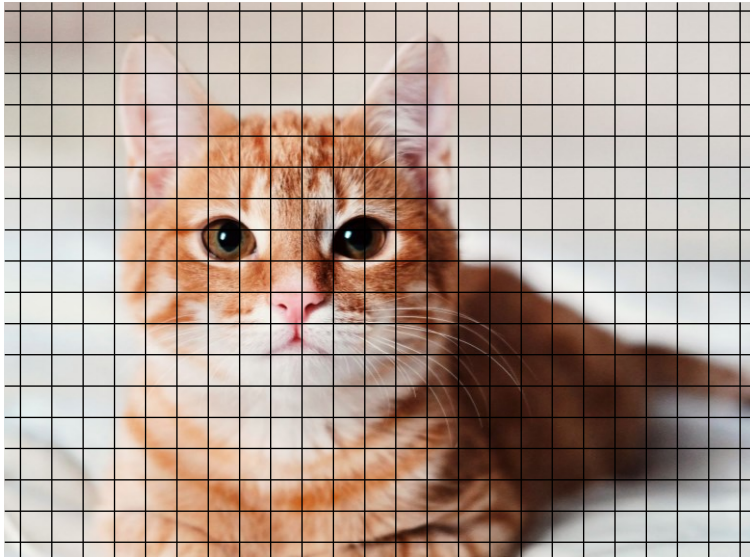


predict next word



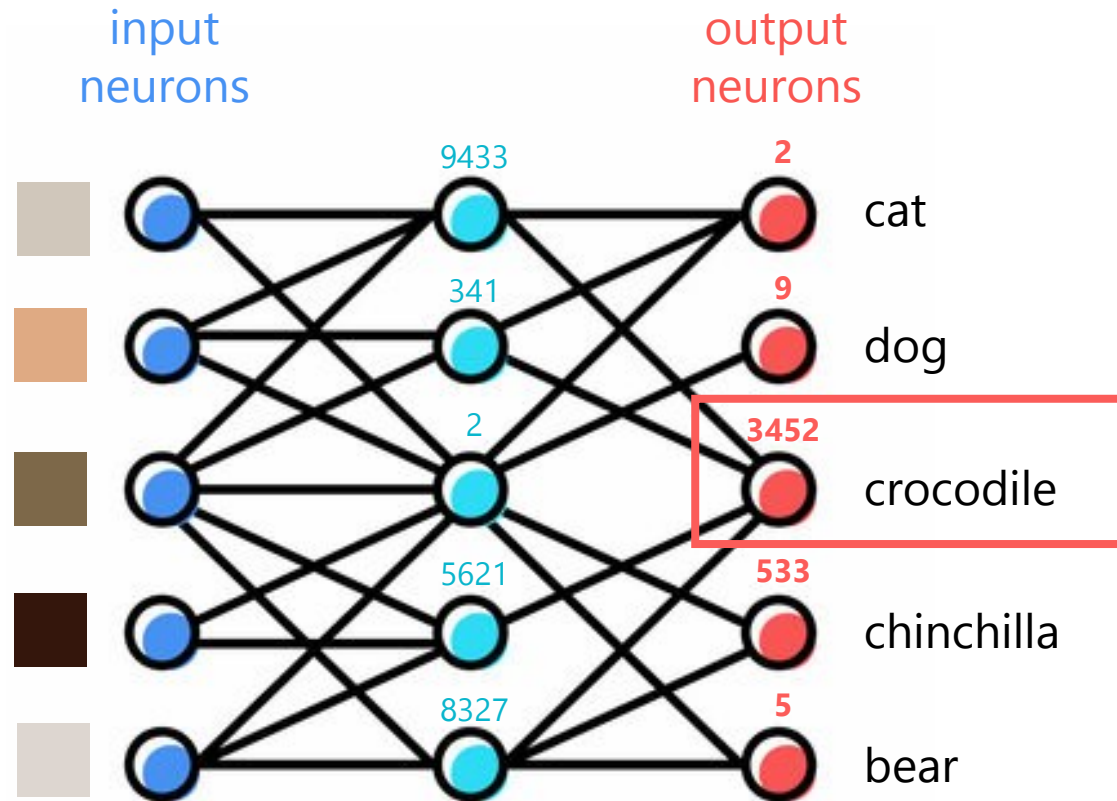
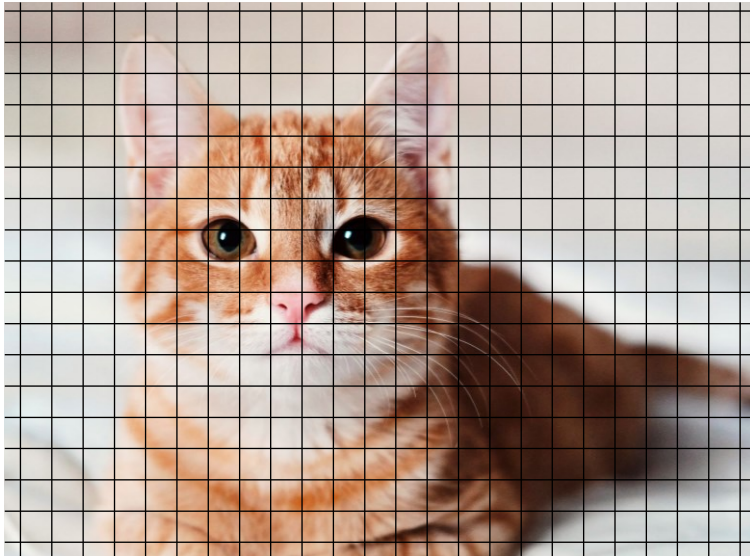
training an

Artificial neural network



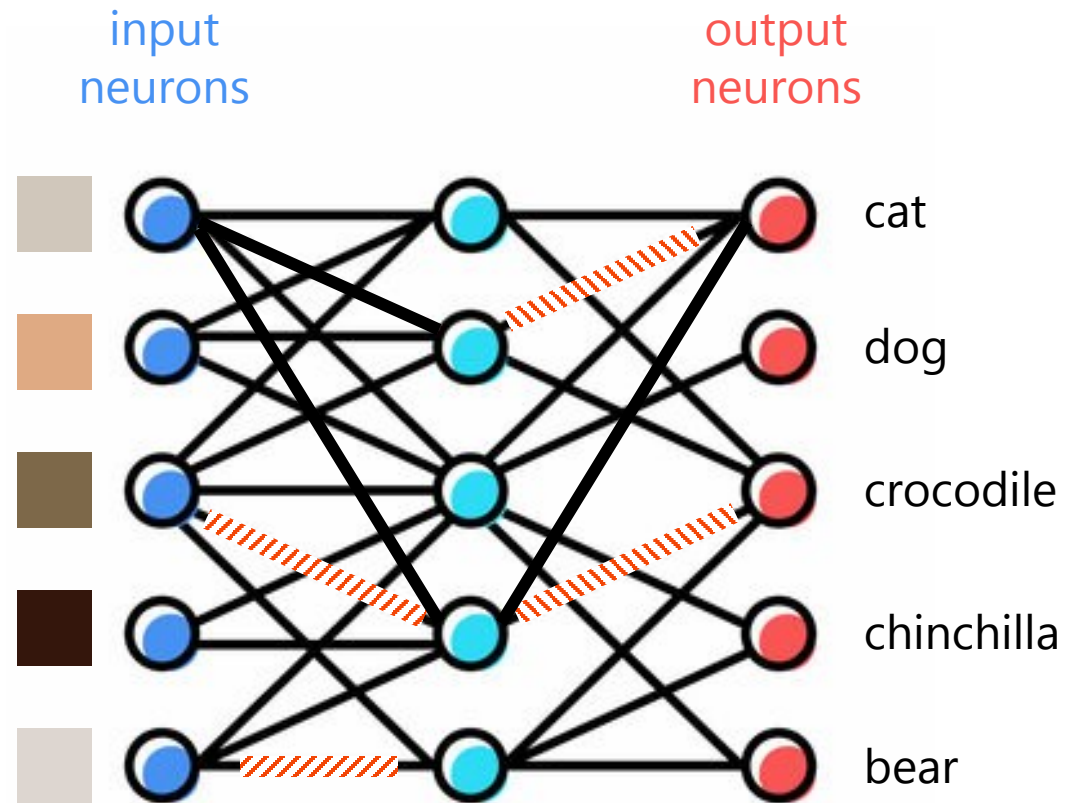
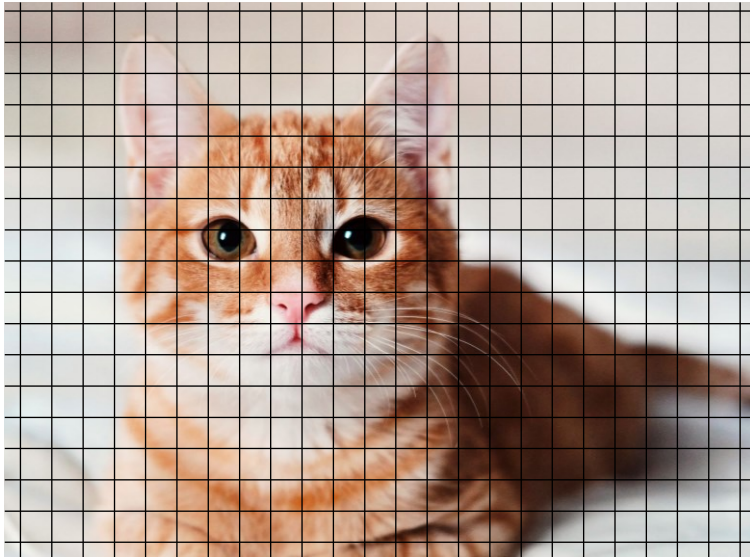
training an

Artificial neural network



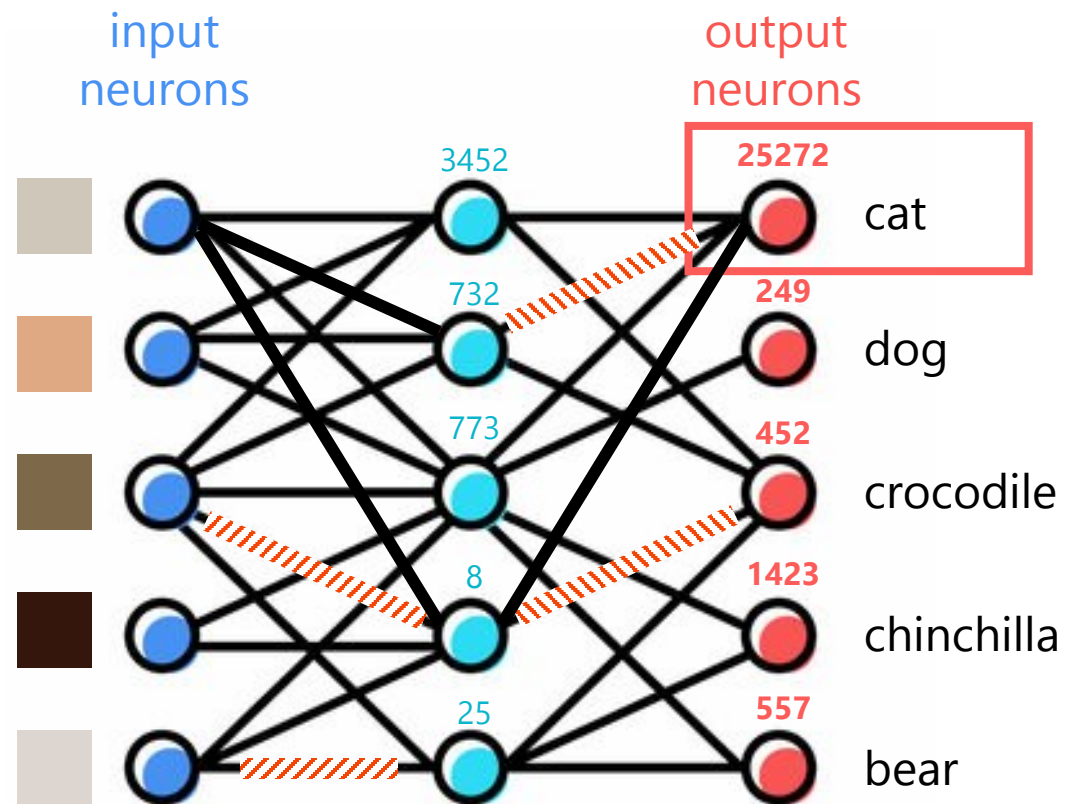
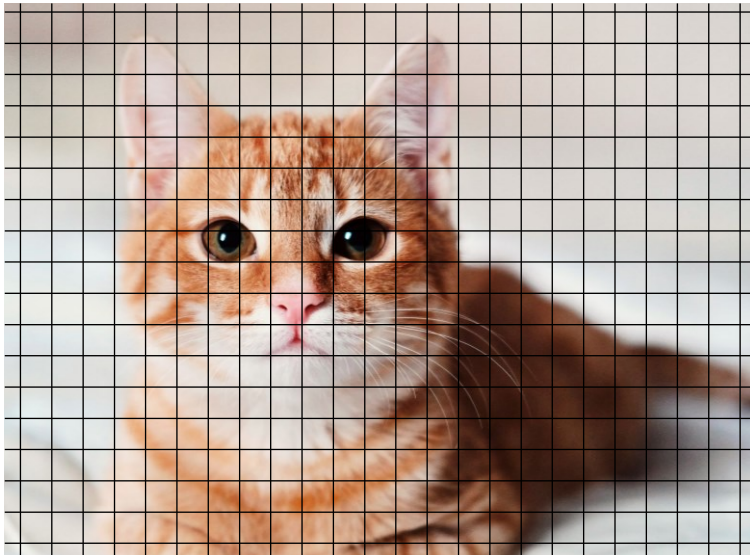
training an

Artificial neural network



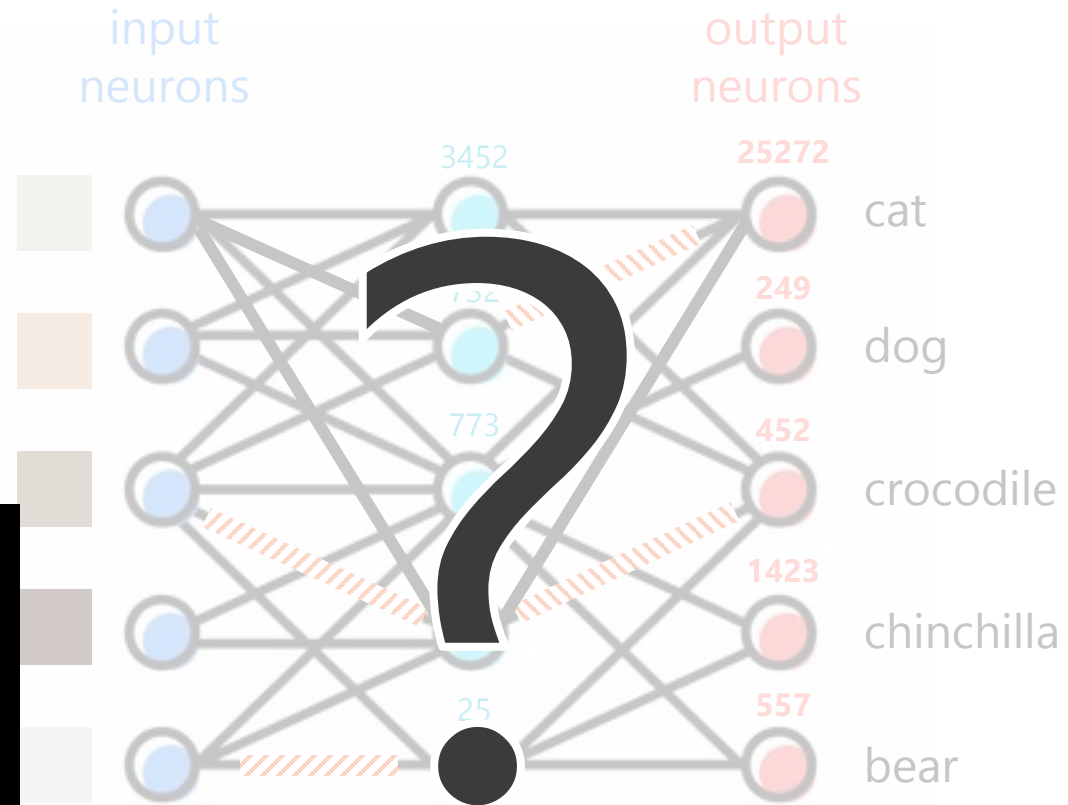
training an

Artificial neural network



training an

Artificial neural network



training an

Artificial neural network

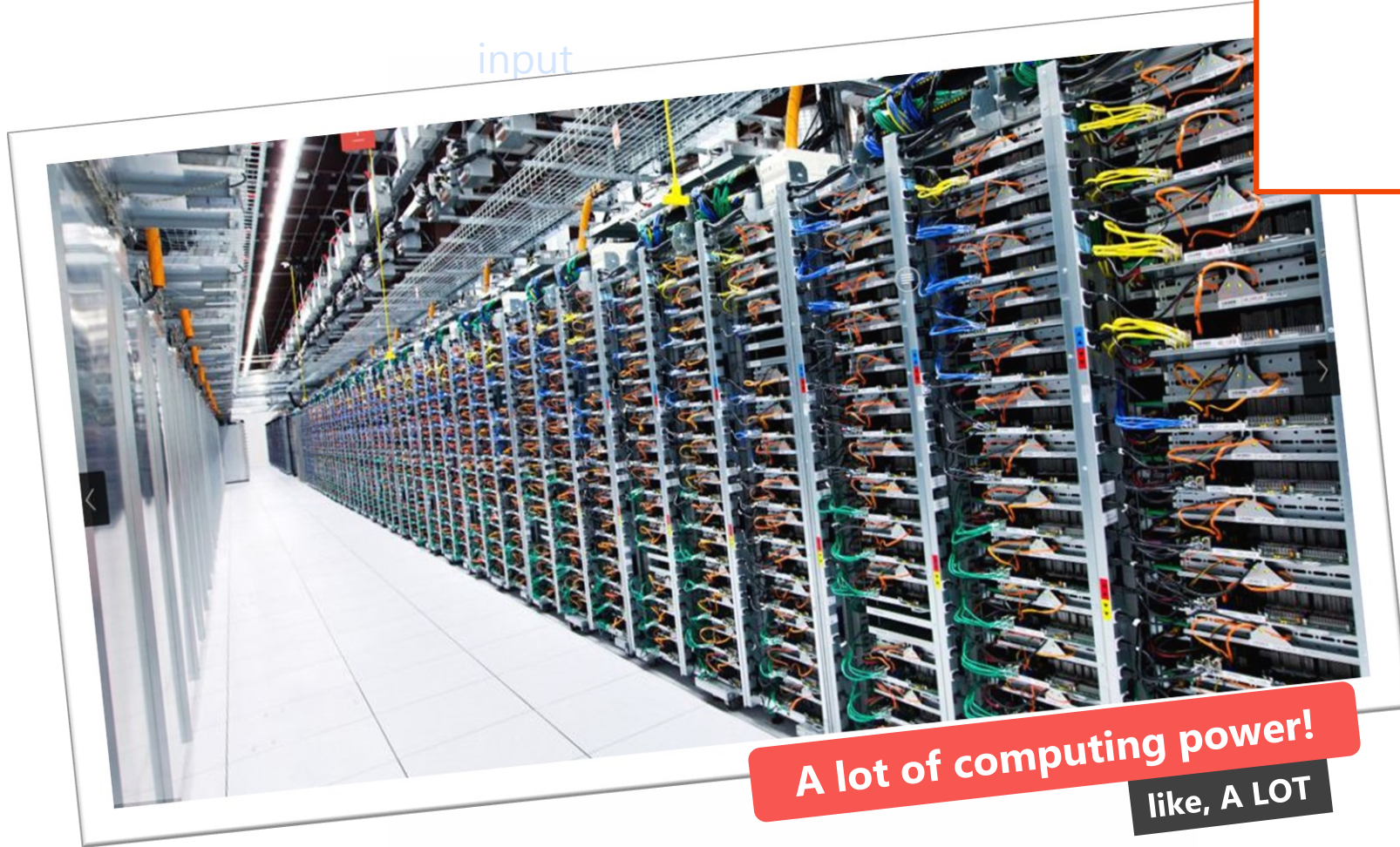
training GPT 3:

1024 GPUs

34 days

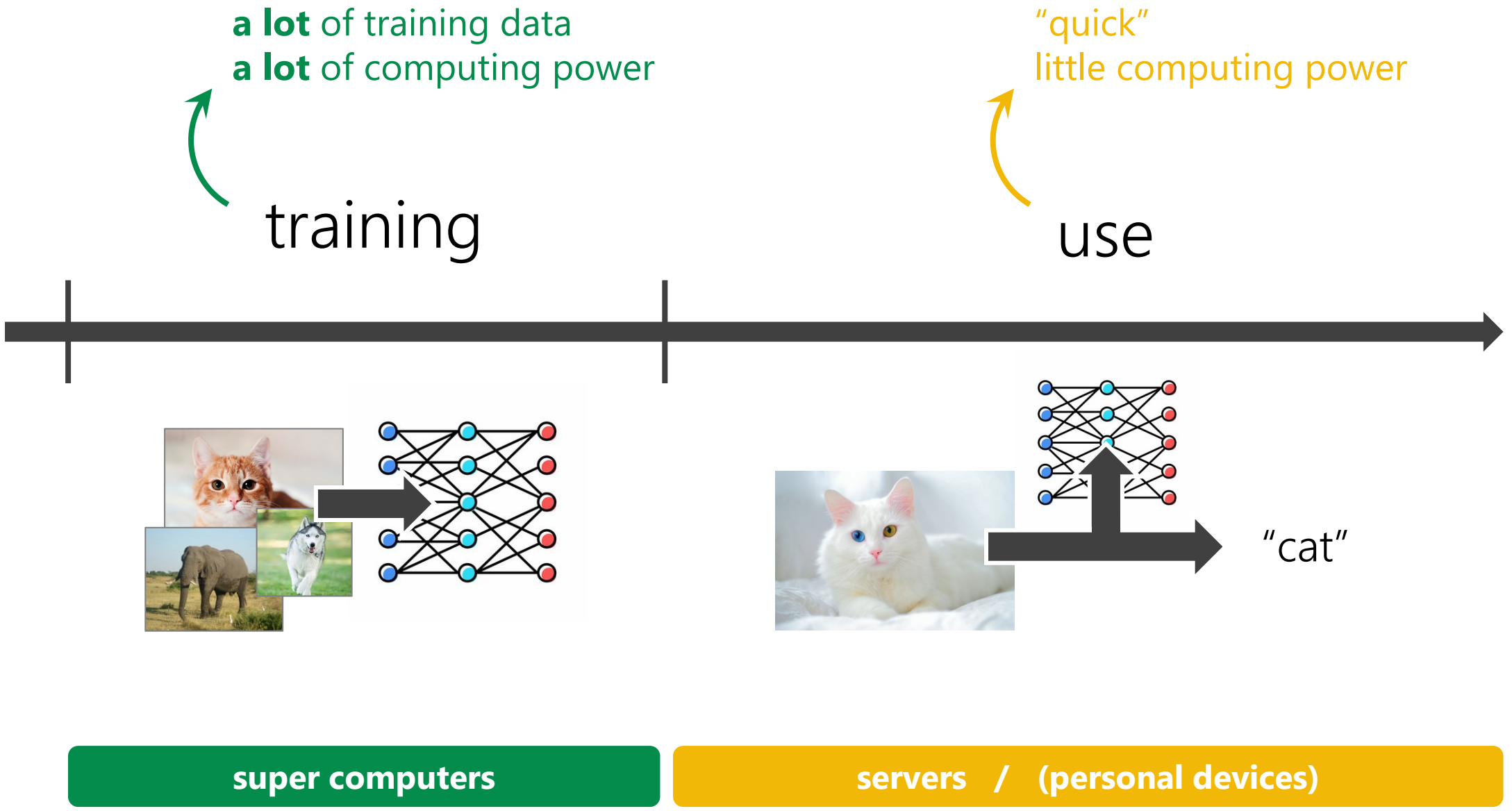
\$4.6M

input



A lot of computing power!

like, A LOT



privacy challenges



inherent to the technology

#1

Rights on training data

AI requires **masses** of data

ChatGPT 3:
300.000.000.000 words (\cong 3.000.000 books)

mostly protected by
intellectual property and **privacy** rights

→ How to ask for everyone's
consent / permission?

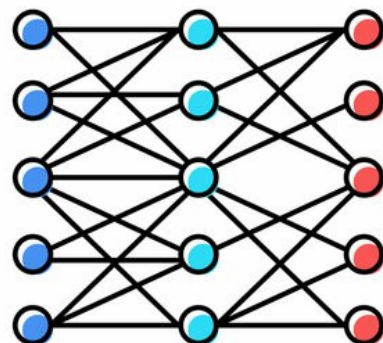
... but wait, how did OpenAI do?



#2

Rectify infringements

learning
is easy:



but how to
forget?

ChatGPT 3 → 1.000.000.000.000 nodes

hard to know
which links and nodes
represent specific training data

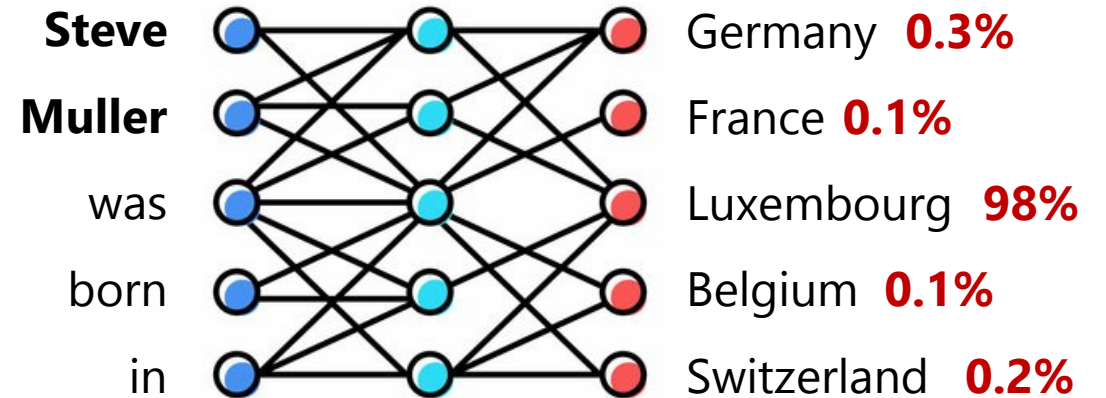
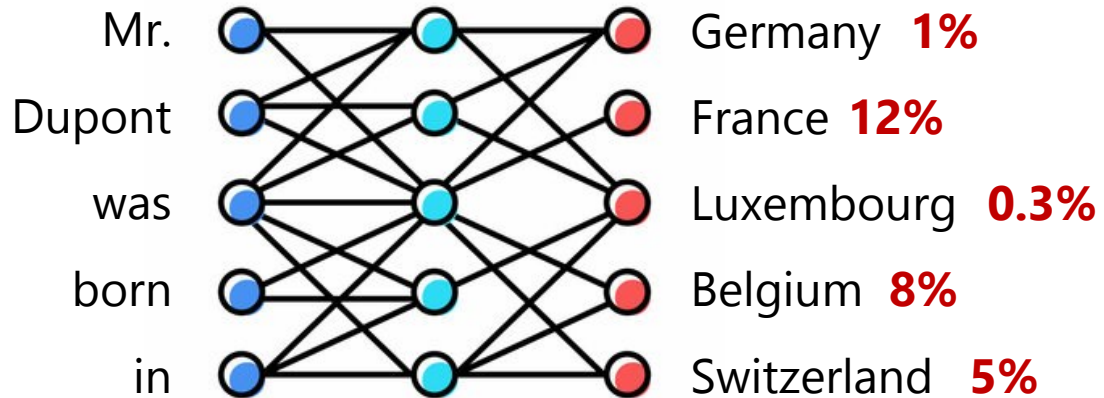
cannot "just" remove
personal data or
copyright material

#3

Learning data can leak

example: membership inference

LLM gives most likely next word,
based on training data



privacy challenges

when AI is applied

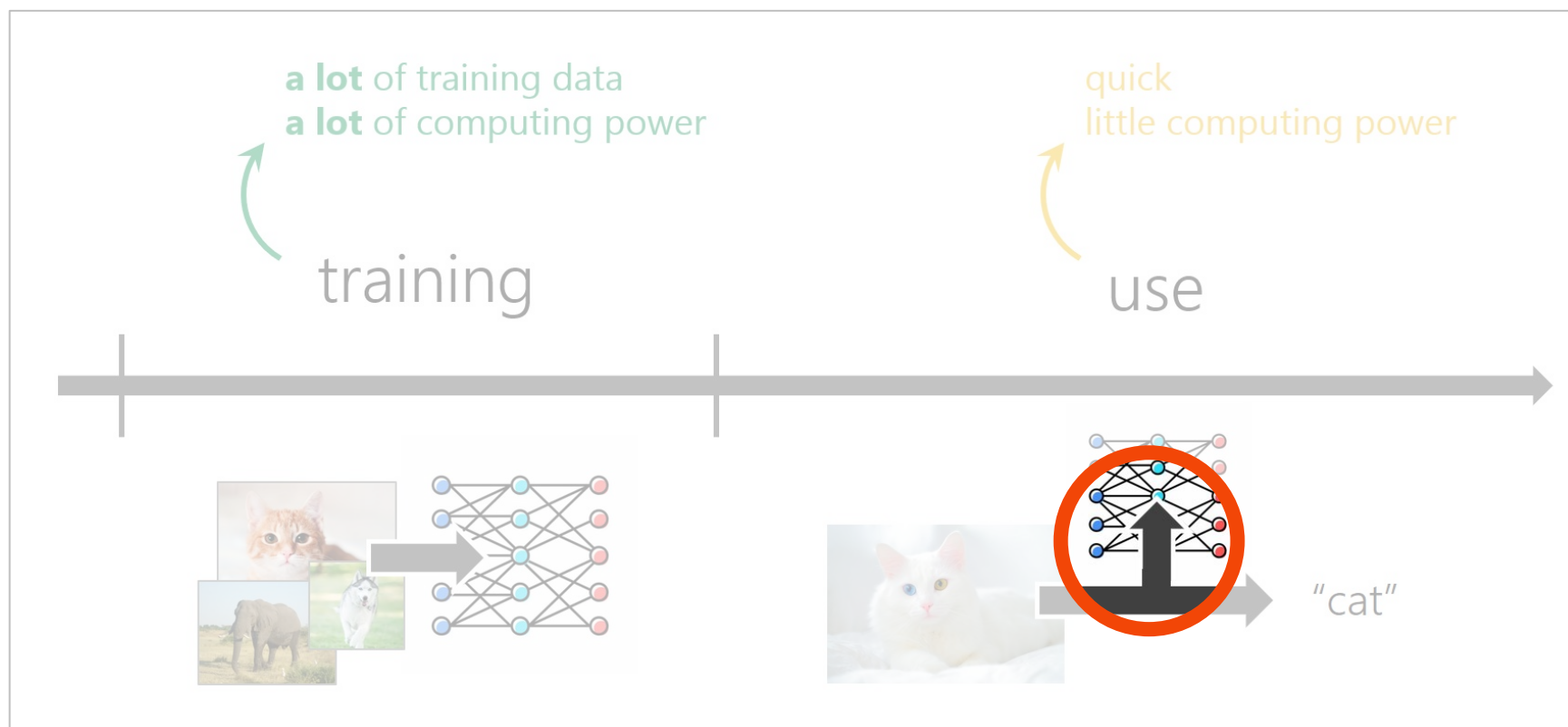


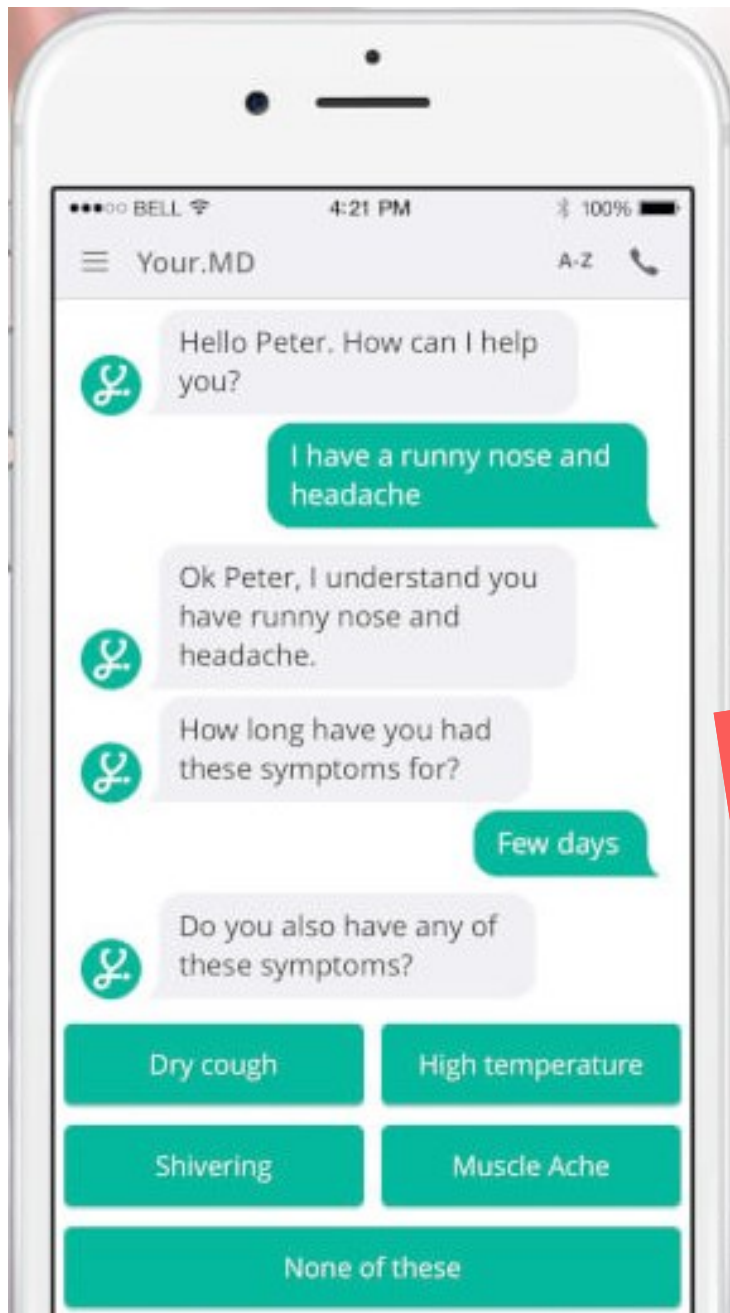
#4

ChatGPT

When you use our non-API consumer services ChatGPT or DALL-E, we may use the data you provide us to improve our models. You can switch off training in ChatGPT settings (under Data Controls) to turn off training for any conversations created while training is disabled or you can submit [this form](#). Once you opt out, new conversations will not be used to train our models.

<https://openai.com/policies/terms-of-use>





ChatGPT

When you use our non-API consumer services ChatGPT or DALL-E, we may use the data you provide us to improve our models. You can switch off training in ChatGPT settings (under Data Controls) to turn off training for any conversations created while training is disabled or you can submit [this form](https://openai.com/policies/terms-of-use). Once you opt out, new conversations will not be used to train our models.

<https://openai.com/policies/terms-of-use>

**your data
might leak
to another user**

a lot of training data
a lot of computing power

quick
little computing power

use

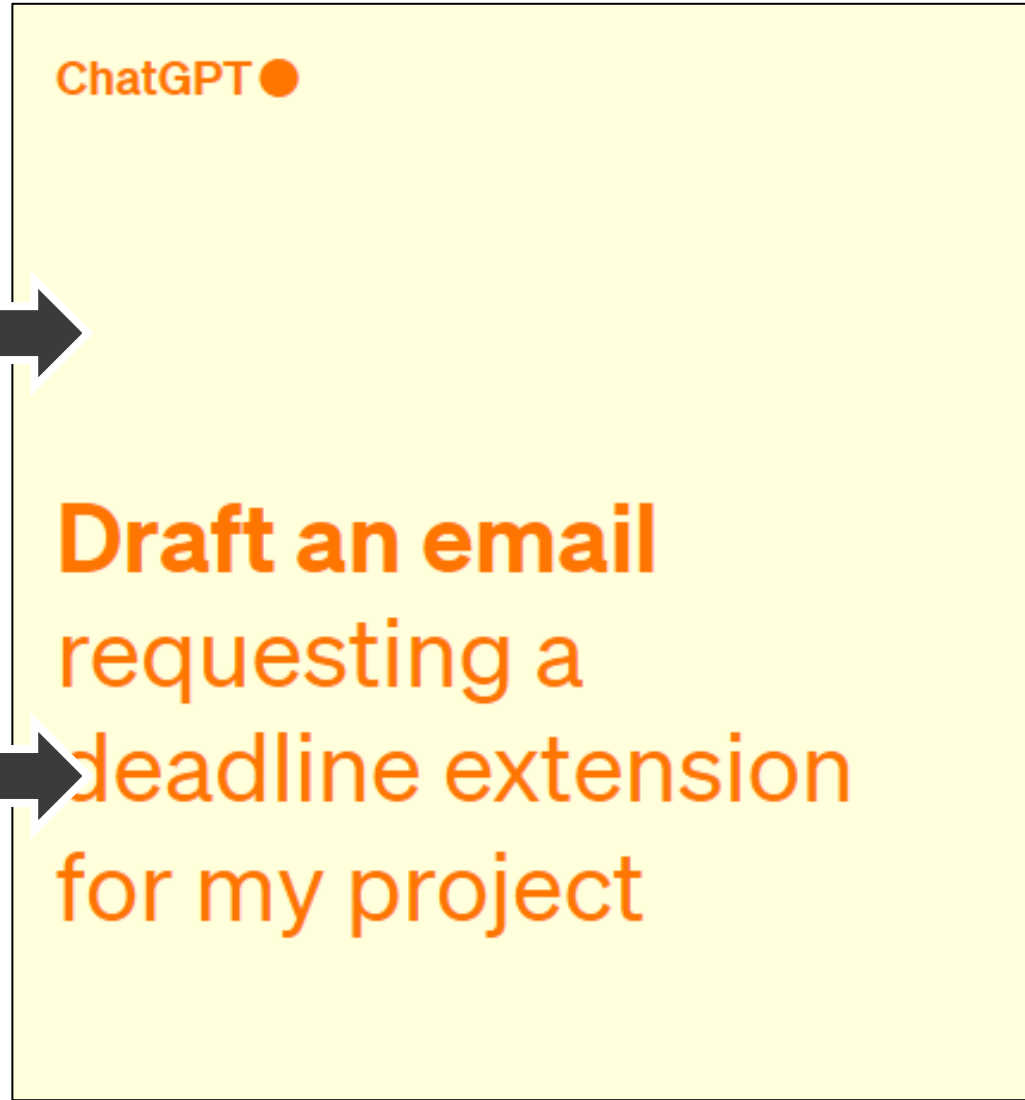




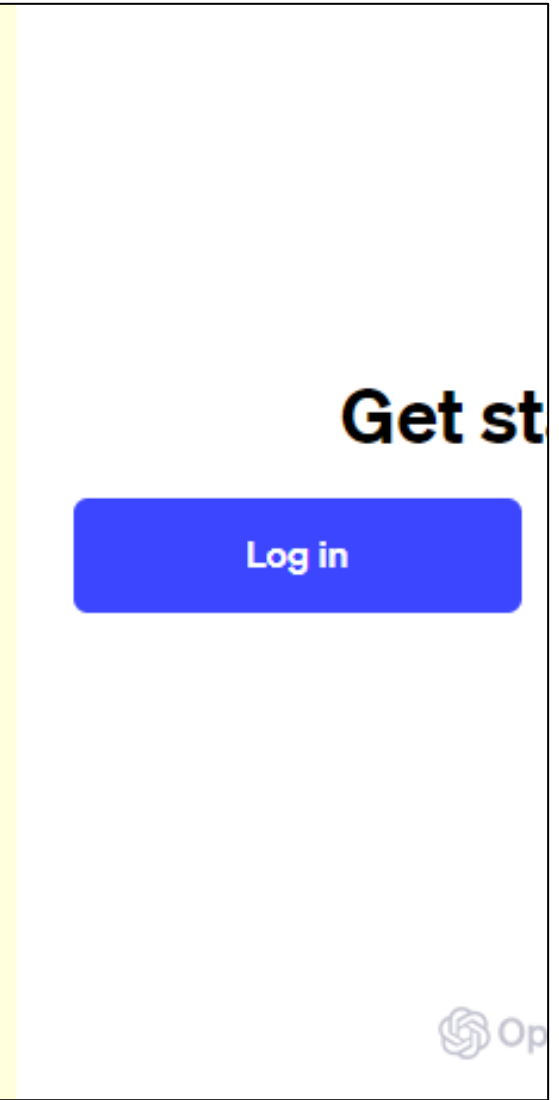
your doctor's
website



health AI
provider



chatbot
provider



#6

Deep fakes

<https://ars.electronica.art/center/en/obama-deep-fake/>

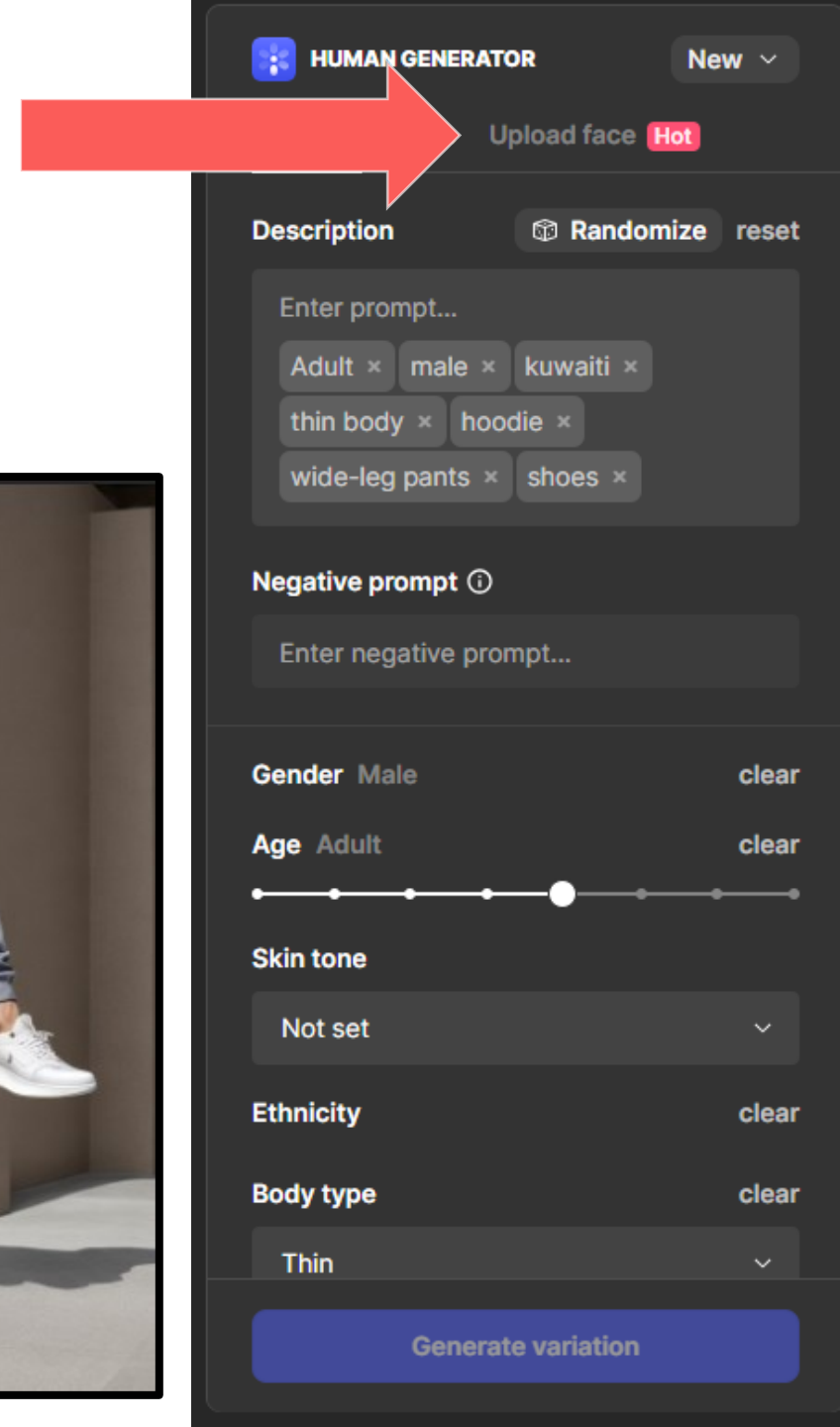


#6

Deep fakes

- reputation damage
- mobbing
- blackmailing
(e.g. deep fake porn)
- faking evidence

other **data leaks**
can suddenly have
higher impacts



#6

Deep fakes

Solutions?

detect AI generated content
with AI

bypass AI content detection **with AI**

trick the bypassing **with AI**

detect tricking of

overcome detection **with**

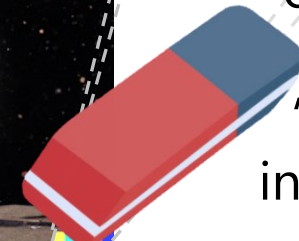
even better detection **with AI**



watermarking

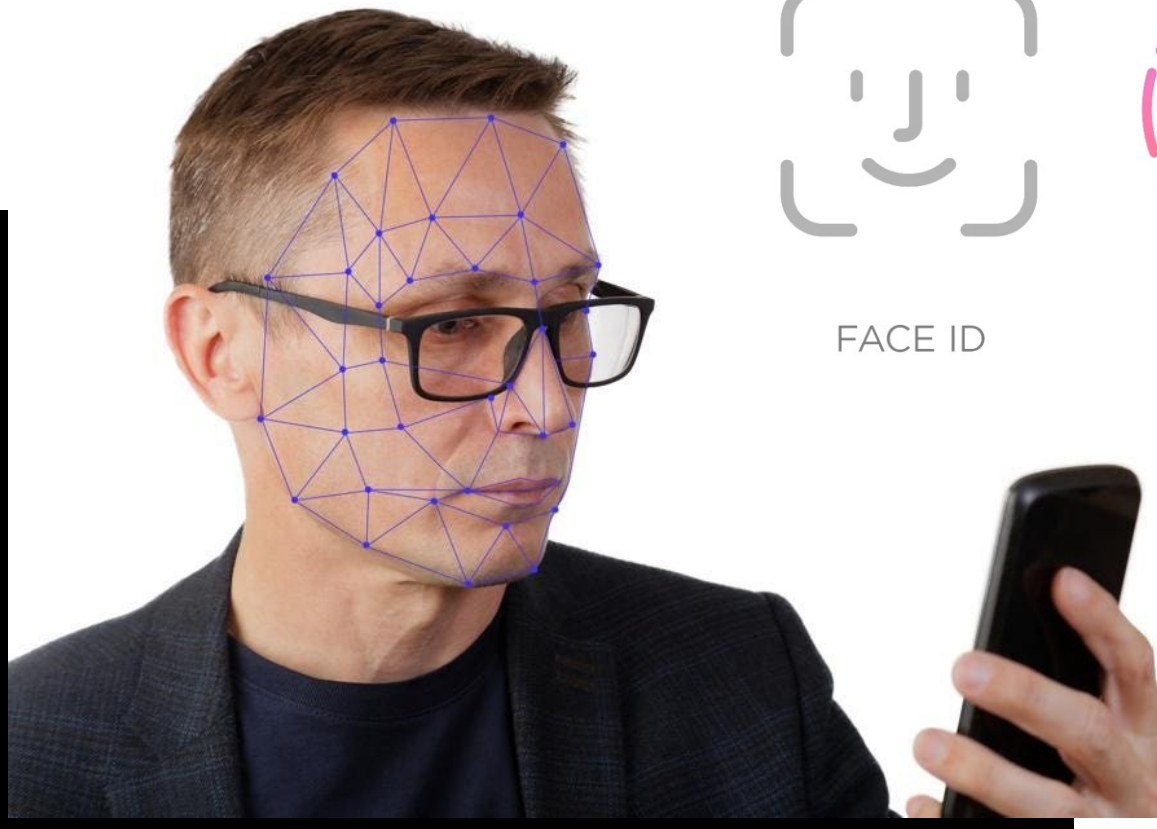
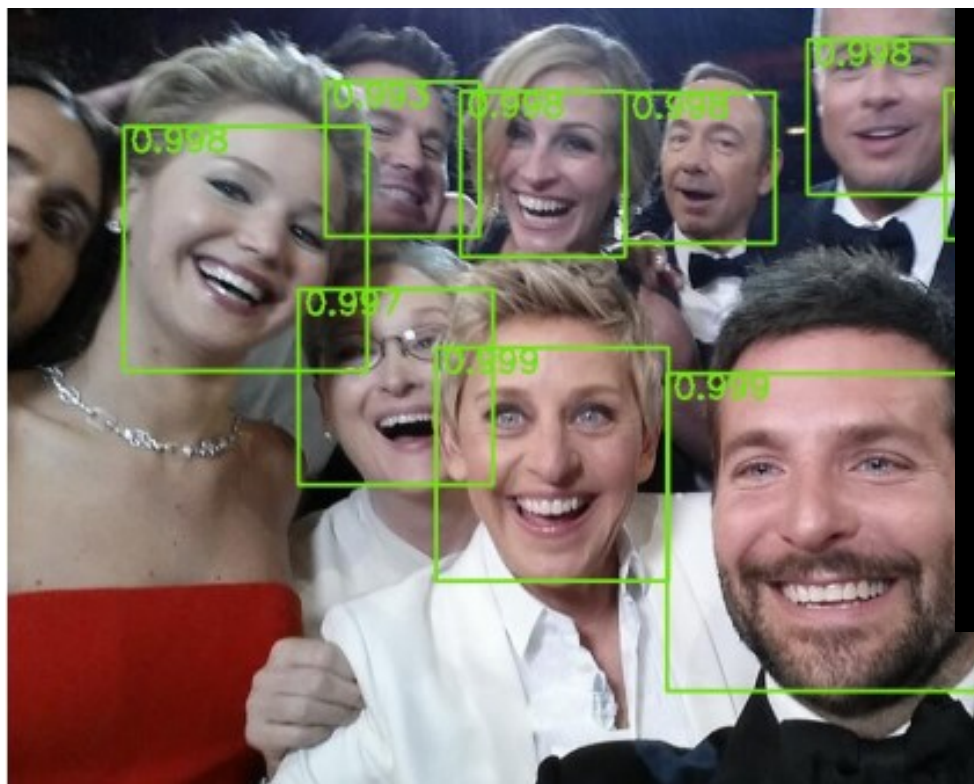


can also be
included
"invisibly"
into a picture



#7

Face detection



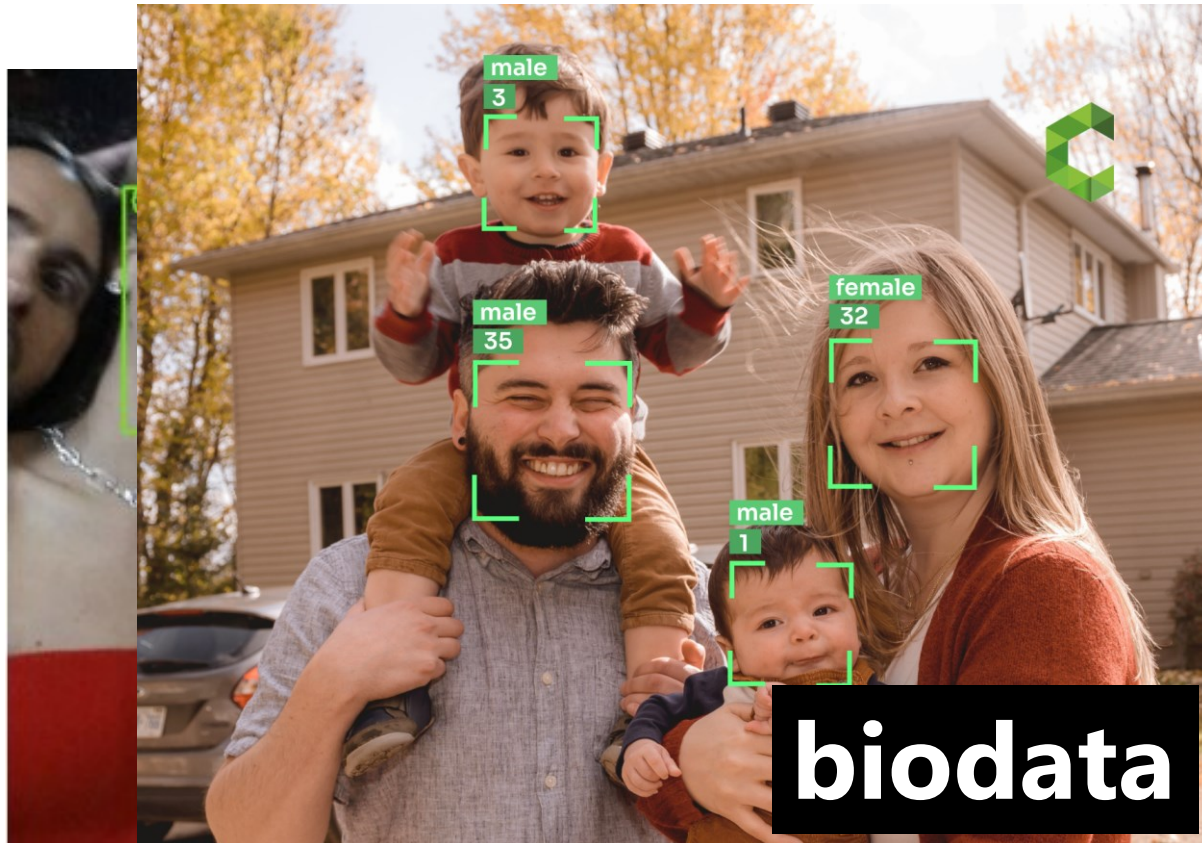
FACE ID



TOUCH ID

#7

Face detection



#7

Face detection




tracking

stalking

surveillance

social
scoring

#7

 Sign in

The Guardian


News

Opinion

Sport

Culture

Lifestyle



Pornography websites will have to check users' ages, under draft guidelines

The list of measures for proving someone is over 18 include: uploading a photo-ID document; facial age estimation technology; contacting your mobile network provider to allow your phone to access adult content; checking age via credit card details; and using “digital identity wallets” that store evidence of a person’s age.

#7

Sign in

The
Guardian

News

Opinion

Sport

Culture

Lifestyle



CNIL

PROTÉGER les données personnelles
ACCOMPAGNER l'innovation
PRÉSERVER les libertés individuelles

I have to check
guidelines

Online age verification: balancing privacy and the protection of minors

22 September 2022

is over 18 include: uploading a
technology; contacting your
none to access adult content;
using “digital identity wallets”



Steve Muller
steve.muller@eco.etat.lu