Intelligent systems for industry, supply chain and environment

LESSON 2

Artificial and human learning, Gestalt, applications and opinions about AI

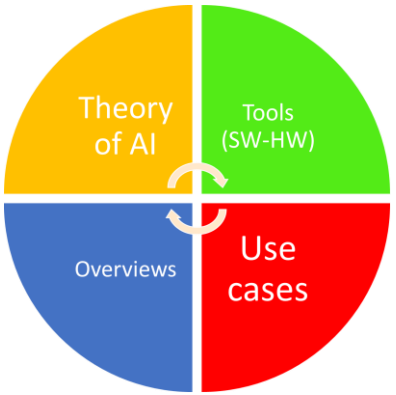
Lesson outline

- Theory
  - “Just” a clever non linear function
  - Input space and internal space
  - Gestalt and current limitation of machine learning
  - Computational intelligence models
  - Explainable Artificial Intelligence
- Overview of AI and ML methods
- Relevant opinions about AI
- The 4 main drivers of AI
- **Main points**

THEORY

What is trained neural network?

Just a non linear function

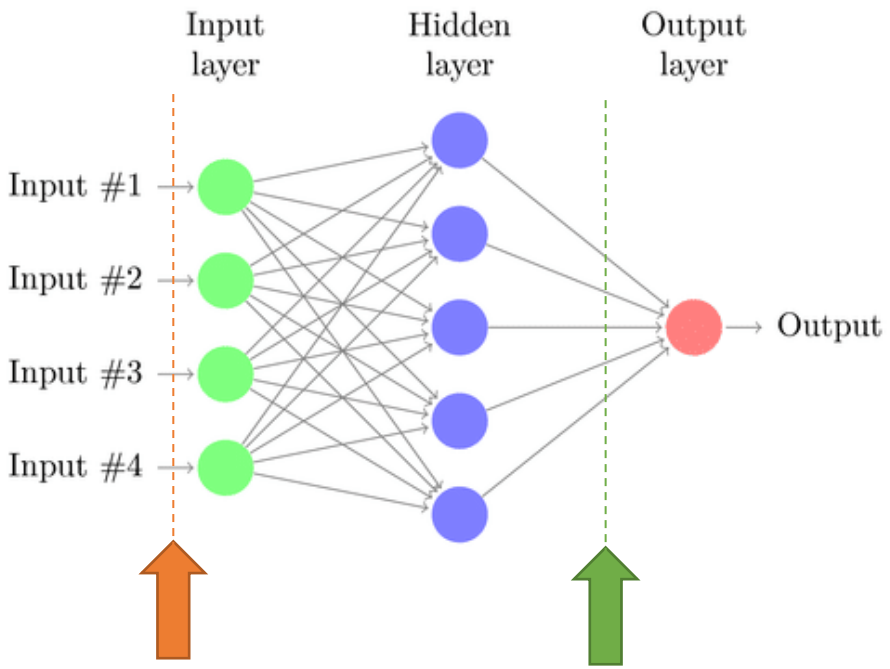


Not magic at all! Just functions

Output:
a single value as
an integer or float,
or a vector/matrix

- A trained model of Machine Learning is just a (non-linear) function
- X = input vector
- Out = output vector

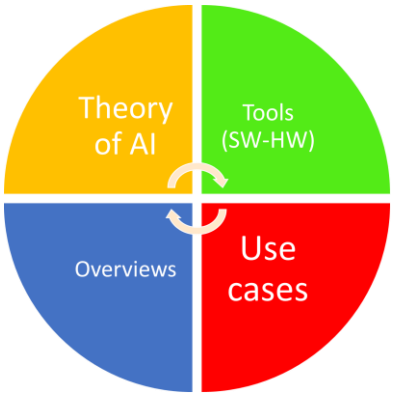
Out=F(X) ≡



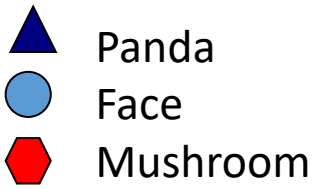
THEORY

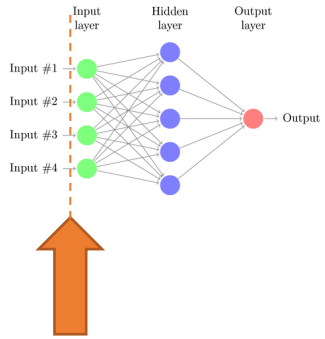
Feature space and decisions

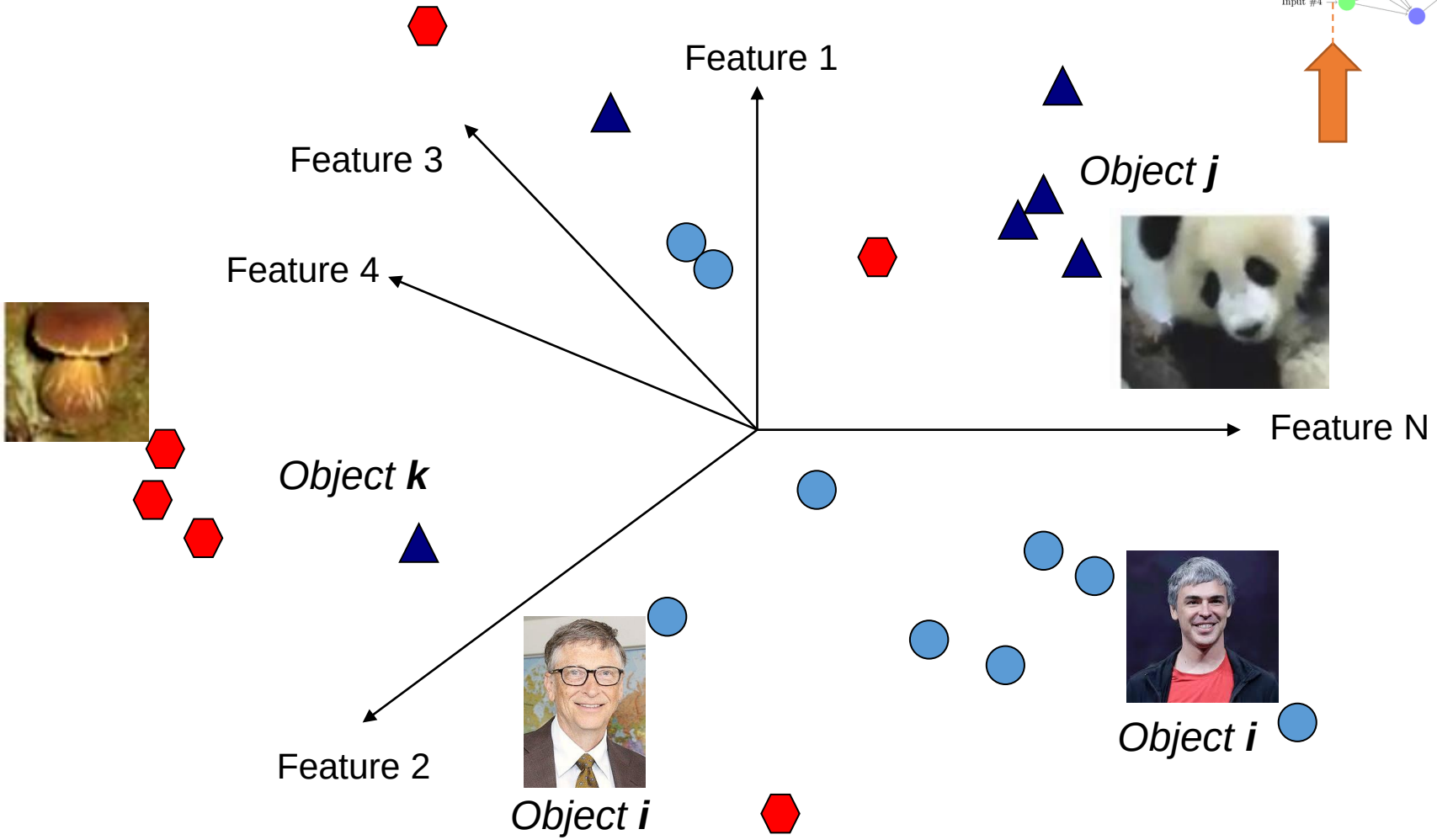
The space where the network
is not rather “thinking” but “measuring”



Input space

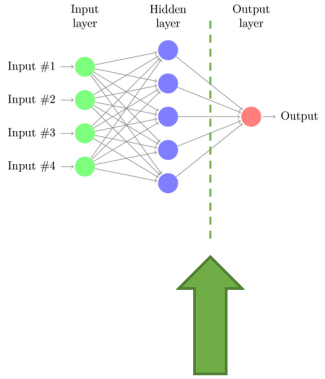

Panda
Face
Mushroom


Input layer
Hidden layer
Output layer
Input #1
Input #2
Input #3
Input #4
Output


Feature 1
Feature 3
Feature 4
Feature N
Feature 2
Object j
Object k
Object i
Object i

Internal feature space

Panda
Face
Mushroom



Feature 1

Feature 3

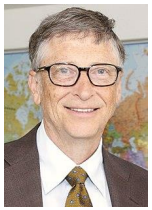
Feature 4

*Object **j***

Feature M

*Object **k***

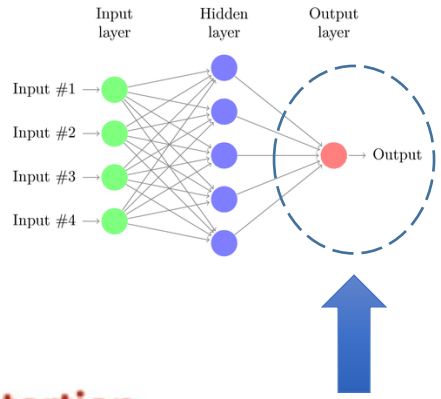
*Object **i***



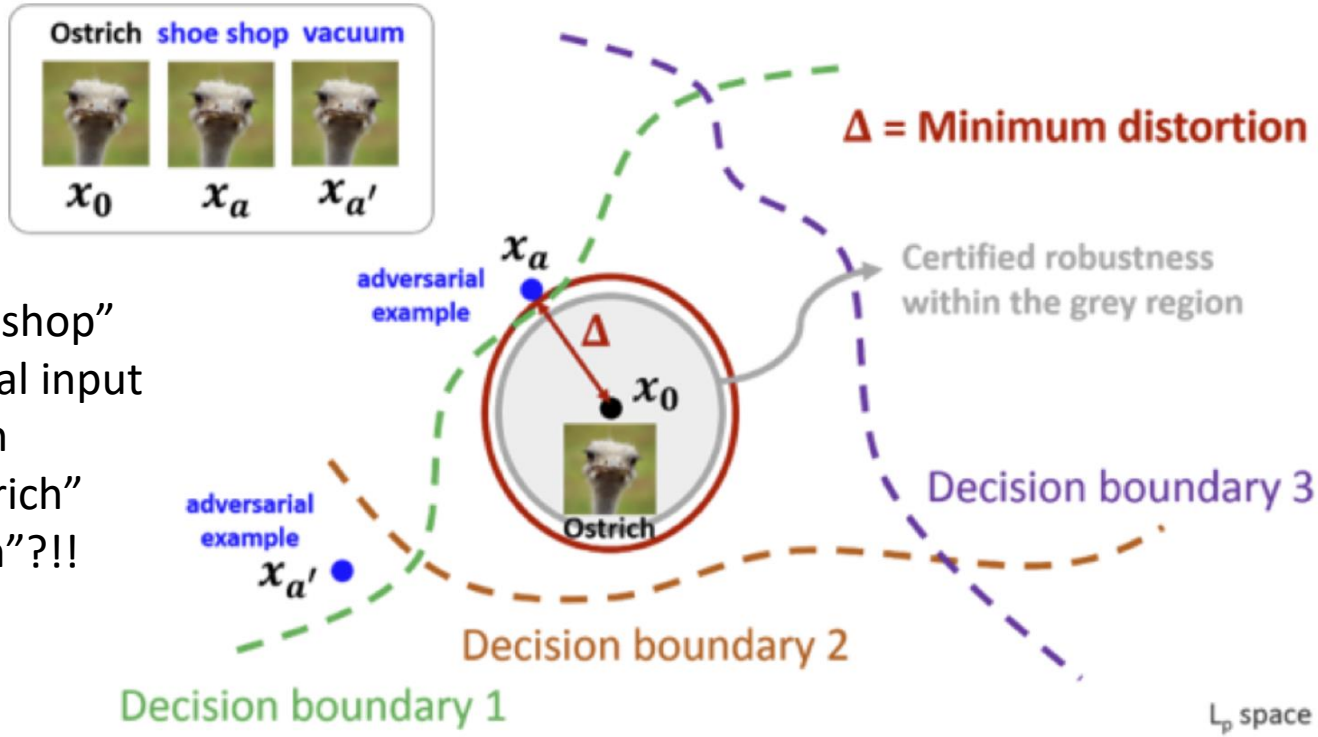
Feature 2

Decision boundaries in the feature space of the neural network

Refer to a NN implementing a Visual Object Detector



Non linear function

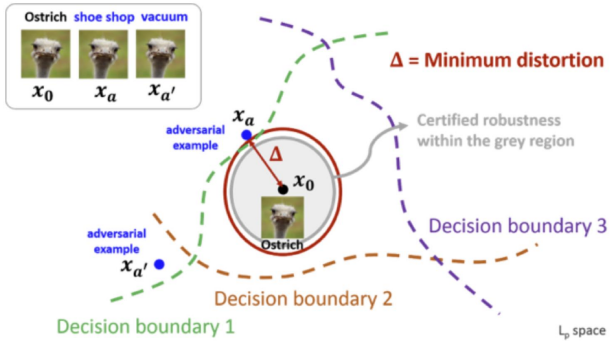


Why a “Shoe shop” with a minimal input distortion can became “Ostrich” or e “Vacuum”?!!

The “chihuahua or muffin” example for deep NN

Chihuahua or muffin?

…but the concept of similarity for the human brain is not equivalent to metrics in deep learning networks.

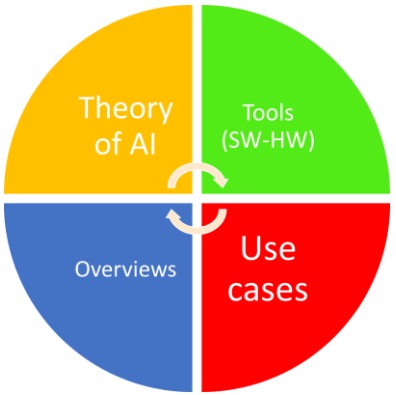


Mariya Yao, Metamaven

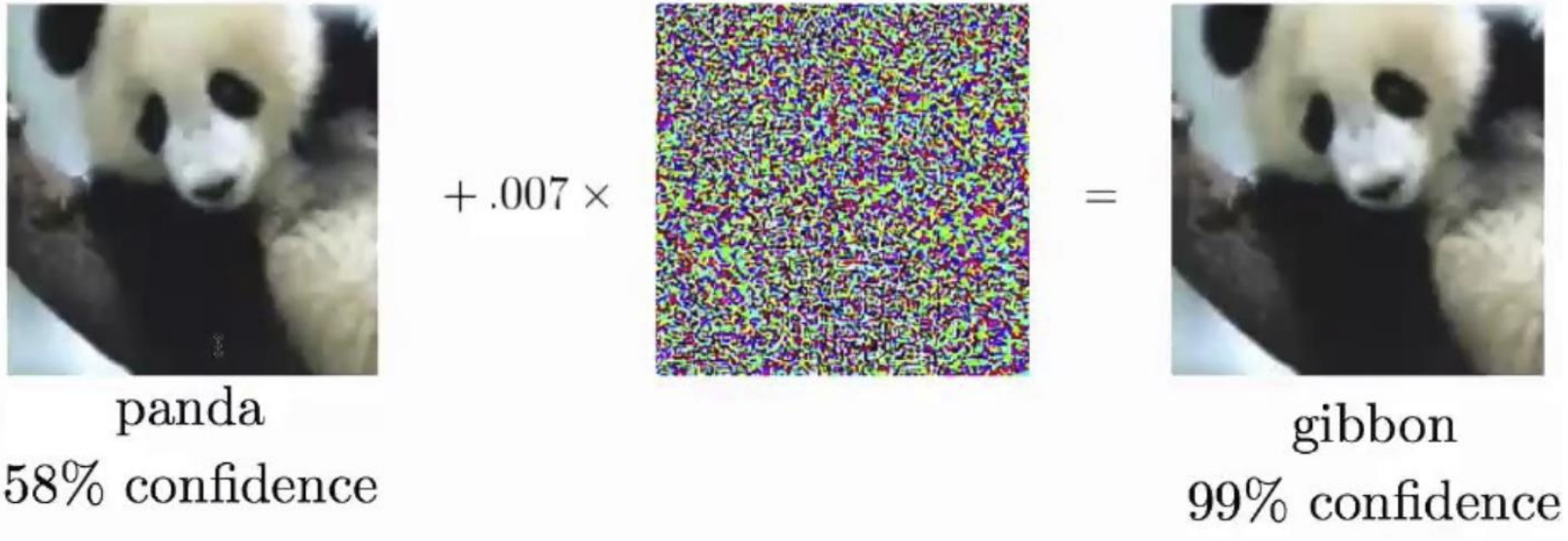
THEORY

Problems: Adversarial attacks

Deep neural models can also be attacked!



Small distortions or noise can produce errors (..adversarial)



Worst case inputs:
choosing the best inputs to be perturbed
by additive designed patterns

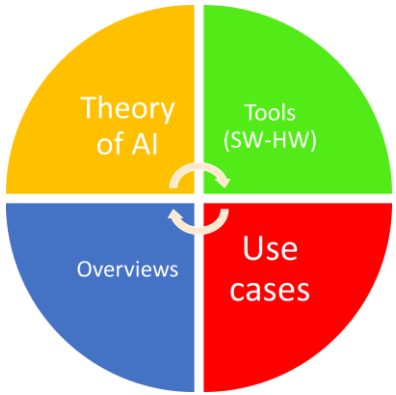
Not all regions of input are conducive to adversarial examples

Adversarial Approaches to Bayesian Learning and Bayesian Approaches to Adversarial Robustness, Ian Goodfellow, NIPS 2016 Workshop on Bayesian Deep Learning, 2016

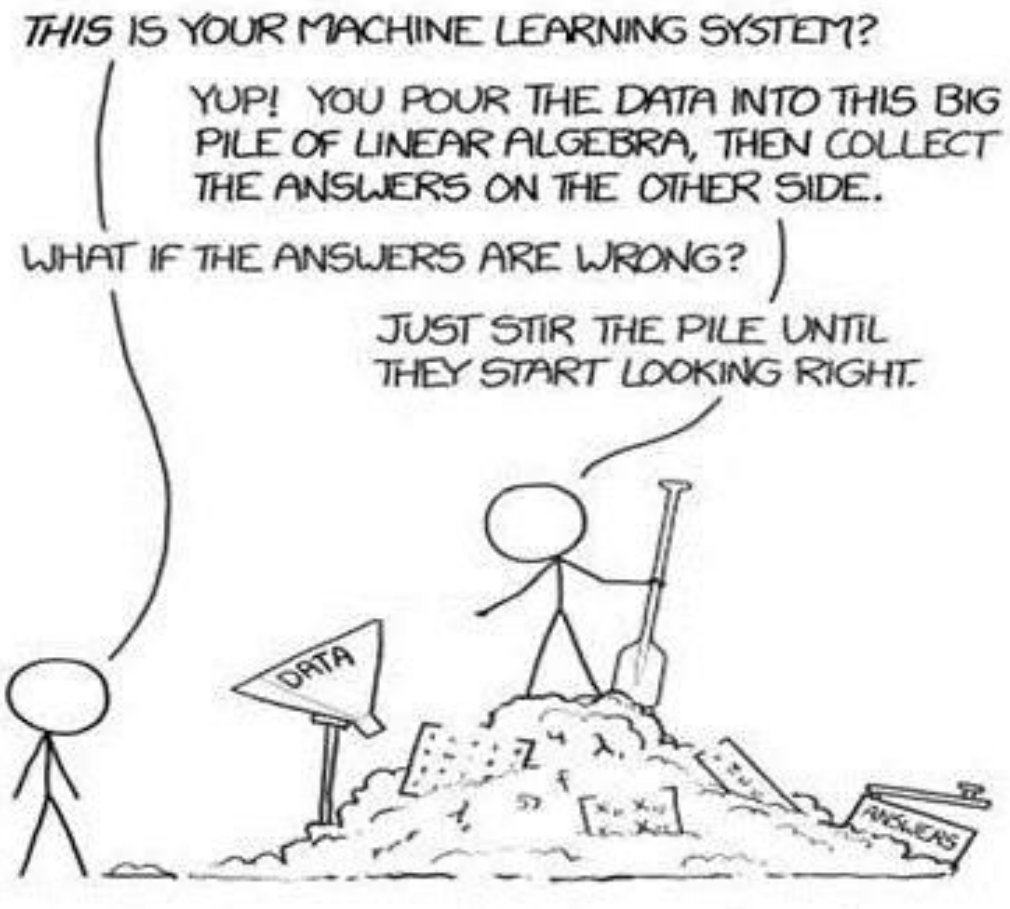
THEORY

Problem:
Minimal Explainability
Explainable Artificial Intelligence

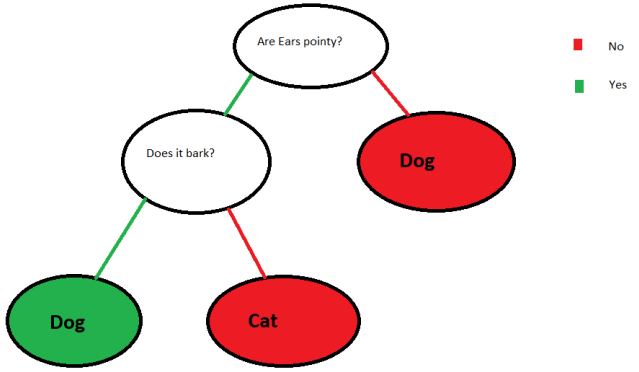
Neural models are like “black boxes”



Humor
or true?

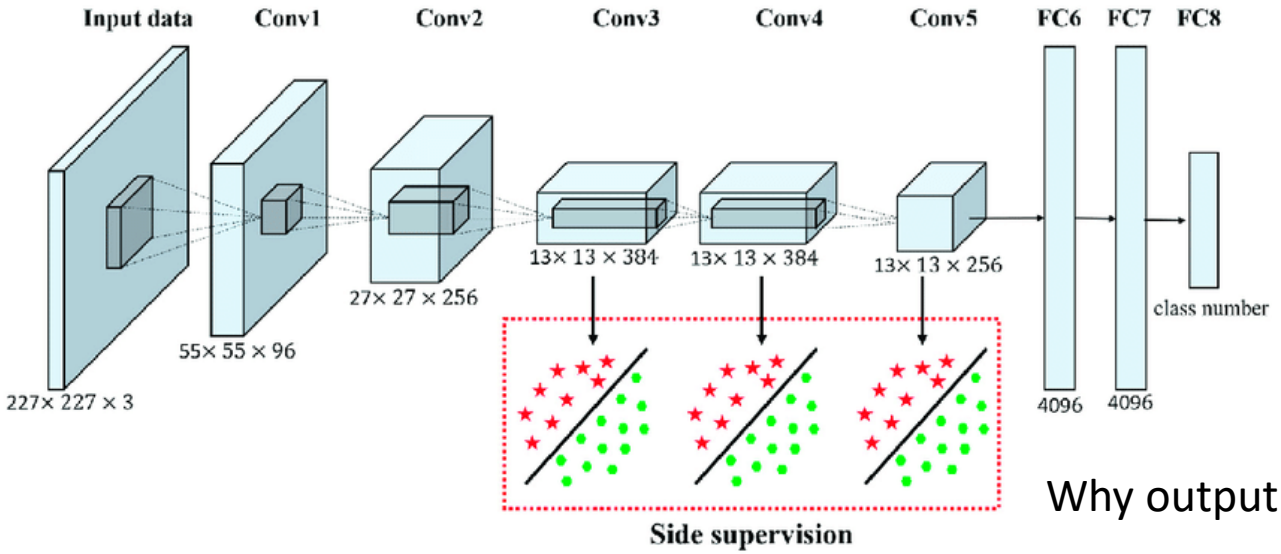


Explainability lack of large NNs



Neural networks
are typically accurate but
un-interpretable.

Decision tree: less accurate but the decision can be explained

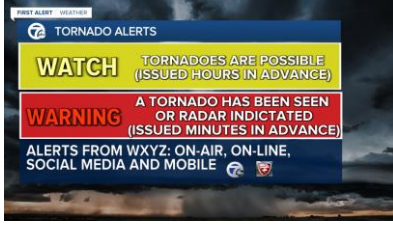


Why output is «cat»?!!

Explainable Artificial Intelligence (XAI)

- XAI is a set of processes and methods that allows human users to
  - comprehend and trust the results and output created by machine learning algorithms.
  - describe the AI model, its expected impact and potential biases

- XAI helps characterize in AI-powered decision making
  - Model accuracy
  - Fairness
  - Transparency
  - Outcomes



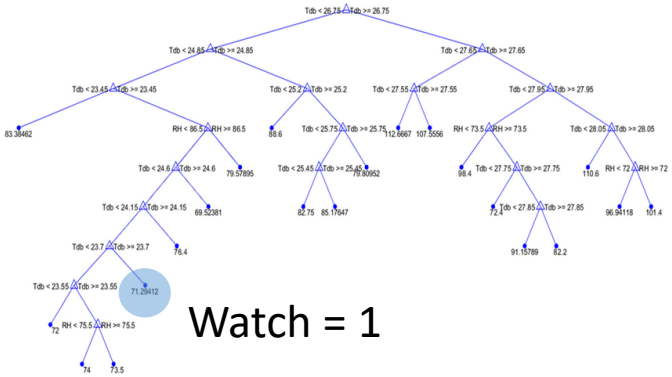
AI vs XAI example

Tornado watch detector

- Classical Decision Tree

  Output: Watch = 1

Processing features



Watch = 1

- XAI Output:

"A tornado watch was predicted in Fairfax county. This has a confidence of 75% of occurring with a minimum and maximum duration of 1-2 hours. This is based on historical data with similar characteristics such as historical frequency, temperature, wind speed, and barometer readings."

Interpretability in AI

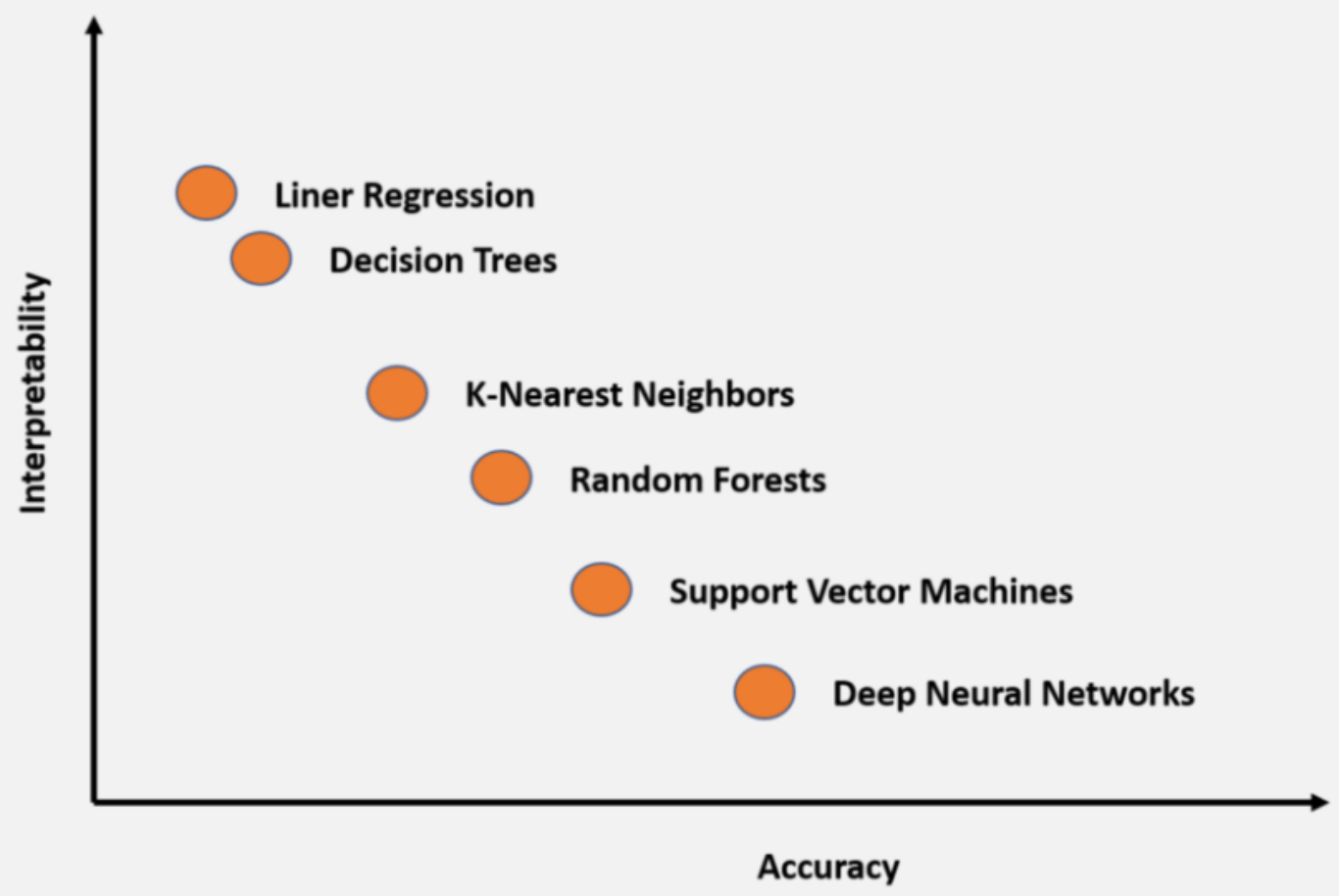
- **Interpretability**
  is the degree to which an observer can understand the cause of a decision.
  For example, the success rate that humans can predict for the result of an AI output

Interpretability-Accuracy Tradeoff

Interpretability

→

Understanding
the cause of a decision

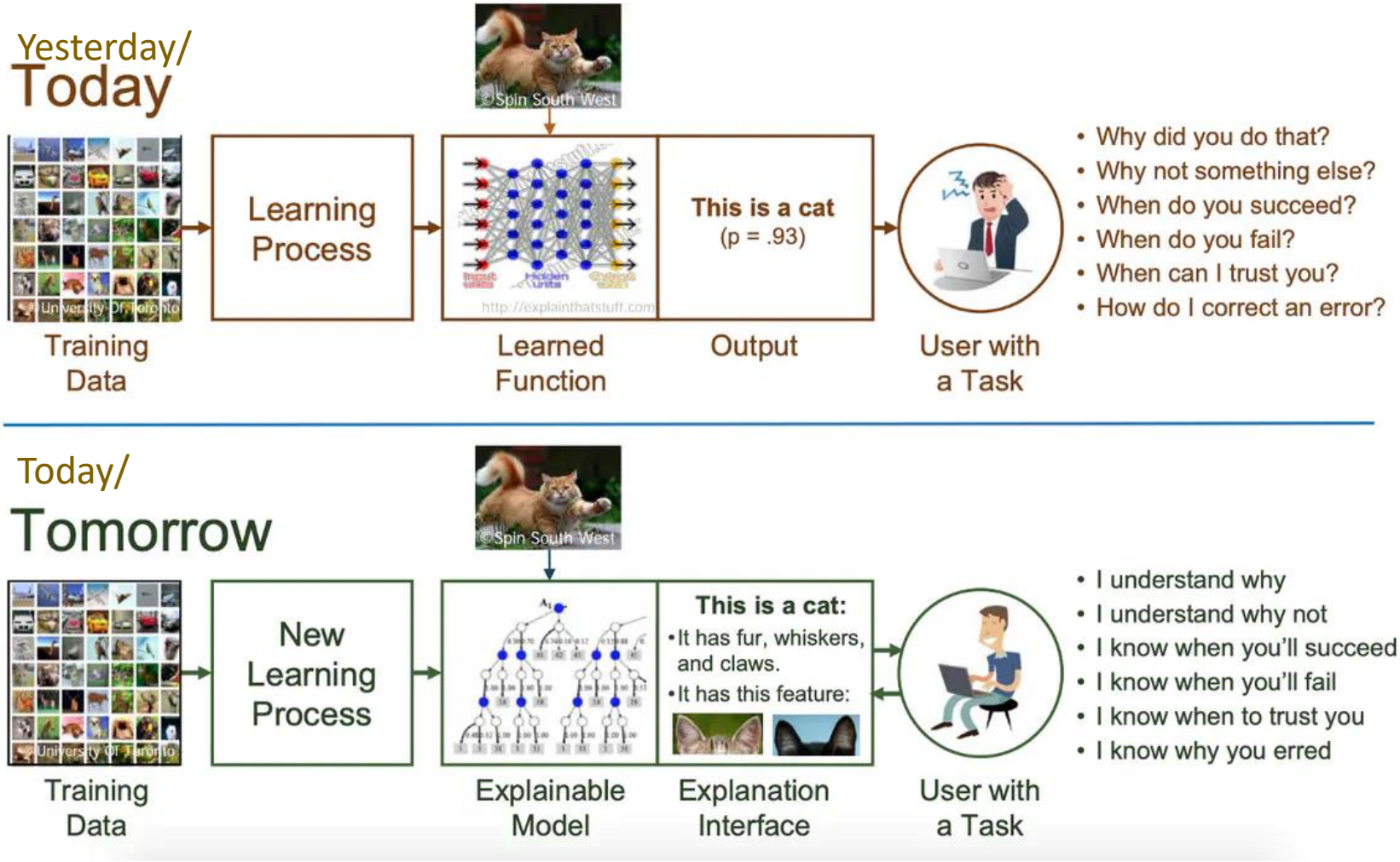


Explainability vs Interpretability in AI

- **Interpretability**
  is the degree to which an observer can understand the cause of a decision. It is the success rate that humans can predict for the result of an AI output.
  It's about Input/Output prediction.

- **Explainability**
  goes a step further and looks at how the AI arrived at the result.
  Understanding the model inside the "box"

Explainable models

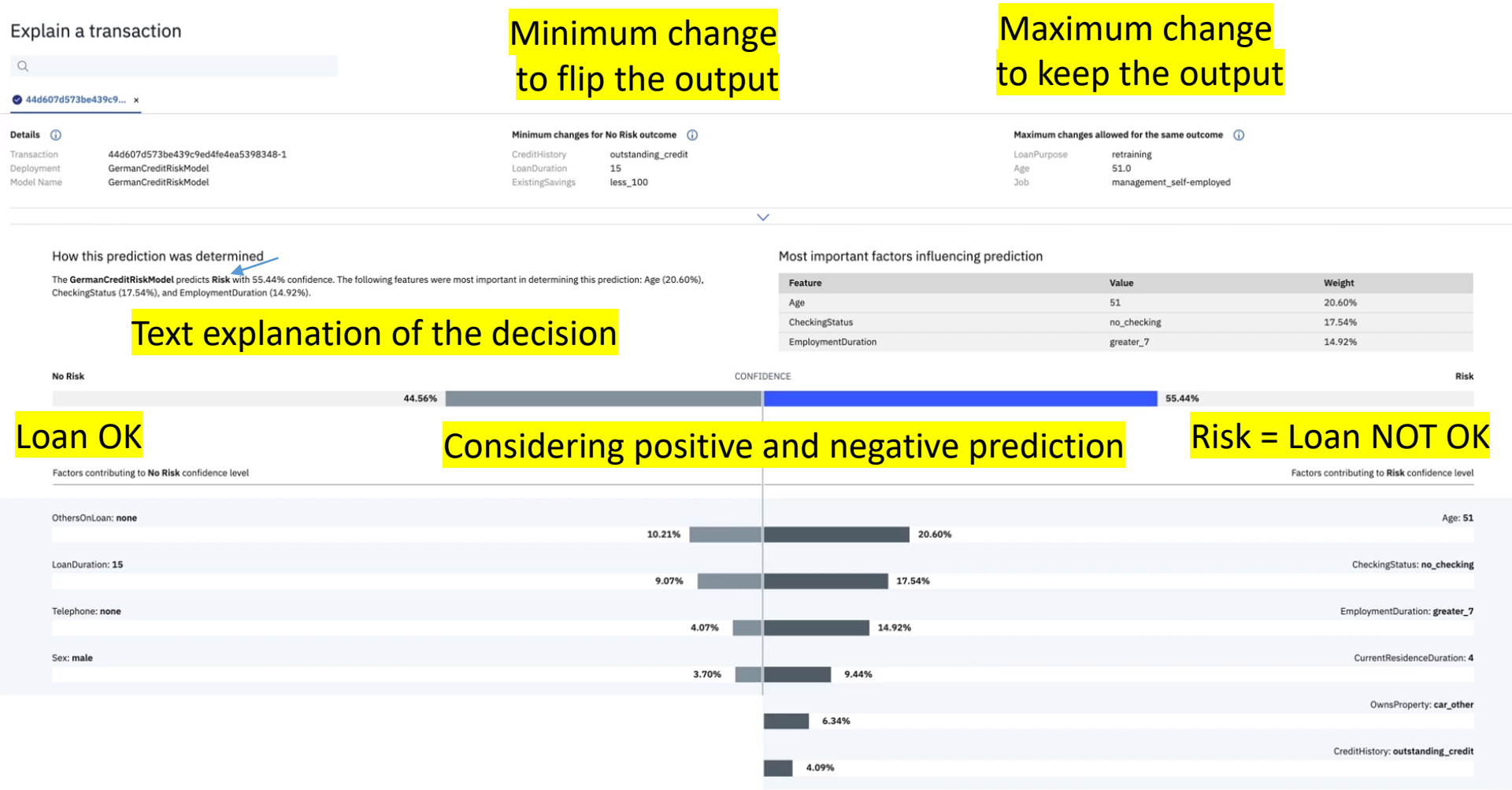


Slide from Yi Mei

Desirable outcomes with XAI

- Fairness and debiasing
  - Manage and monitor fairness
  - Scan your deployment for potential biases
- Model drift mitigation
  - Analyze your model and make recommendations based on the most logical outcome
  - Alert when models deviate from the intended outcomes
- Model risk management
  - Quantify and mitigate model risk.
  - Get alerted when a model performs inadequately.
  - Understand what happened when deviations persist

XAI: Loan approval example

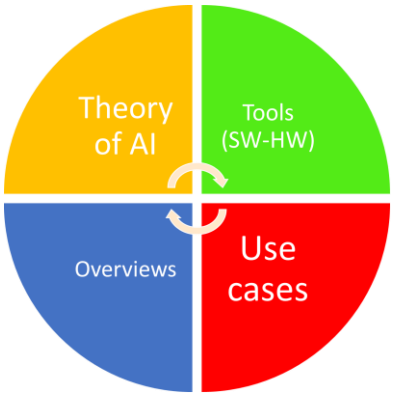


https://www.ibm.com/it-it/watson/explainable-ai

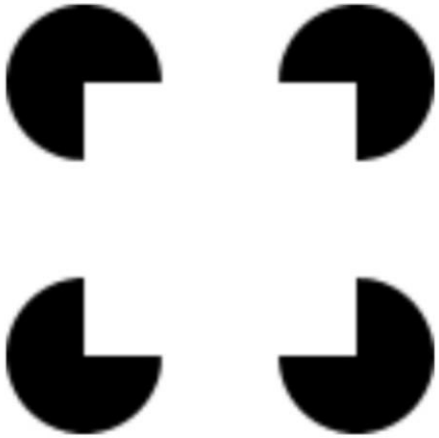
THEORY

The human Gestalt

Why we are so far to the human brain…



Human brain: Gestalt



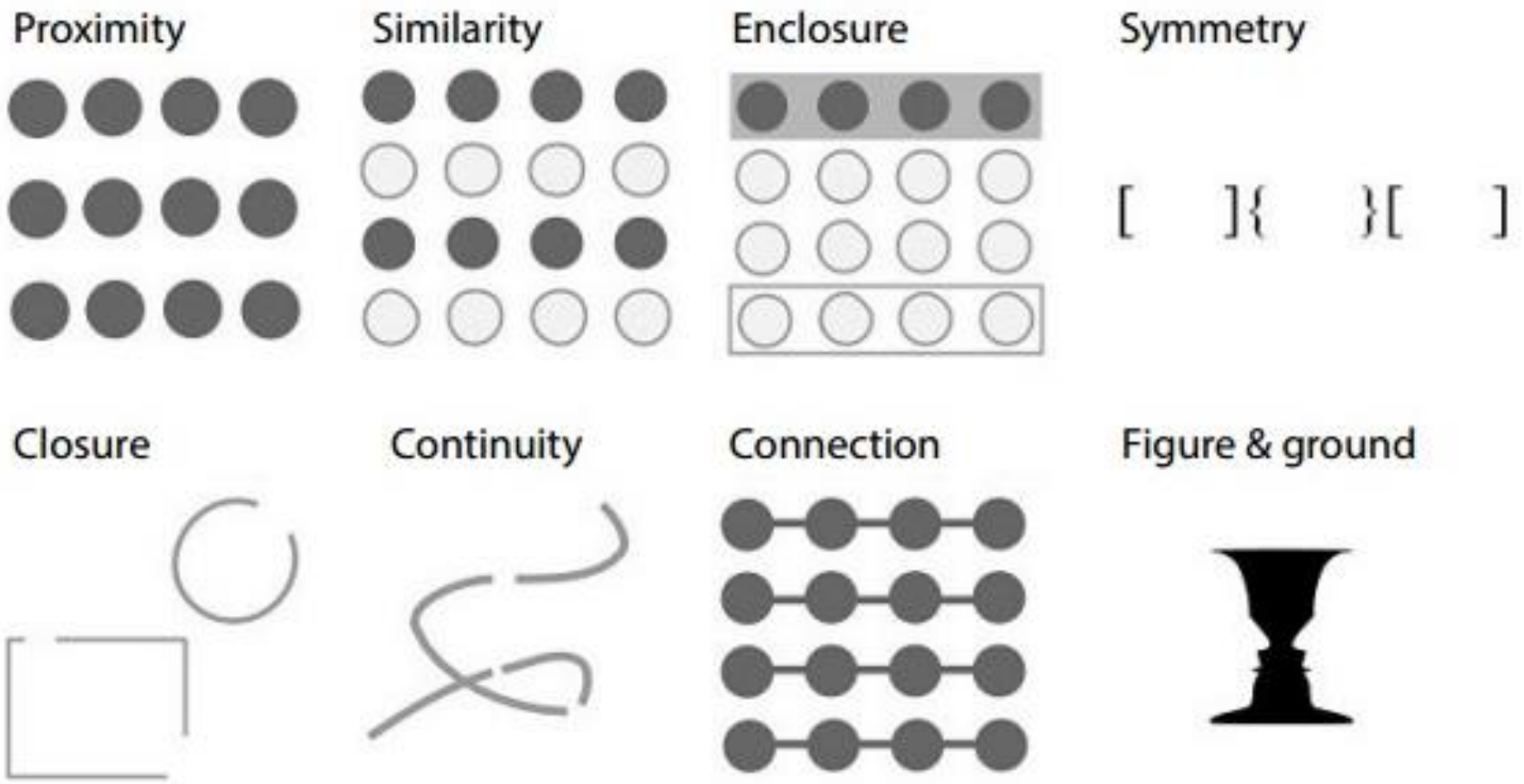
- “Gestalt” (*): from a German word that roughly means *pattern* or *form*.
- “Whole is greater than the sum of its parts”
- “Learning ***is more than just invoking mechanical responses from learners***”.
- The experiences and perceptions of learners have a significant impact on the way that they learn.
- This feature of the human brain is quite difficult to be mimicked from current AI models

(*) from Graf Christian von Ehrenfels, Austrian philosopher (1859-1932)

Gestalt examples: foreground and background

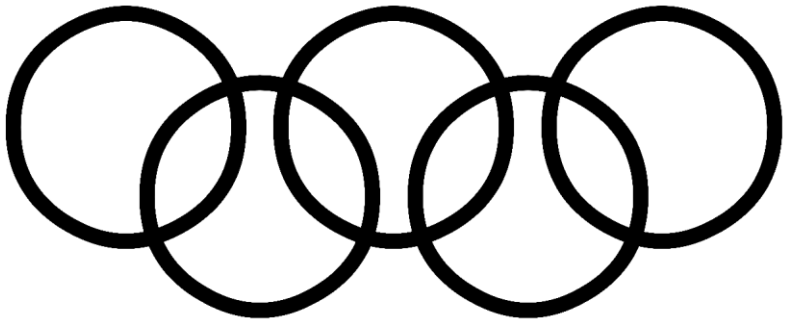
A part of the image can be «just noise» for one pattern
or «the most salient feature» for the other

Gestalt: laws of perceptual organization (examples)



Gestalt: Law of Pragnanz ("good figure")

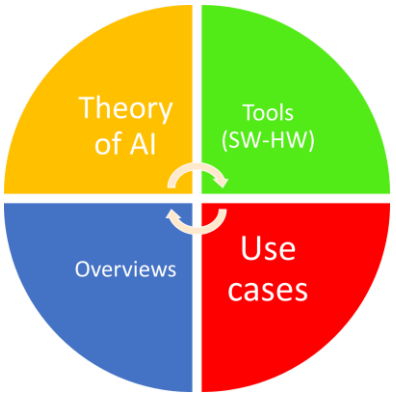
- That objects in the environment are seen in a way that **makes them appear as simple as possible**.
- You see the image below as a series of overlapping circles rather than an assortment of curved, connected lines.



THEORY

What is an intelligent system?

Structure, Properties, Examples,
Pattern matching



Artificial Intelligent Systems

- SW programs or SW/HW systems designed to perform **complex tasks** employing strategies that mimic some **aspect of human thought**

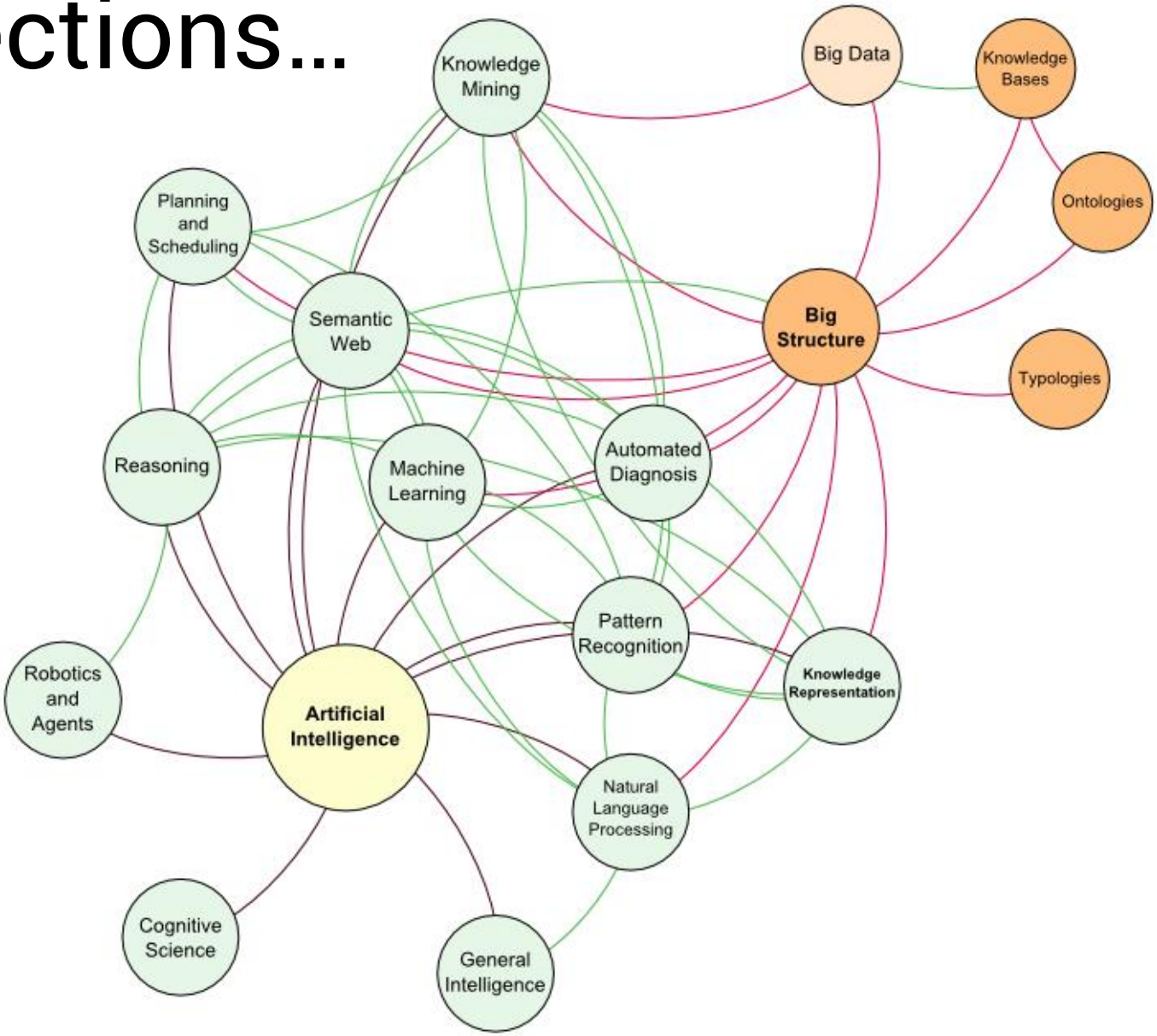
- One can debate endlessly about whether a certain system is intelligent or not

- The key is **evolution**: it is intelligent if it can **learn** (even if only a limited sense) and/or get better in time

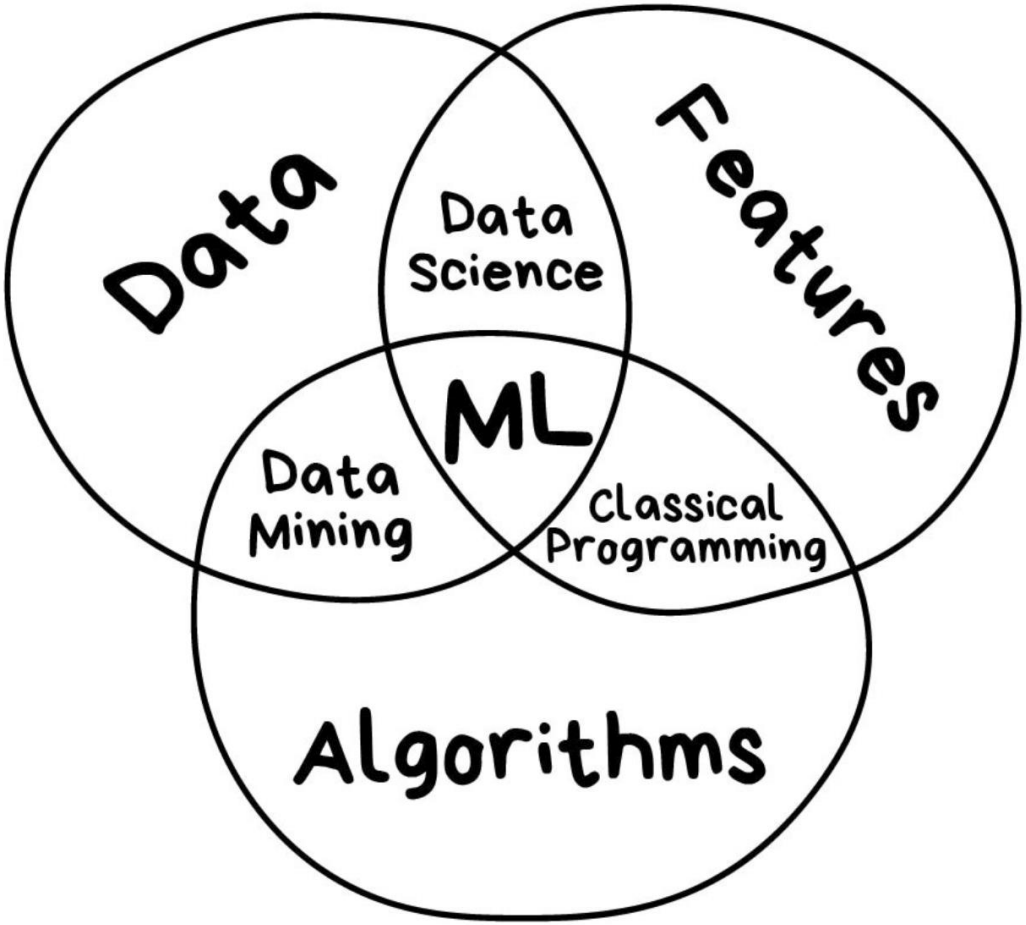
Artificial Intelligence (AI) vs Machine learning (ML)

- AI
  is the broader concept of machines being able to carry out tasks in a way that we would consider “smart”
- ML
  a current application of AI based around the idea that we should really just be able to give machines access to data and let them learn for themselves

AI connections…



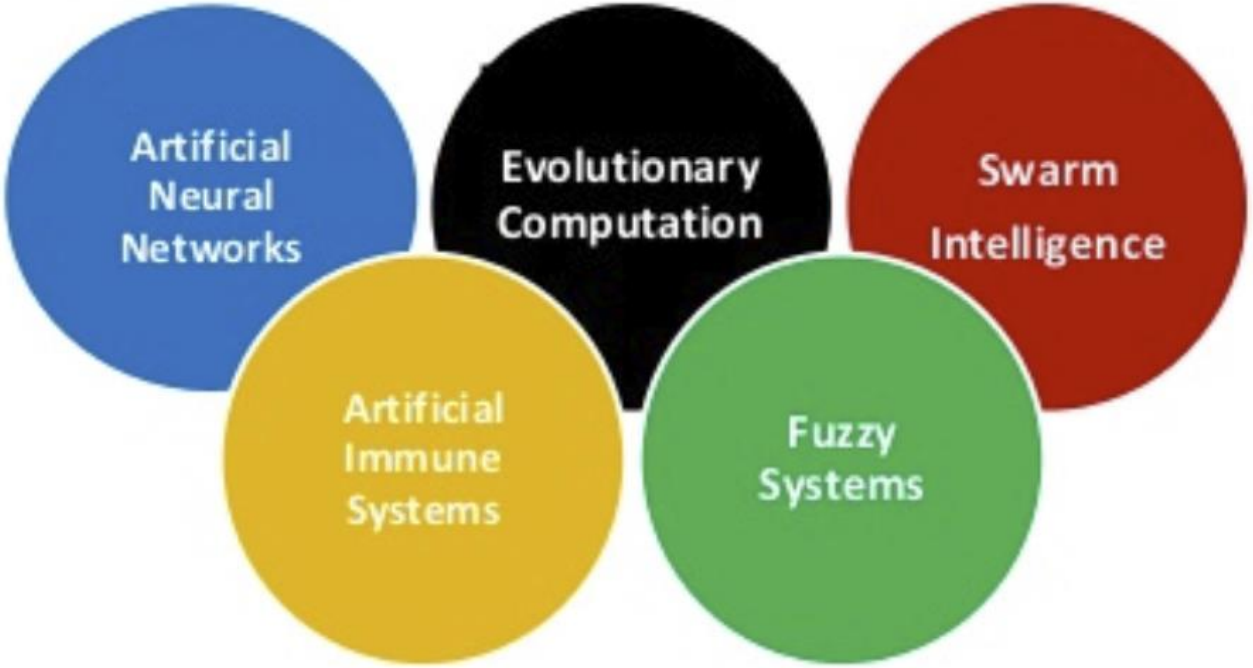
ML «ingredients»



https://vas3k.com

Computational intelligence

- According to Engelbrecht (2006)



Computational Intelligence: An Introduction
Andries P. Engelbrecht

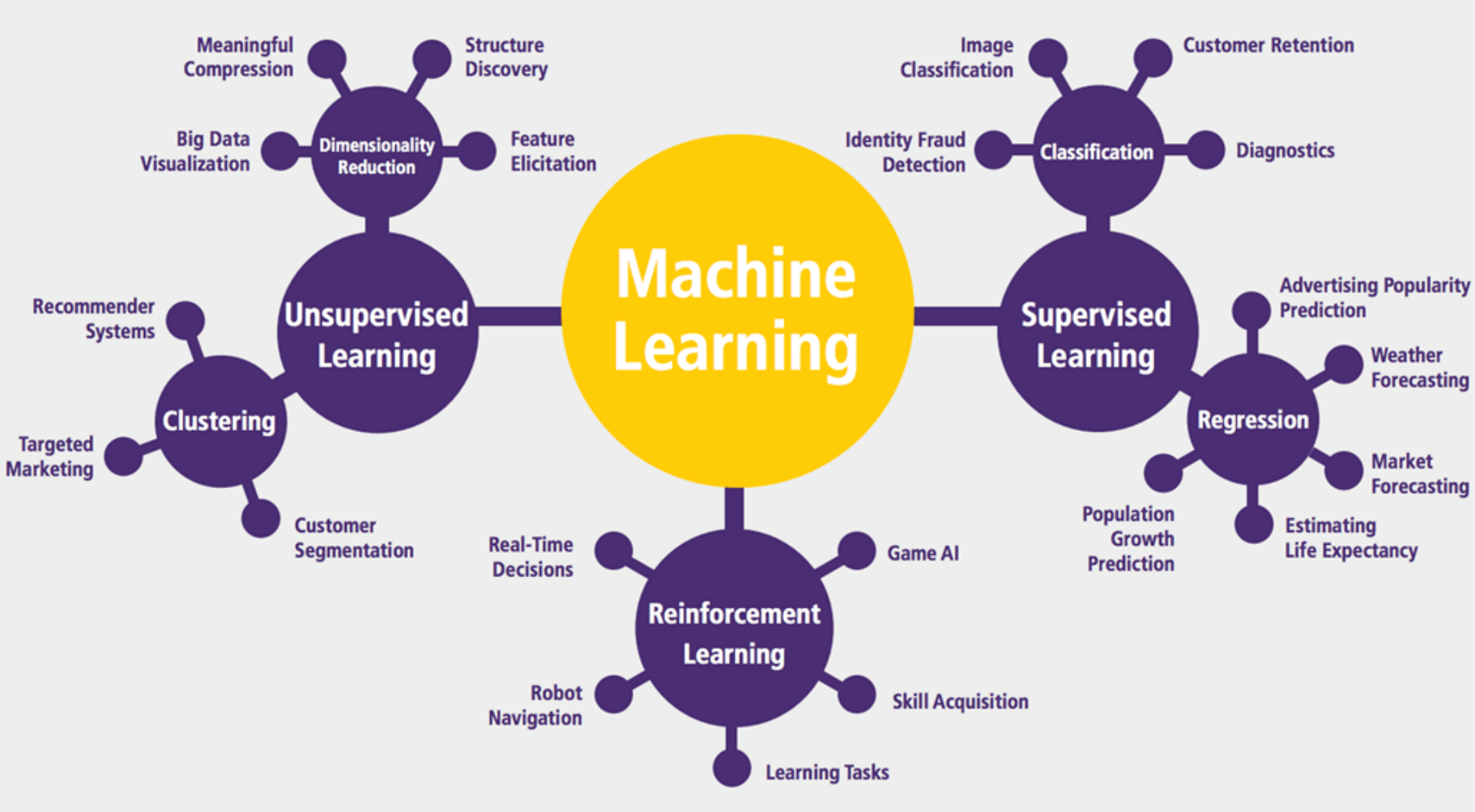
Computational intelligence

- According to the *IEEE computational intelligence society* (current definition):

"theory, design, application, and development of *biologically* and *linguistically* motivated **computational paradigms** emphasizing **neural networks, connectionist systems** (*), **genetic algorithms**, **evolutionary programming**, **fuzzy systems**, and **hybrid intelligent systems** in which these paradigms are contained"

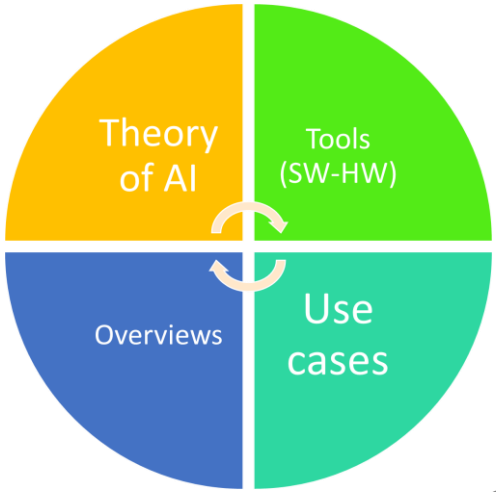
(*) represents mental or behavioral phenomena as emergent processes of interconnected networks of simple units.

Machine Learning taxonomy



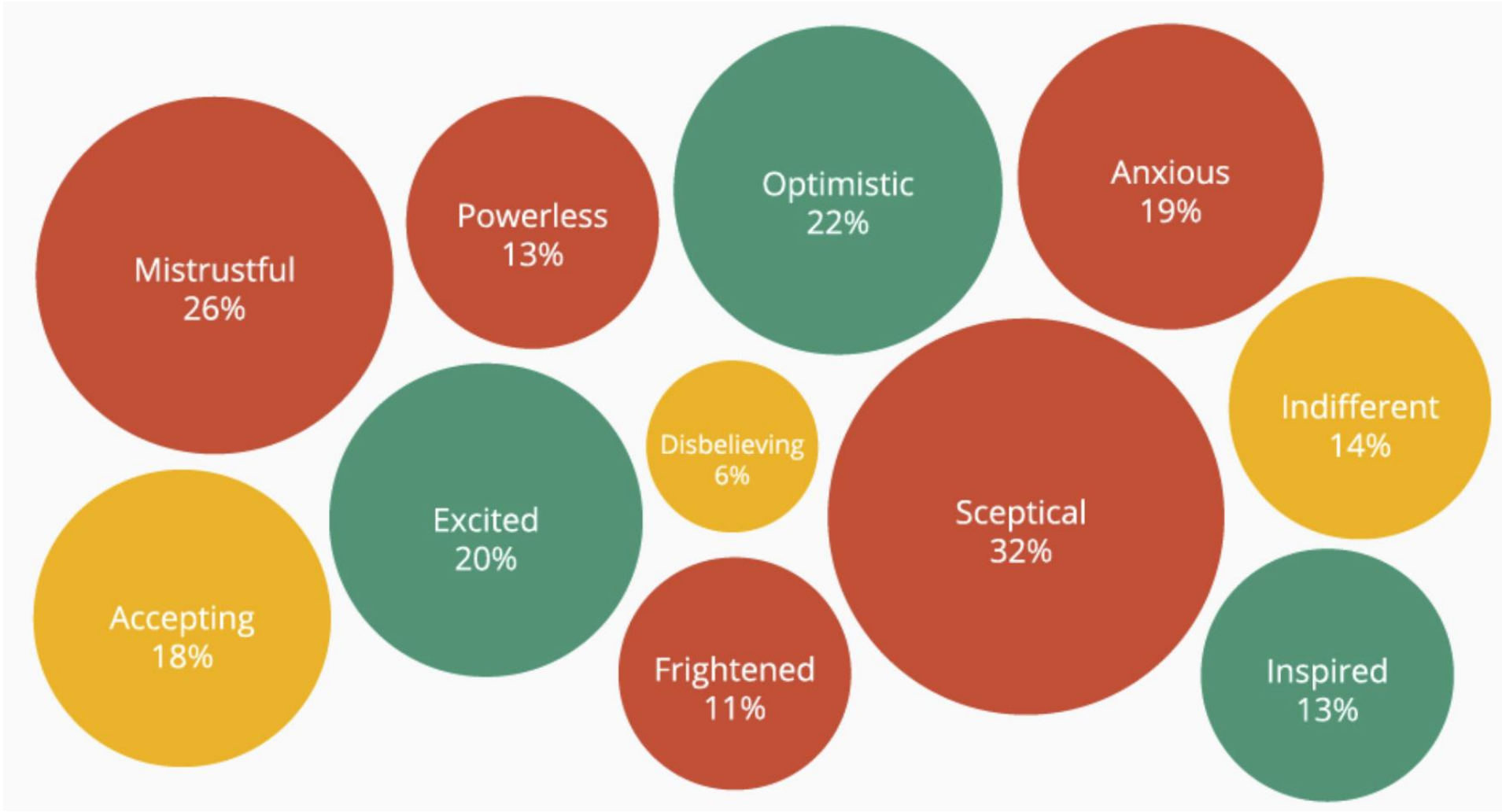
Overview

Opinions about AI in users and experts



Mixed!

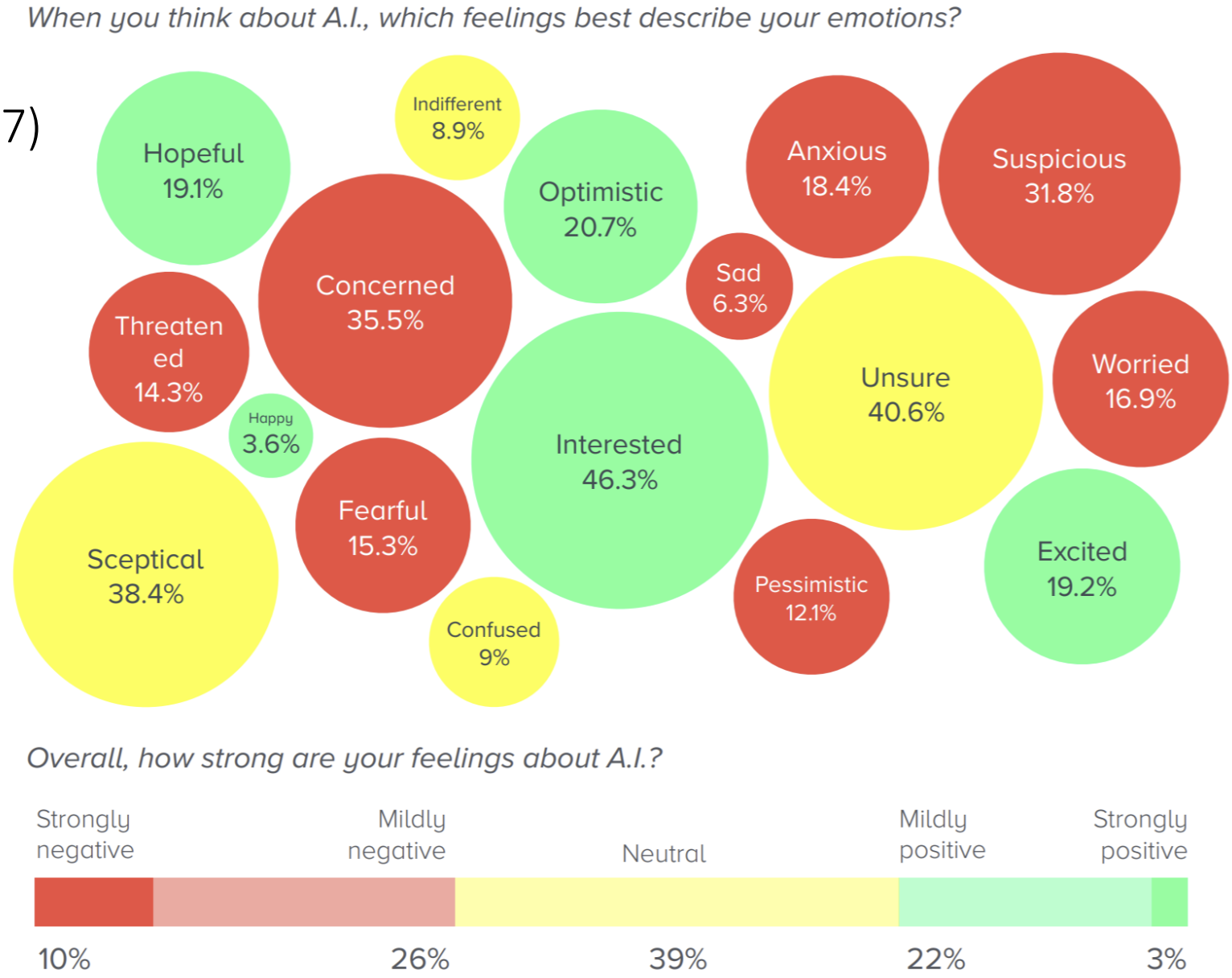
Feeling about AI (2000) in UK



Based on a >2000 UK user survey

UK
(October 2017)

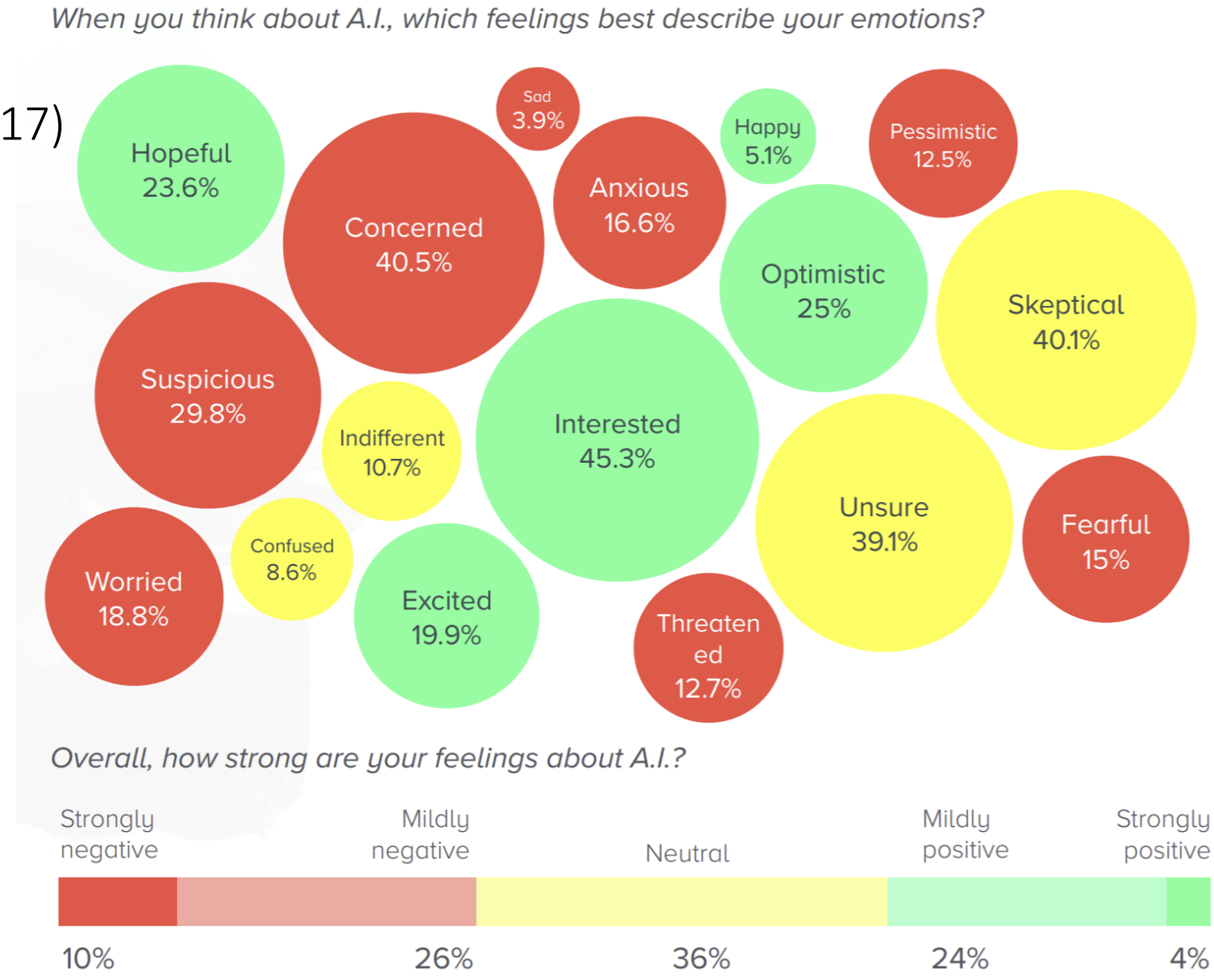
Mixed! again



https://think.syzygy.net/ai-report/uk

USA
(October 2017)

Mixed!
The same…



https://think.syzygy.net/ai-report/uk

Some (positive) opinions

- “A **breakthrough** in machine learning would be worth ten Microsofts”, Bill Gates, Chairman, Microsoft
- “Machine learning is the **next Internet**”, Tony Tether, Director, DARPA
- Machine learning is the **hot new thing**”, John Hennessy, President, Stanford

Let's have a look also to different opinions (quite different opinions)…

Opinions: Stephen Hawking

- Artificial intelligence (AI) will be “either the best or worst thing” for humanity.
- “The development of full artificial intelligence could spell the end of the human race….It would take off on its own, and re-design itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded.”

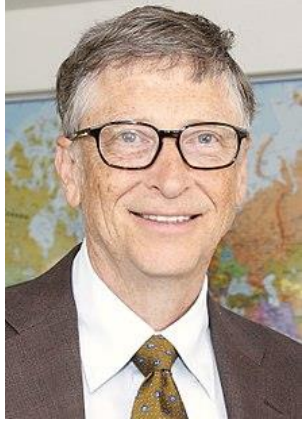
Larry page

“Artificial intelligence would be the ultimate version of Google. The **ultimate search engine** that would **understand everything on the web**. It would understand exactly what you wanted, and it would give you the right thing. We're nowhere near doing that now. However, we can get incrementally closer to that, and that is basically what we work on.”

Elon musk

- “The pace of progress in artificial intelligence (I’m not referring to narrow AI) is incredibly fast. Unless you have direct exposure to groups like Deepmind, you have no idea how fast—it is growing at a pace close to exponential. The risk of **something seriously dangerous happening** is in the five-year timeframe. **10 years at most**.”
- “I’m increasingly inclined to think that there should be some **regulatory oversight**, maybe at the national and international level, just to make sure that we don’t do something very foolish. I mean with artificial intelligence we’re summoning the demon.”

Bill Gates (recent opinion)



- “Humans should be worried about the threat posed by artificial intelligence.”
- “I am in the camp that is concerned about artificial intelligence. **First the machines will do a lot of jobs for us** and not be super intelligent. That should be positive if we manage it well. A few decades after that though the intelligence is strong enough **to be a concern**”.
- “I agree with Elon Musk and some others on this and don’t understand why some people are not concerned.”

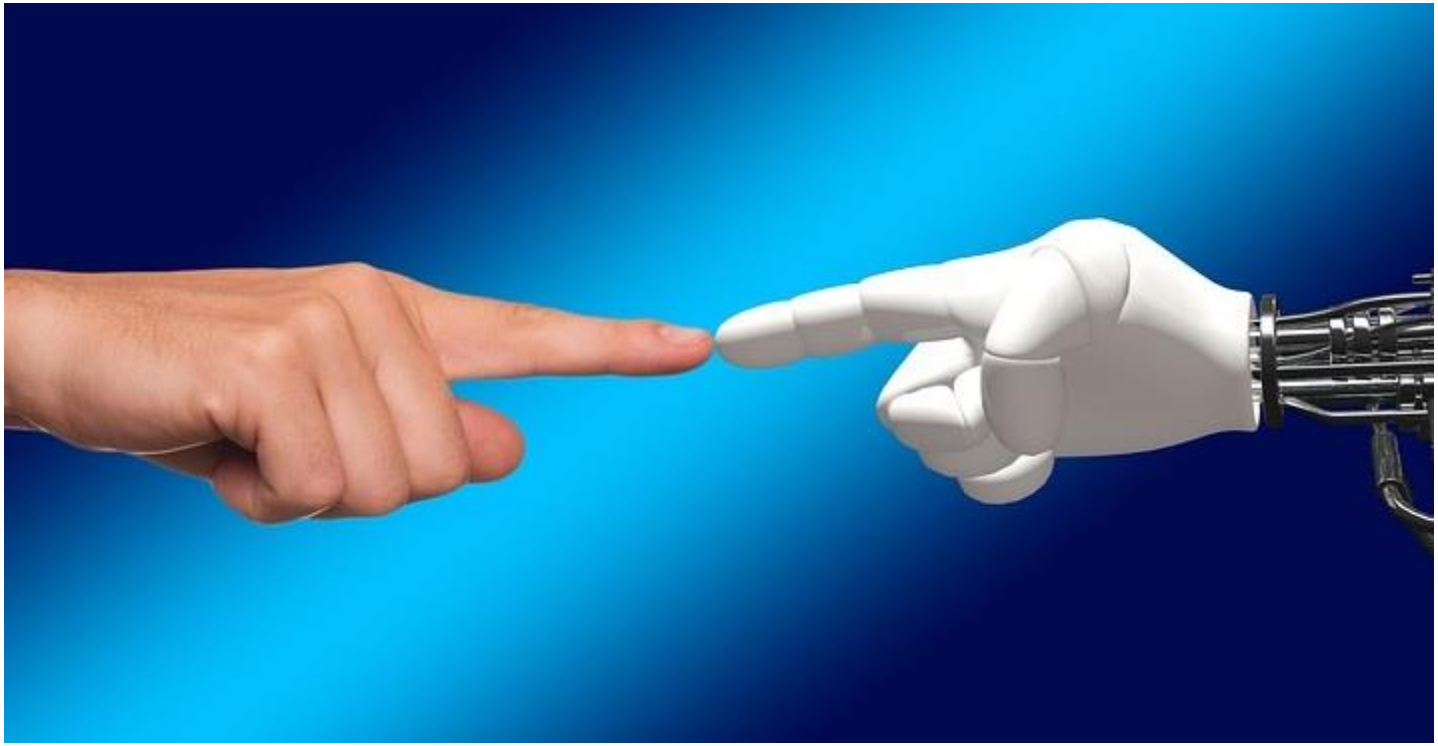
Andrew Ng

Computer scientist, investor, and writer
With >167000 citations….

- Worrying about AI evil superintelligence today is like worrying about overpopulation on the planet Mars. We haven't even landed on the planet yet!

- AI is the new Electricity

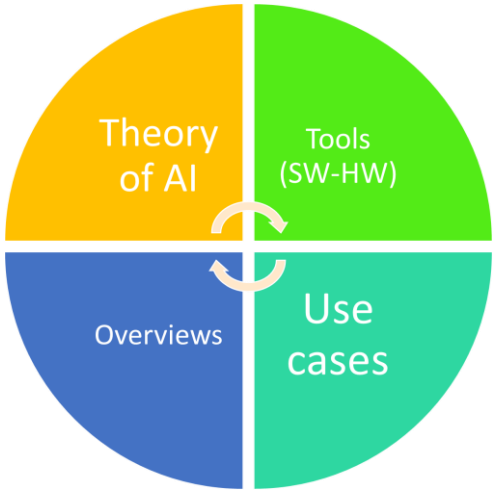
What is your opinion?



Write it down …. We'll go back to that

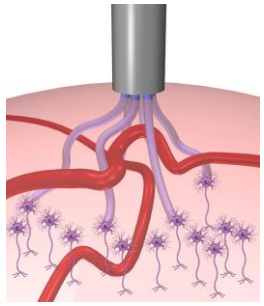
Overview

Usability and maturity

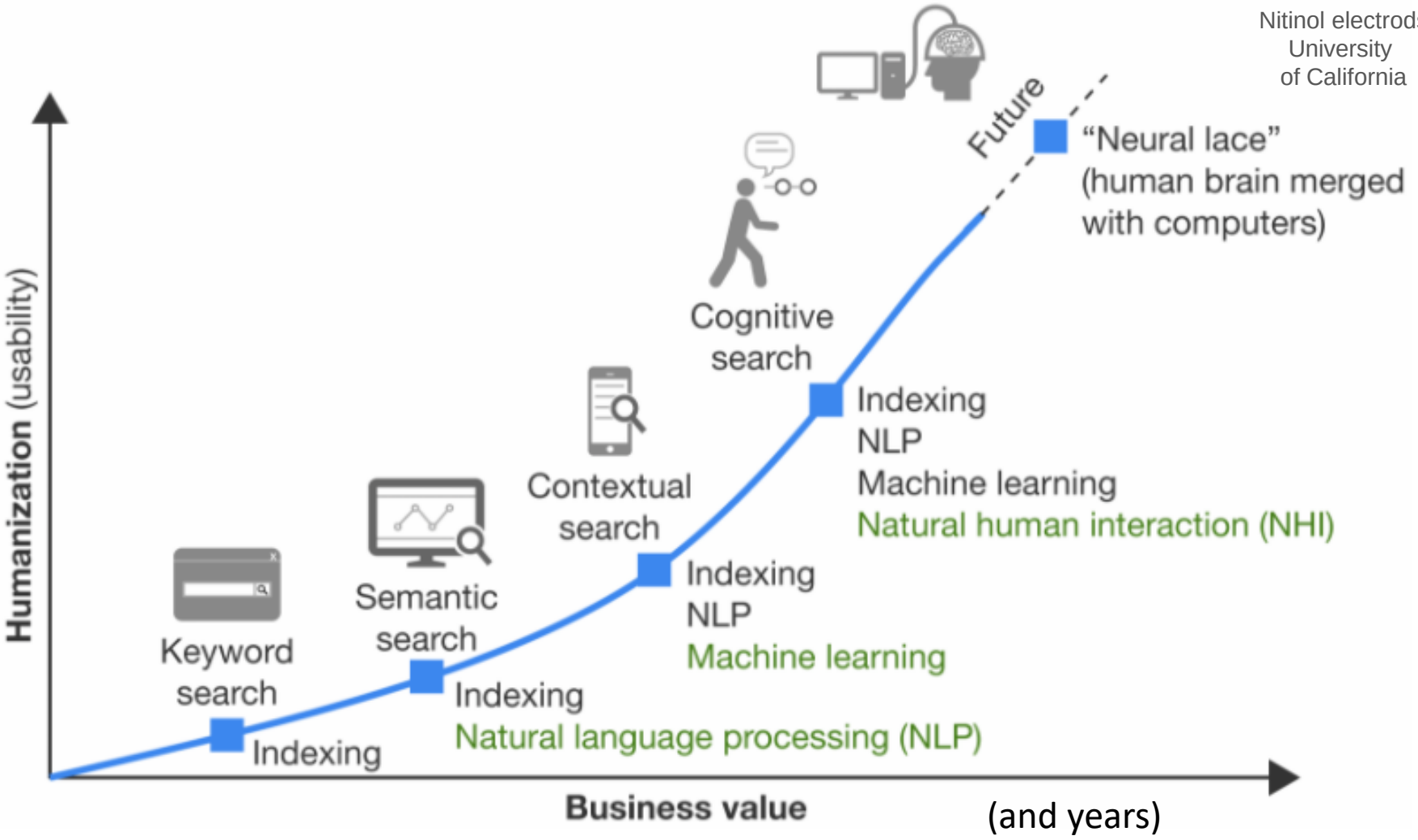


Usability in ML and Business value

Neuralink
E. Musk

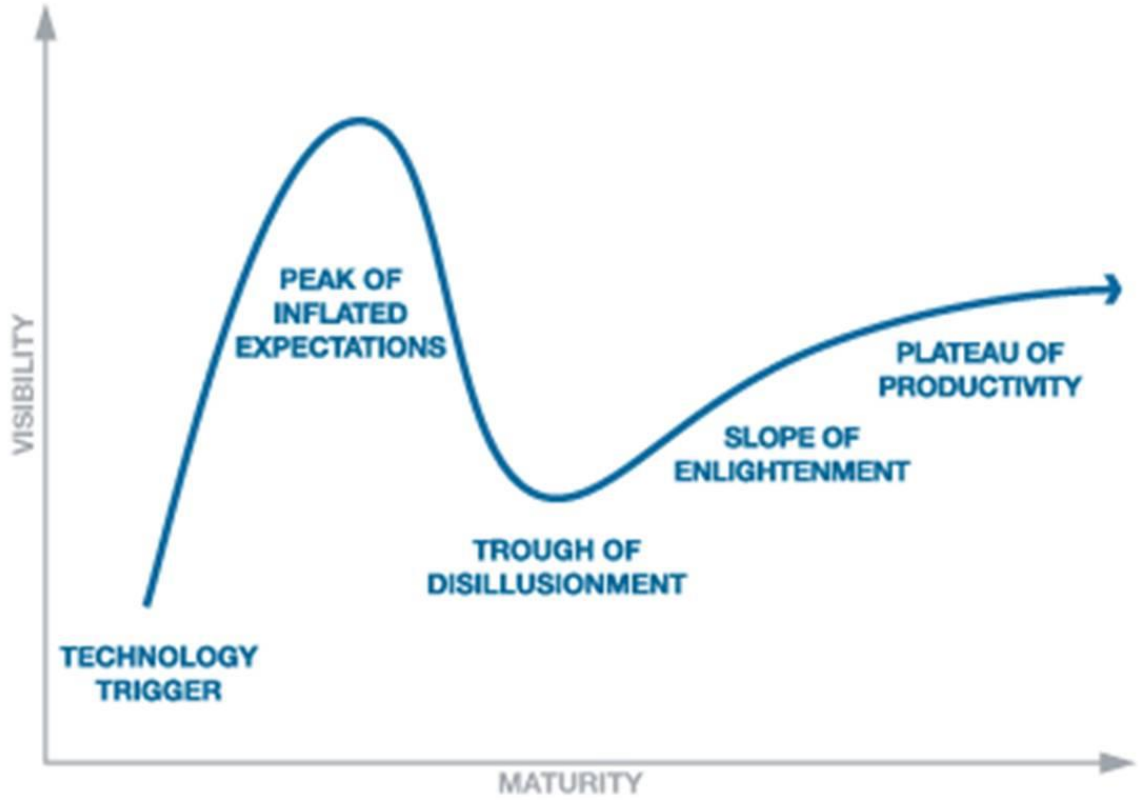


Nitinol electrods
University
of California



The Gartner “Hype cycle”

(Expectations)



(Time)

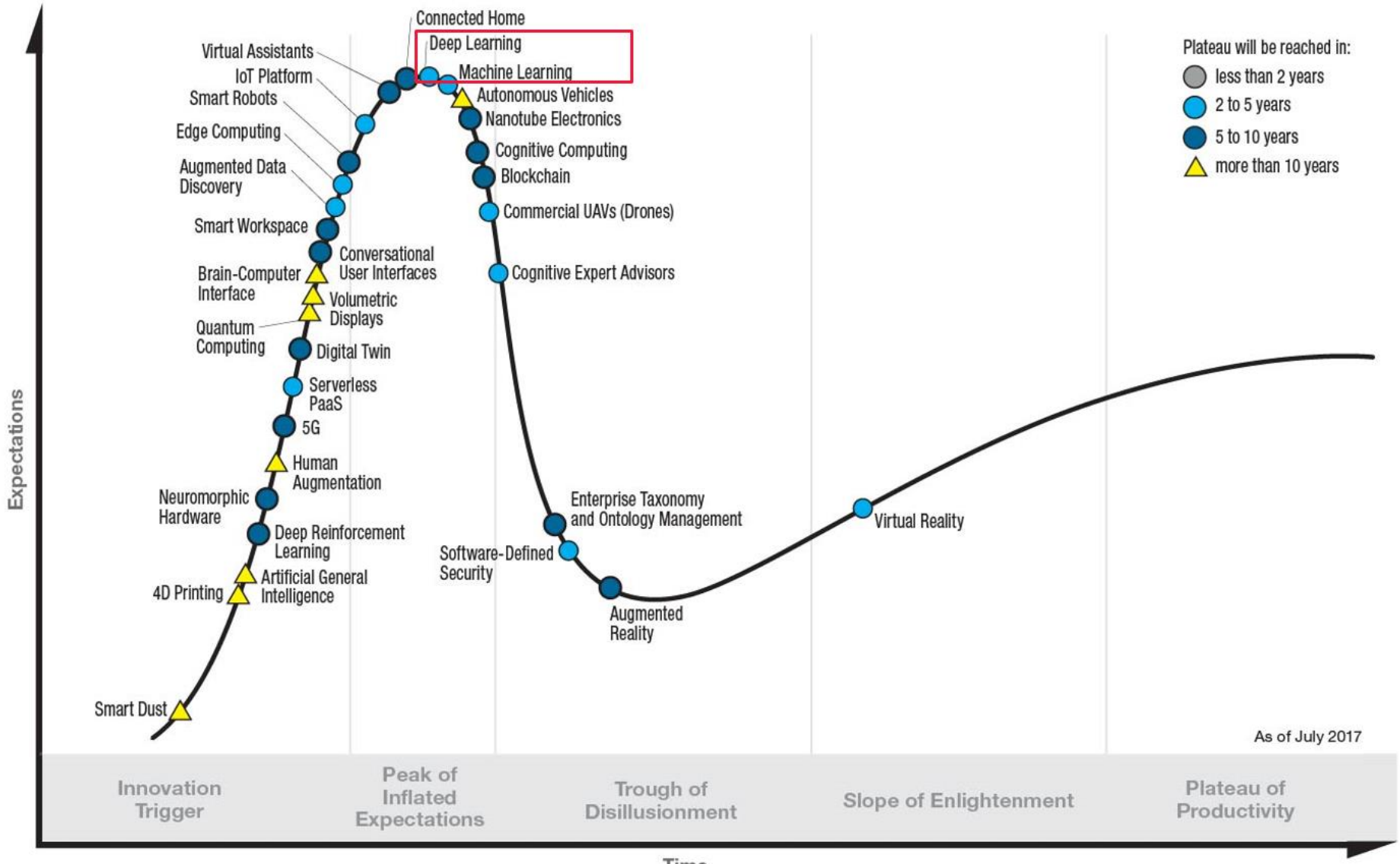
The plot represents the maturity, adoption, and social application of specific technologies.

Ok **but**….

- is not a cycle
- is not scientific in nature,
- no changes over time in the speed at which technology develops
- … and many other
- But’s is interesting…

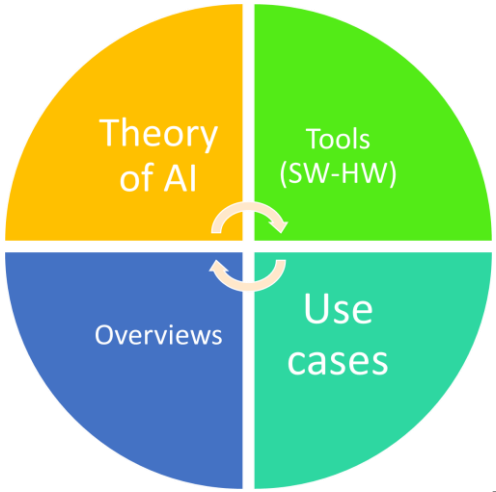
Gartner: American research, advisory and information technology firm

The AI/ML hype cycle graph



Toolboxes

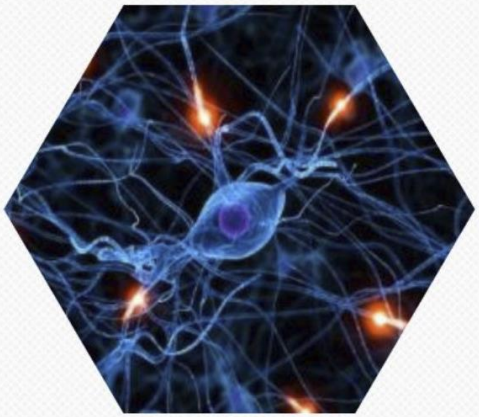
The 4 main drivers



The three main drivers for AI

More Data

Better Models
and Algorithms



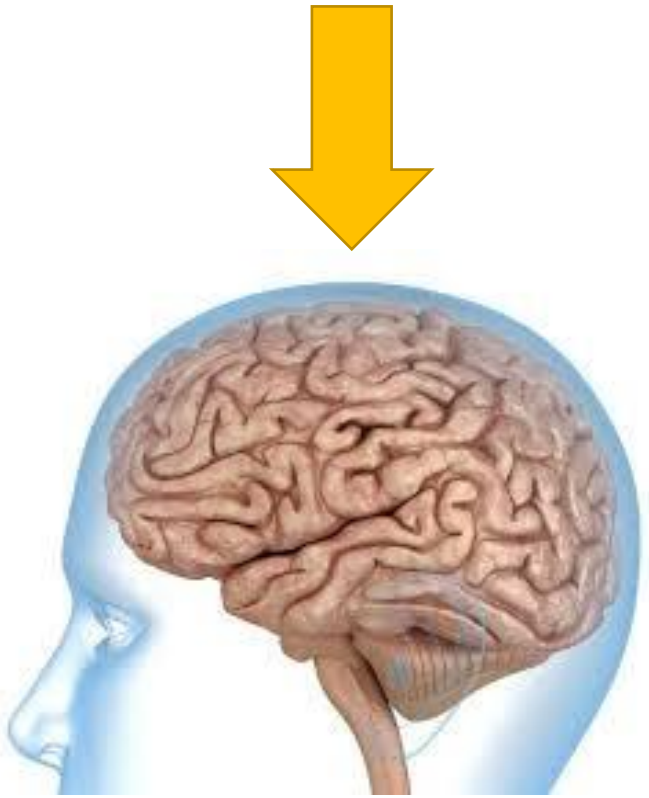
Powerful GPU Accelerators

Especially for Deep Learn

Keep it in mind…

The 4° driver

Intelligent systems for industry, supply chain and environment



- The designer's brain!
  - A priori knowledge
  - Data selection
  - Data Filtering and Enhancing
  - Model selection
  - Learning technique choice
  - Experiment Design
  - Avoid brute force
  - Improve Explainability
  - Hybrid systems
    - Dividi et impera

Main points

- AI is now almost everywhere (hype) and still improving but
  - It is not known what “exactly” a n. network is learning
  - Some relevant human features are not (yet) available
- Small perturbation of the inputs can provide a large change in the output of a neural network
  - **Again**: it is not clear what the network is learning
- Opinions about AI are quite different
- The 4th driver (the designer’s brain) is the most important one!