HSLU Lucerne University of Applied Sciences and Arts
School of Computer Science
& Information Technology

The History and Development of Machine Learning

Lecture Series in Artificial Intelligence & Machine Learning

Dr. Donnacha Daly
Head of Bachelor in Artificial Intelligence & Machine Learning
Member of the Algorithmic Business (ABIZ) Research Lab

donnacha.daly@hslu.ch | hslu.ch/bachelor-ai | www.abiz.ch

Topics Covered in Today's Lecture

A. Machine Learning in the Context of Artificial Intelligence (AI)

B. The Connectionists: Artificial Neural Networks (ANN's), Machine Learning (ML) and Deep Learning (NN)

C. Three Pillars of Machine Learning: Data, Algorithms & Processing Power

D. Machine Learning Today: Progress, State-of-the-Art & Challenges

Part A

Machine Learning in the Context of Artificial Intelligence (AI)

What is Machine Learning?

*"… a field of study that gives **computers** the ability to learn without being explicitly programmed"*

(- Arthur Samuel, 1959)

*"… the study of **computer algorithms** that allow computer programs to automatically improve through experience"*

(- Tom Mitchell, 1997)

*If you want to learn more about ML, check out the really excellent online lectures by Andrew Ng**

Artificial Intelligence

Completing tasks which normally require human level intelligence

Machine Learning

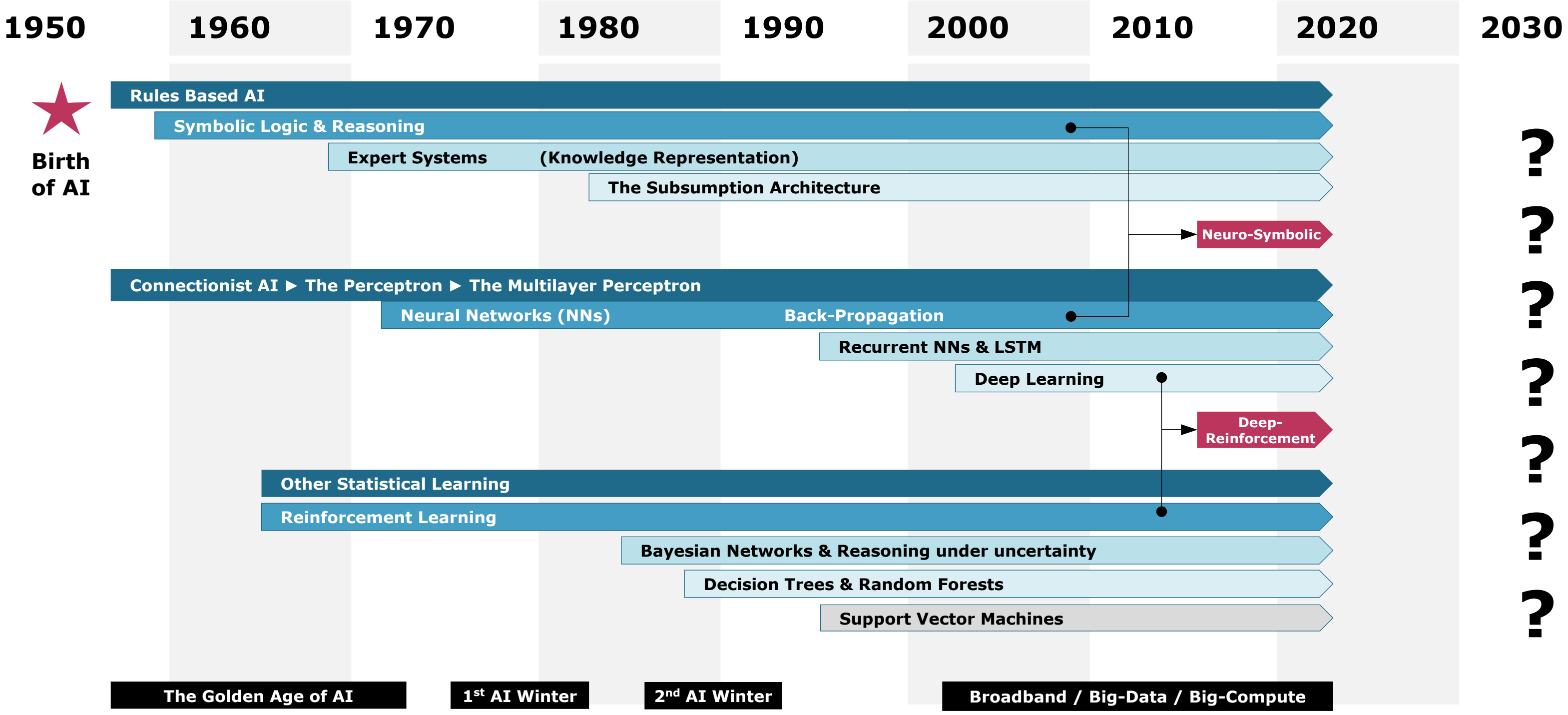
Statistical methods enabling machines improve with experience

Deep Learning

Many-layered learning-networks progressively extracting higher level features from raw input

* https://www.youtube.com/playlist?list=PLLssT5z_DsK-h9vYZkQkYNWcItqhlRJLN

A Brief History of AI / ML



Dr. Donnacha Daly
donnacha.daly@hslu.ch

The Origins of AI

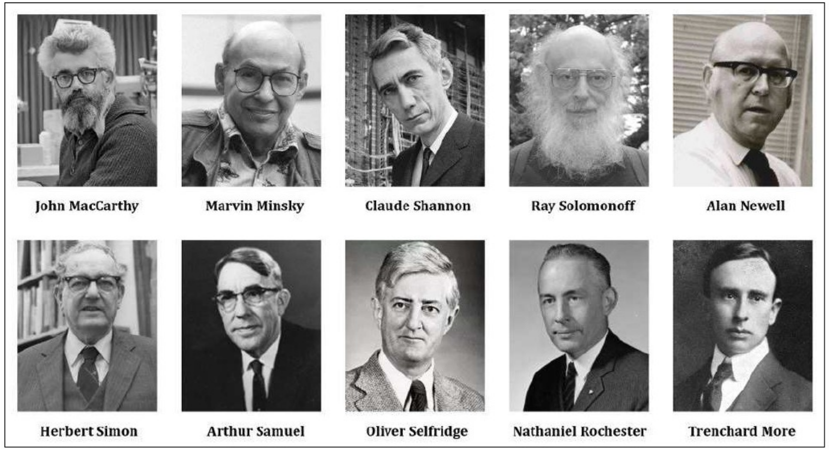
1945 Worlds first programmable computer: ENIAC

1950 Alan Turing: **The Turing Test**. Can a machine pass itself off as a human in its responses?

1956 **Dartmouth conference**: official birth of AI

1960's The Golden Years: Interpret the human brain as a machine

Dartmouth 1956 - The Founding Fathers of AI



"The study [of AI] is to proceed on the basis of conjecture that every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it"

https://medium.com/rla-academy/dartmouth-workshop-the-birthplace-of-ai-34c533afe992

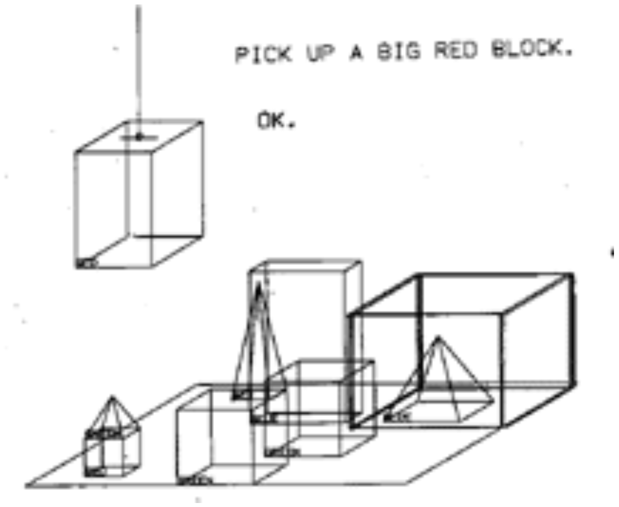
The Golden Years: Logic & Reasoning

1956 Newell & Simon : 38 Proofs from Pinicipia Mathematica

1963 DARPA gives MIT $2.2M for machine aided cognition

1964 IBM Handwriting Recognition Demo at NY World Fair
Write Date → Get Headlines
Translate English Russian texts

1967 Marvin Minsky: "Within a generation, the problem of creating AI will substantially be solved"



1970 Terry Winnograd (MIT): Natural language processing for moving blocks in a Microworld

http://hci.stanford.edu/~winograd/shrdlu/

The First AI Winter

1970 Minsky: In 3-8 years we will have a machine with the general intelligence of an average human

70-73 AI mainly based on rule-based programming, fails to deliver on most promises

1973 U.S. and British Governments stop funding undirected AI research

74-80 The first **AI Winter**

Shakey, the First "Intelligent" Robot

The Stanford Research Institute 1966-1972

***Moravec Paradox**: "It is easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility"*

https://thenewstack.io/remembering-shakey-first-intelligent-robot/

The Second Wave: Expert Systems

1970's Research on **Expert Systems** which augments reasoning with **Knowledge**

1980's Large-scale deployment: Two thirds of Fortune 500 companies used the technology in daily business.

1985 Over $1B Spent on AI research & development by more than 150 companies

Expert Systems: Success Criteria

Problem type criteria

task involves mostly symbolic processing;
test cases are available;
problem task is well-bounded (i.e. taking a few minutes to a few hours to solve);
task is required to be performed frequently;
written materials exist explaining the task;
task requires only cognitive skills;
experts agree on the solutions.

Expert criteria

an expert exists;
the expert is cooperative;
the expert is articulate;
the expert's knowledge is based on experience, facts and judgement;
other experts exist in the task.

Domain area personnel criteria

a need exists for developing an expert system for that task;
the task would be provided with the necessary financial support;
top management supports the project;
the domain area personnel have realistic expectations on the use of an expert system;
users would welcome the expert system;
the knowledge is not politically sensitive or controversial.

"Expert Systems: A Short Introduction", Jay Liebowitz, 1995, Engineering Fracture Mechanics

The Second AI Winter

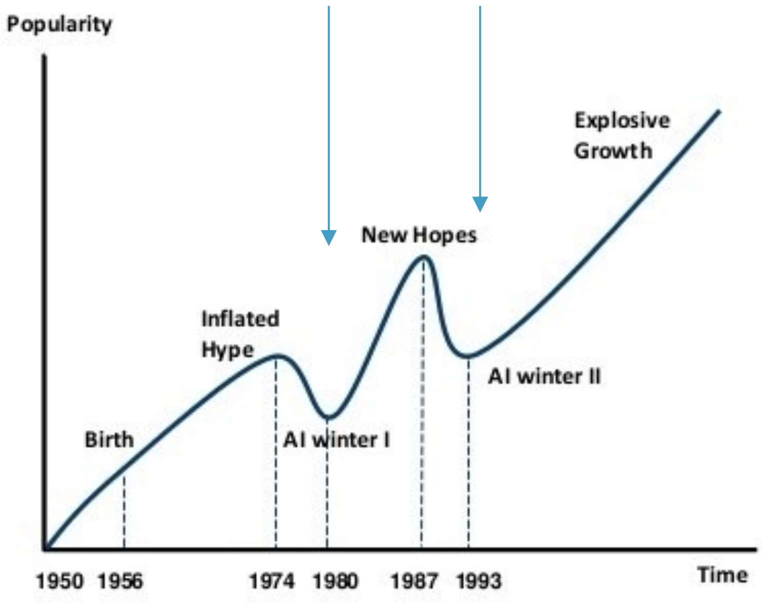
1987 Collapse of market for specialized AI hardware (LISP Machines)

87-89 DARPA Funding Slashed by one third

89-92 Most deployed systems fall into disuse

Steven Pinker *"The main lesson of the first 35 years of AI research is that the hard problems are easy and the easy problems are hard"*

"AI Winters"



https://www.actuaries.digital/2018/09/05/history-of-ai-winters/
"Early Expert Sytems, Where are they Now?" T.G. Gill, MIS Quarterly, 1995

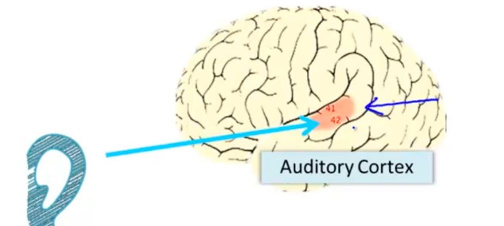
Part B

The Connectionists:
Artificial Neural Networks (ANN’s),
Machine Learning (ML) and Deep Learning (NN)

How do Humans and Animals Learn?

There is a hypothesis that there is **one type of learning**, used by every part of the brain

Example: in animal experiments, when optic nerves are connected to the auditory cortex: the animal learns to see!



Auditory cortex learns to see

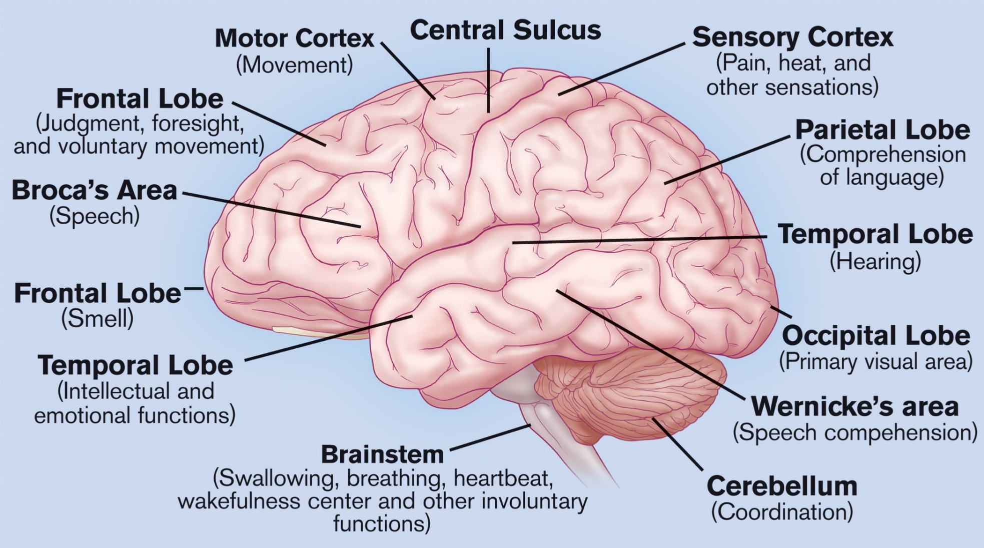
Example: In 1848 Phinneas Gage had a 3-foot rod go through his cheekbone and out the top of his skull. He lived another 11 years and learned to adapt to his condition.

Example: Some blind people learn to echolocate like bats, making clicks with their mouths.

Can we explain learning by examining structures in the brain?

https://www.wpr.org/how-phineas-gages-freak-accident-changed-brain-science
https://www.newscientist.com/article/2145962-this-is-how-some-blind-people-are-able-to-echolocate-like-bats/#ixzz6b9VWpdHW

The Human Brain



Weight:
1.3 – 1.5kg

Neurons:
~100 Billion (~10^{11})

Connections:~150 Trillion (~10^{14})

Connectivity:
Each neuron connects to 1k-10k others

All parts of the brain are made of the same neural tissue. It is a highly connected network. The original **Connectionists** believed that this was at the center of human learning and intelligence and they set out to copy it within computers.

https://youtu.be/m3U1_Zv4_Ik

Networks that Learn

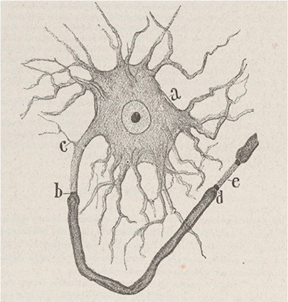
1887 Santiago Ramón y Cajal at Uni. Barcelona discovers **neurons** as fundamental to the nervous system

1943 McCulloch & Pitts **model the neuron** as a **logic unit**. Show that network of such neurons can do any computation.

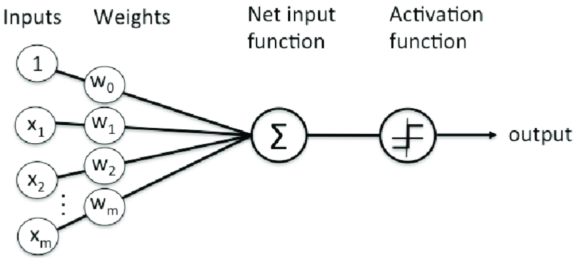
1949 Donald Hebb conceives an update Rule for **learning**: *Neurons that fire together, wire together.*

1958 Frank Rosenblatt takes the ideas into computer research: he invents the **Perceptron**, a single layer **neural network**

Cajal's sketch of a Neuron



Rosenblatt's Perceptron



https://www.brainfacts.org/brain-anatomy-and-function/cells-and-circuits/2018/santiago-ramon-y-cajal-the-first-modern-neuroscientist
https://vitalflux.com/java-implementation-for-rosenblatt-perceptron/

The Difficult Birth of Neural Networks

1969 Minsky & Papert claim much scientific writing on Perceptrons is without scientific value

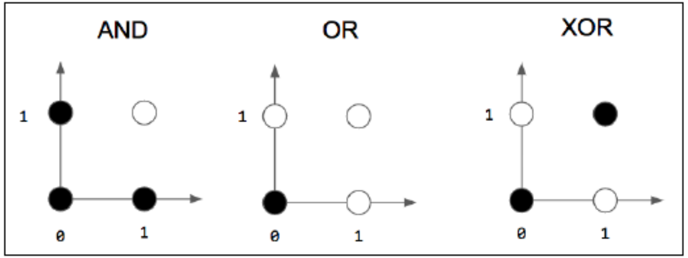
70's Rosenblatt dies. Rule-based programming dominates. Funding crashes. First AI-Winter.

80's **Neural networks** show promise and make comeback. By '91 there are >10k NN researchers in the U.S. alone.

1987 Yann LeCun proposes modern form of **back-propagation** learning algorithm for ANNs

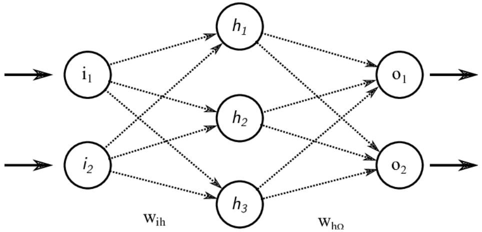
1993 The Second AI Winter! ☹

The perceptron is a supervised binary linear classifier. Can you separate two groups using a straight line?



AND, OR are linearly separable; XOR is not

Neural networks with hidden layers can implement non-linear classification



Deep Learning and the AI Boom

2005 Geoffry Hinton exposes internal function of multi-layered Neural networks with many hidden units

2010 Affordable GPUs give boost to large Neural Networks

2015 Machines outperform humans at image recognition. Facebook launches DeepFace (9 layers, 120M params)

2016 AlphaGo from DeepMind beats world champion Lee Sodol at Go

2020 Open-AI releases GPT-3 language model with 175 Billion parameters

Turing Award 2018

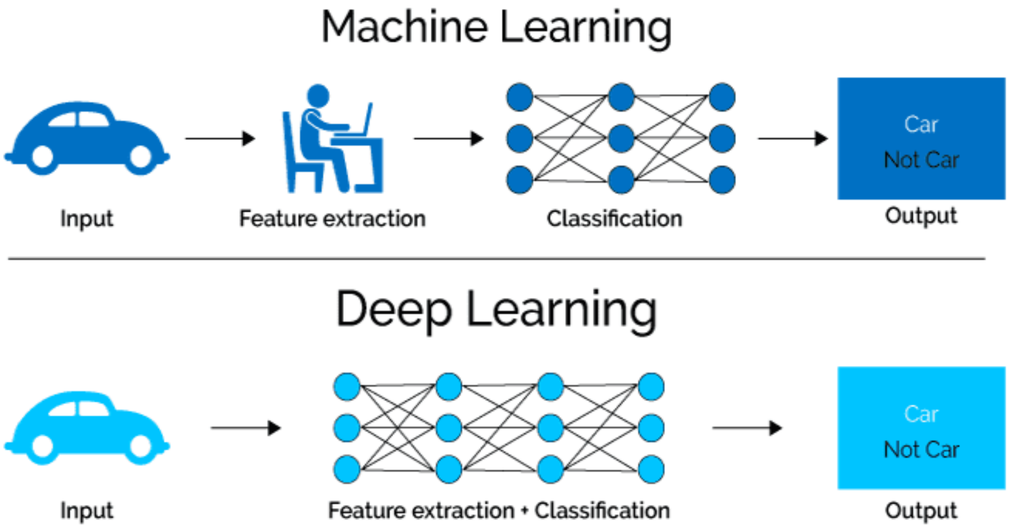
Key Contribution: *Generative Adversarial Networks*

Key Contribution: *Backpropagation allows NNs discover their own internal representations of data*

Key Contribution: *Convolutional Neural Networks*

https://awards.acm.org/about/2018-turing
"Artificial Intelligence in the 21st Century", Kong et al., IEEE Acess, 2018

Deep Learning Using Many Layers



The core idea of deep learning is to exploit the ability of large neural networks to learn features within the data relevant to the task at hand

To do this we need lots of data, sophisticated algorithms and lots of processing power

https://medium.com/@kuzuryu71/deep-learning-vs-machine-learning-7db5cfd75321

PART C

Three Pillars of Machine Learning: Data, Algorithms & Processing Power

How Much Data is Generated Every Day?

- 100 Million Uploads to Instagram
- 500 Million Tweets on Twitter
- 3.5 Billion Photos Taken
- 5 Billion Web Searches
- 65 Billion Messages on Whatsapp
- 300 Billion Emails

This is mostly well known consumer data. There are also vast amounts of business data being captured in digitalization programs and other public data being captured by government agencies.

- **Machine learning algorithms can be trained to higher performance levels with labelled data.**
- **This is more expensive, and far more scarce than typical unstructured and unlabeled data**
- **But large, open-source, labelled data sets are available…**

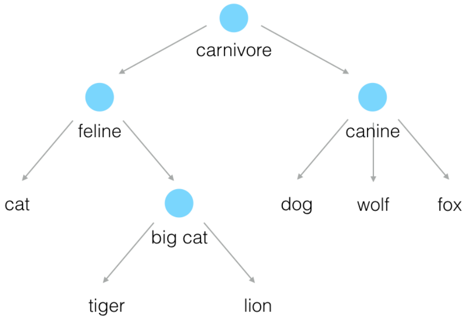
https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

Hochschule Luzern
Computer Science &
Information Technology

Example: Open Source Image Data

ImageNet

- Biggest and best known, out of Stanford
- **14M labelled images**
- 1 Million images with bounding box notations
- 22k image categories,

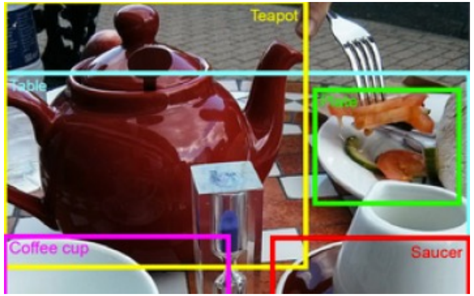


Coco Dataset

- **Co**mmon **O**bjects in **Co**ntext: backed by Microsoft and Facebook
- A large-scale object detection, segmentation, and captioning dataset
- **~200k labelled images**

Open Images

- Released by Google
- **~9M labelled images**
- object bounding boxes
- object segmentation
- visual relationships



https://zbigatron.com/the-top-image-datasets/

Open Source Algorithms Are Also Available

Hochschule Luzern
Computer Science &
Information Technology

Tensorflow

Google's TensorFlow is JavaScript-based and comes equipped with a wide range of tools and community resources that facilitate easy training and deploying Deep Learning

TensorFlow

Pytorch

Open-source DL framework from **Facebook**. Designed to speed process from prototyping to production. It has a C++ frontend atop a Python interface

PyTorch

Sonnet

Developed by **DeepMind**, Sonnet is a high-level library designed for building complex neural network structures in TensorFlow.

Sonnet

Keras

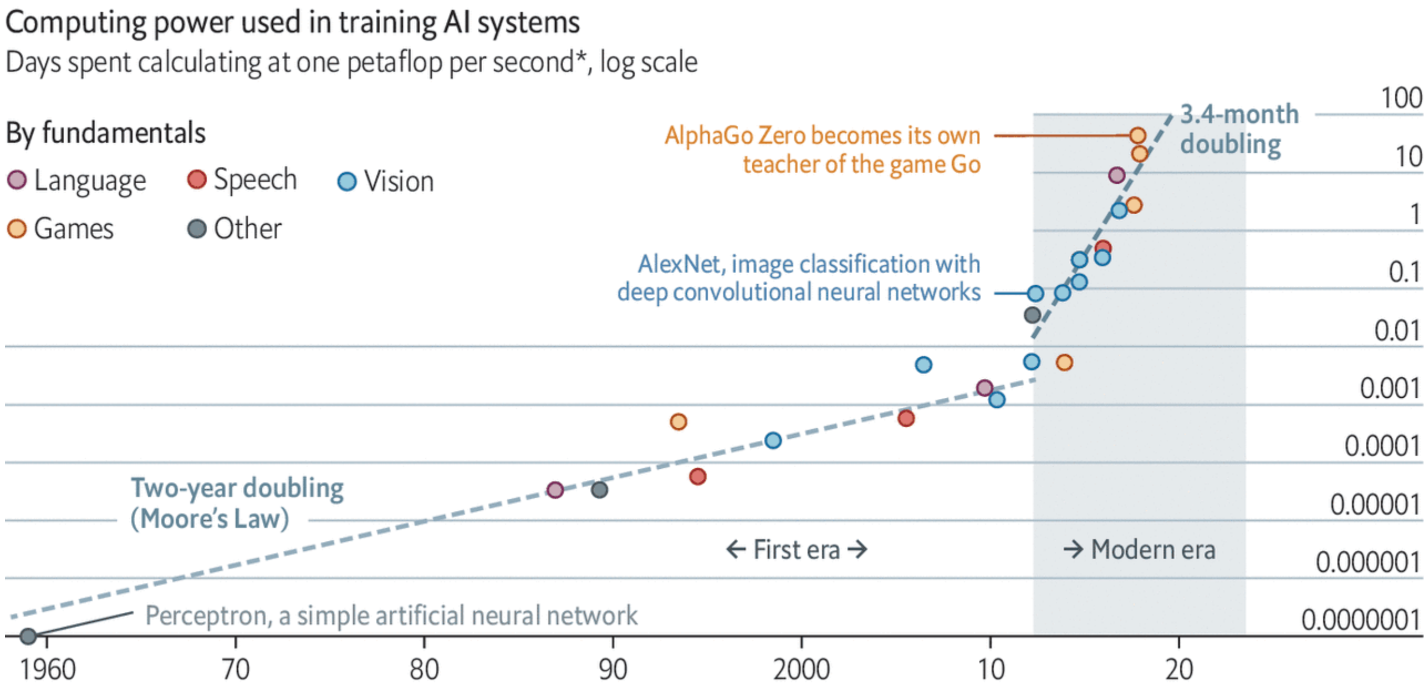
Keras is fast and open source!. It can process massive volumes of data while accelerating the training time for models. Provides a Python interface to the TensorFlow library.

Keras

There are many more frameworks. It's worth to check around depending on the needs of your project.

https://www.upgrad.com/blog/top-deep-learning-frameworks/

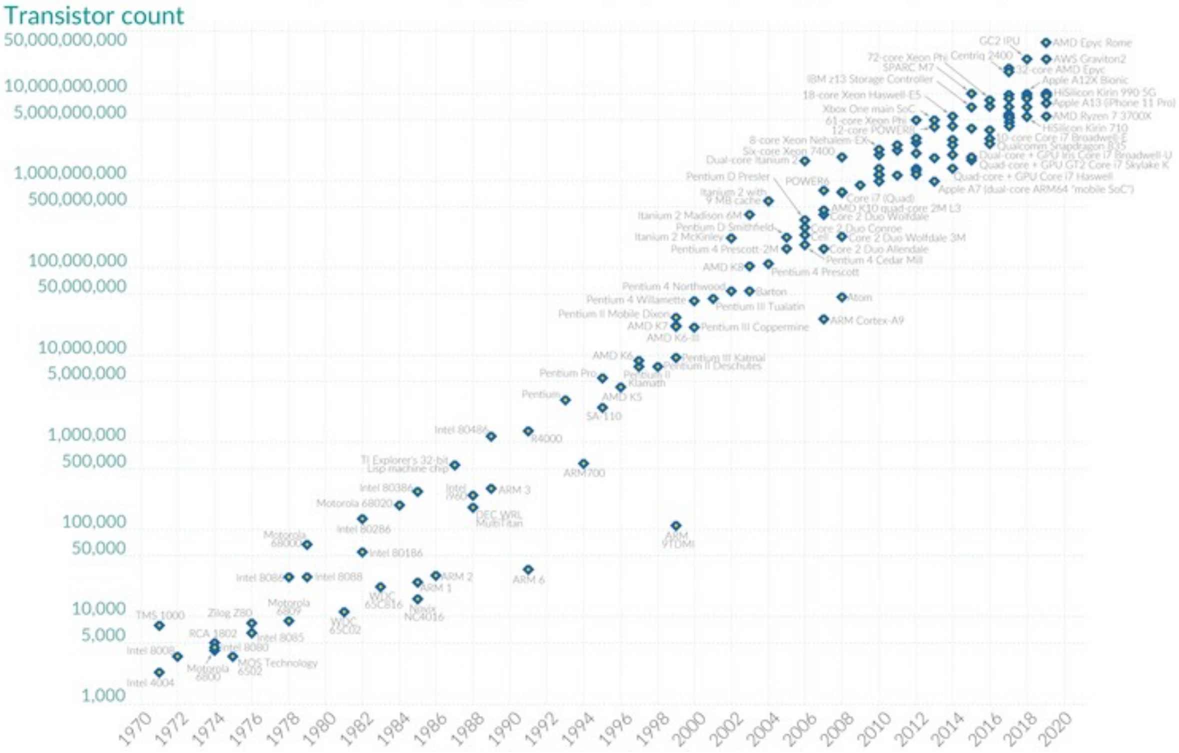
Deep Learning is Hungry for Computation



https://www.economist.com/technology-quarterly/2020/06/11/the-cost-of-training-machines-is-becoming-a-problem

Luckily we have Moore’s Law

Hochschule Luzern
Computer Science &
Information Technology



Moore's law is the observation that the number of transistors in a dense integrated circuit (IC) doubles about every two years.

https://en.wikipedia.org/wiki/Moore%27s_law

Various Options for Hardware Acceleration

Hochschule Luzern
Computer Science &
Information Technology

CPU

A Central Processing Unit is what you will find in your laptop or PC. General purpose, not so efficient.

E.g. Intel

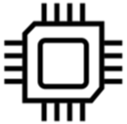
GPU

Graphics Processing Unit, well suited for tensor calculations found in neural nets. Widely used.

E.g. NVidia

FPGA

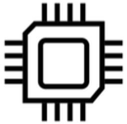
Field programmable gate array. Expensive to program, but lightning fast! Worth it for some applications.



E.g. Xilinx

ASIC

Application specific integrated circuit. Dedicated silicon. Can’t be beaten for speed. Very costly!



E.g. Google TPU

Quick and cheap **Development Time** *Long and expensive*

Long! Runs more slowly **Execution Time** *Short! Runs very fast*

Graphics Processing Units (GPU's)

Hochschule Luzern
Computer Science &
Information Technology

	Cores	Clock Speed	Memory	Price	Speed
CPU (Intel Core i7-7700k)	4 (8 threads with hyperthreading)	4.2 GHz	System RAM	$385	~540 GFLOPs FP32
GPU (NVIDIA RTX 2080 Ti)	3584	1.6 GHz	11 GB GDDR6	$1199	~13.4 TFLOPs FP32

CPU: Fewer cores, but each core is much faster and much more capable; great at sequential tasks

GPU: More cores, but each core is much slower and "dumber"; great for parallel tasks

vs.

http://vision.stanford.edu/teaching.html

Processing Power Available in the Cloud

Colab

Google Colab allows you to write and execute Python in your browser with zero configuration, free access to GPUs, and easy sharing

AWS

Amazon Web Services enables deep learning in the cloud at scale. You can quickly train custom AI models, with new algorithms.

Azure

GPUs are now part of **Microsoft**'s cloud platform, ideal for the neural networks underpinning much of modern machine learning

IBM Cloud

IBM Cloud offers GPUs on bare metal and virtual servers. Design and deploy neural networks at scale within **IBM Watson Studio**

The international cloud computing landscape is highly dynamic and changing rapidly.
Switzerland has its own national providers such as Switch.ch for universities

HSLU Lucerne University of Applied Sciences and Arts

Part D

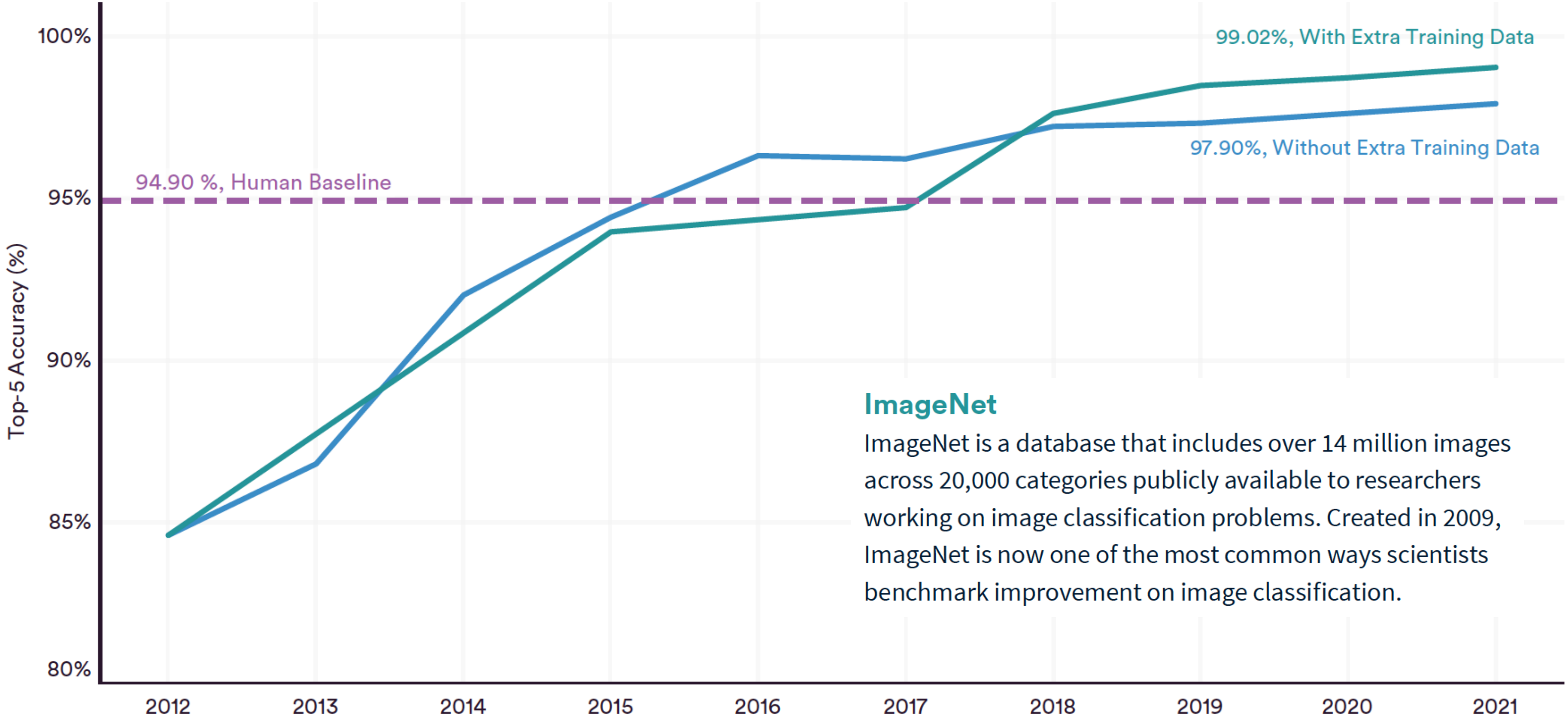
Machine Learning Today:
Progress, State-of-the-Art &
Challenges

Artificial intelligence is…

the ability of machines to complete tasks …
which normally require a human level of intelligence…
better faster and cheaper than humans can

IMAGENET CHALLENGE: TOP-5 ACCURACY

Source: Papers with Code, 2021; arXiv, 2021 | Chart: 2022 AI Index Report



ImageNet

ImageNet is a database that includes over 14 million images across 20,000 categories publicly available to researchers working on image classification problems. Created in 2009, ImageNet is now one of the most common ways scientists benchmark improvement on image classification.

Figure 2.1.3

SQUAD 1.1 and SQUAD 2.0: F1 SCORE

Source: SQuAD 1.1 and SQuAD 2.0, 2021 | Chart: 2022 AI Index Report

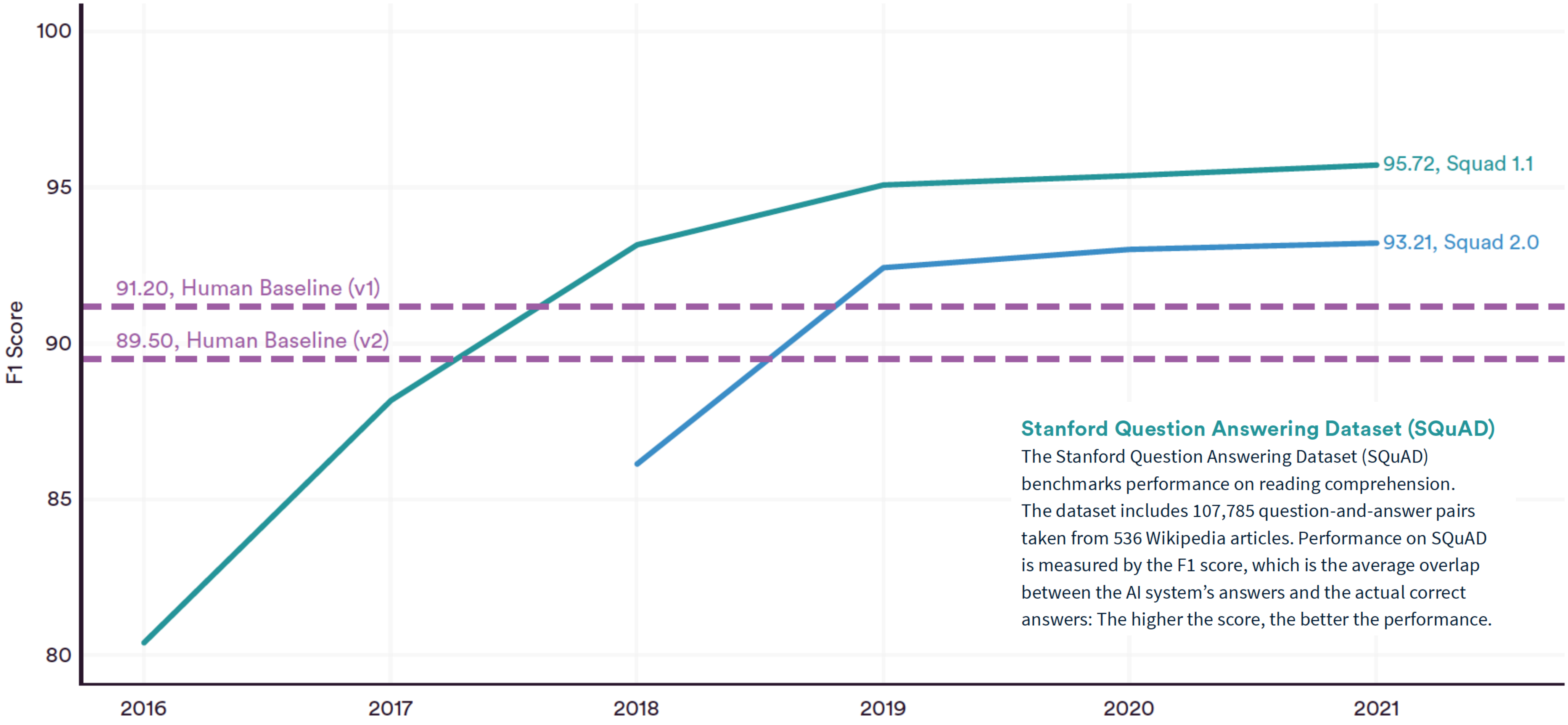


Figure 2.3.4

ABDUCTIVE NATURAL LANGUAGE INFERENCE (aNLI): ACCURACY

Source: Allen Institute for AI, 2021 | Chart: 2022 AI Index Report

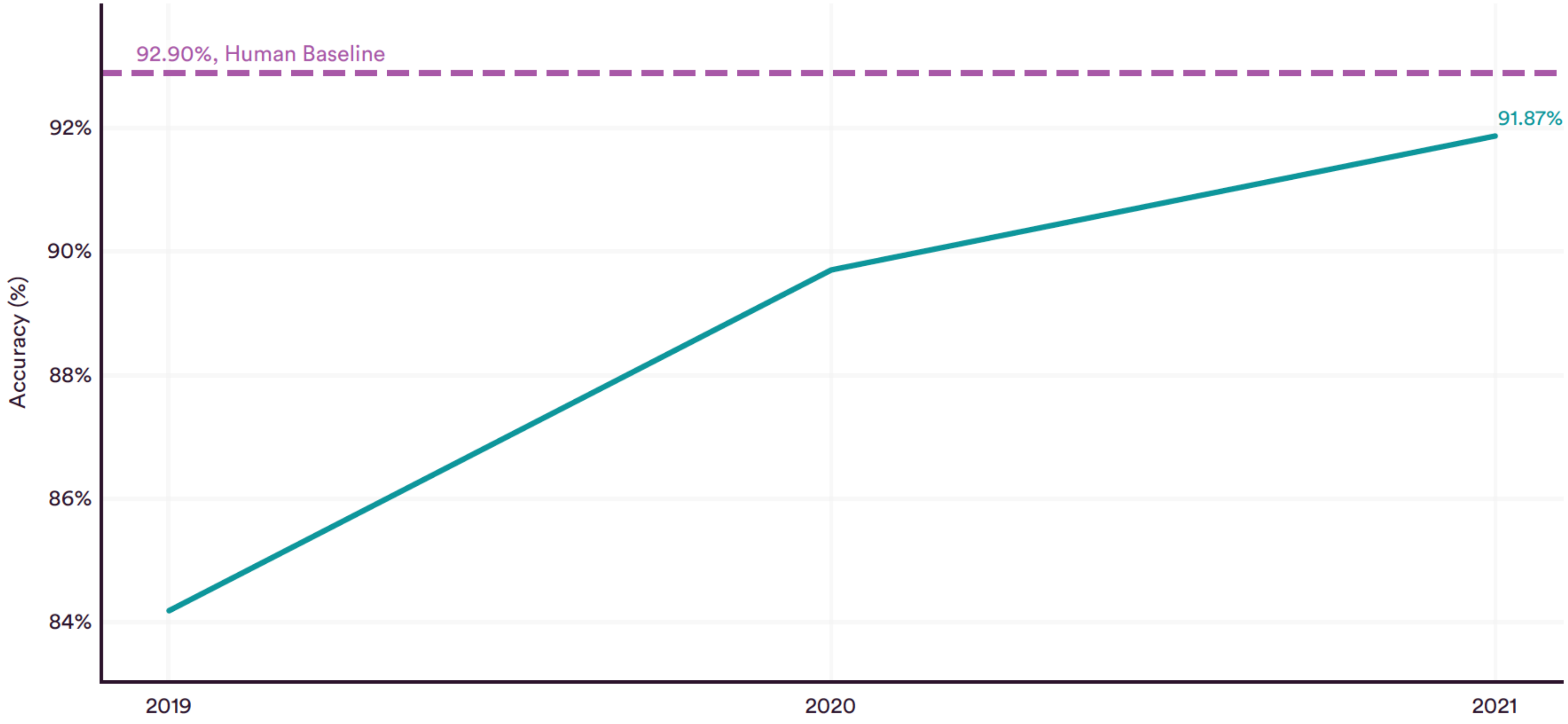


Figure 2.3.12

CHESS SOFTWARE ENGINES: ELO SCORE

Source: Swedish Computer Chess Association, 2021 | Chart: 2022 AI Index Report

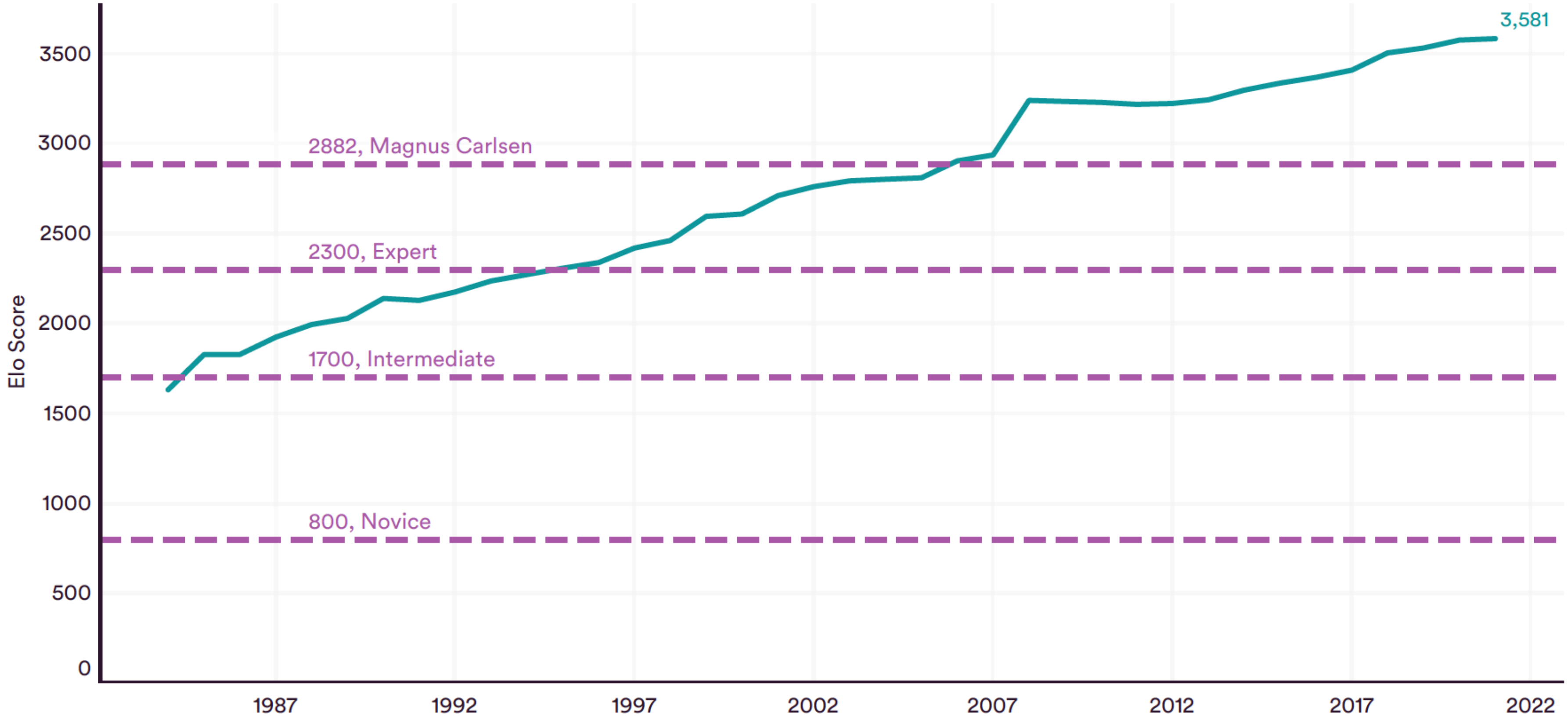


Figure 2.6.4

NUMBER of AI PUBLICATIONS in the WORLD, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

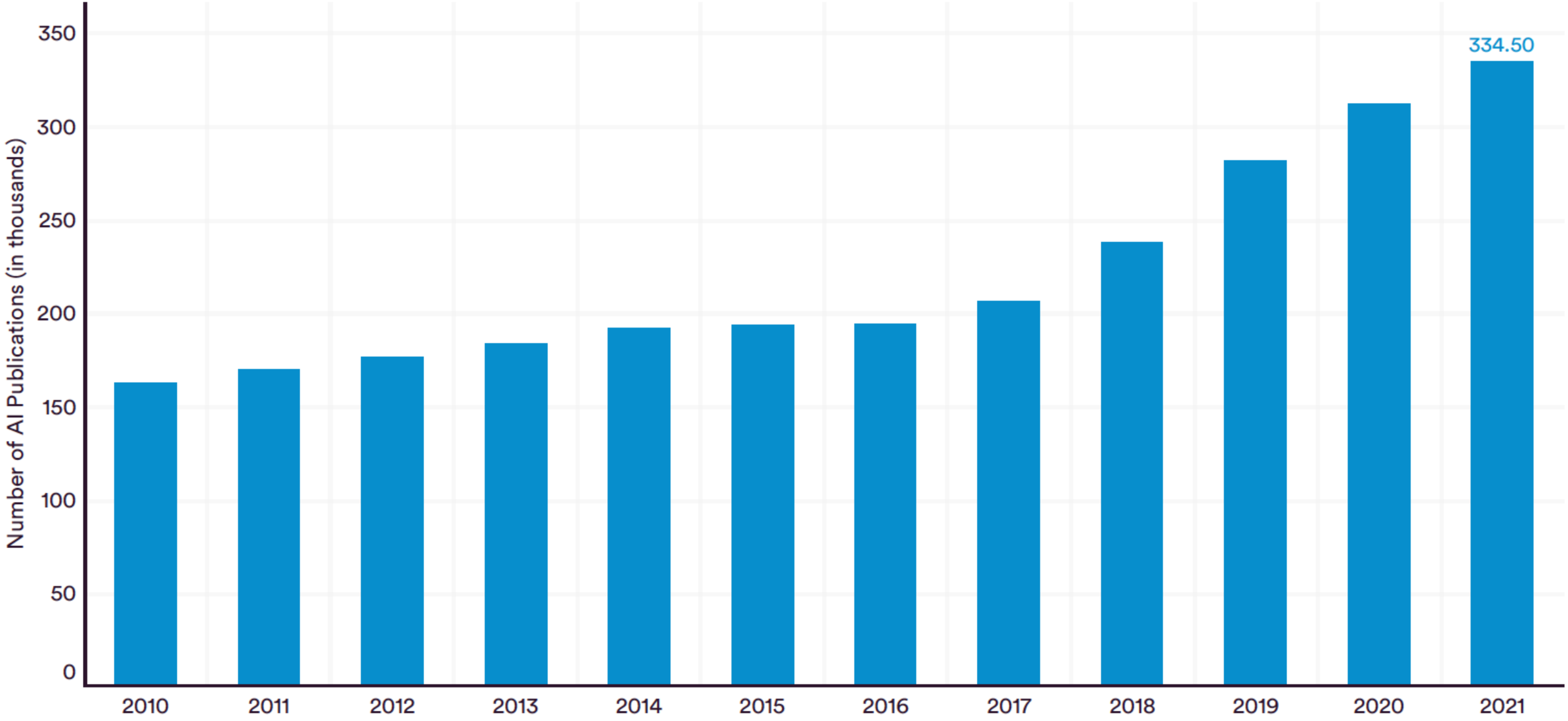


Figure 1.1.1

NUMBER of AI PATENT FILINGS, 2010–21

Source: Center for Security and Emerging Technology, 2021 | Chart: 2022 AI Index Report

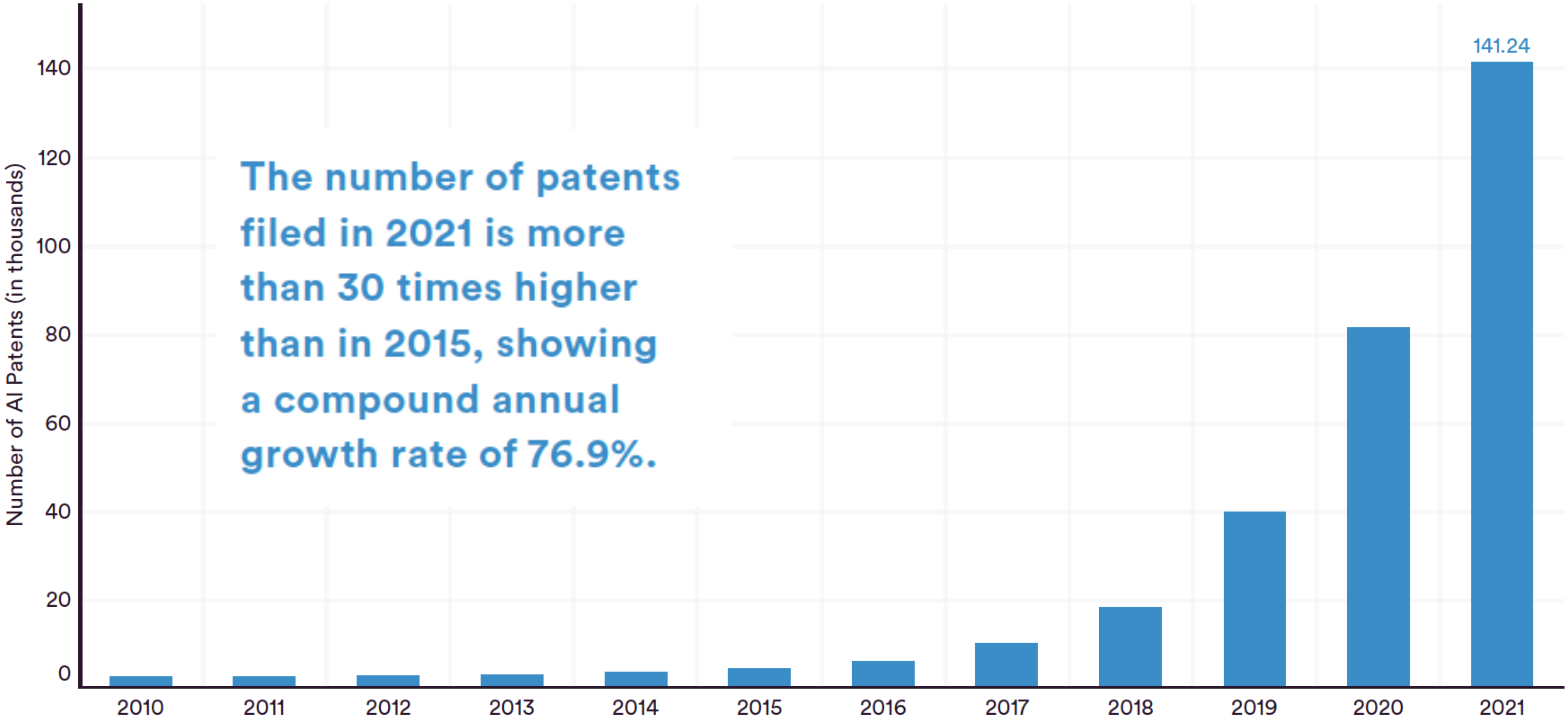
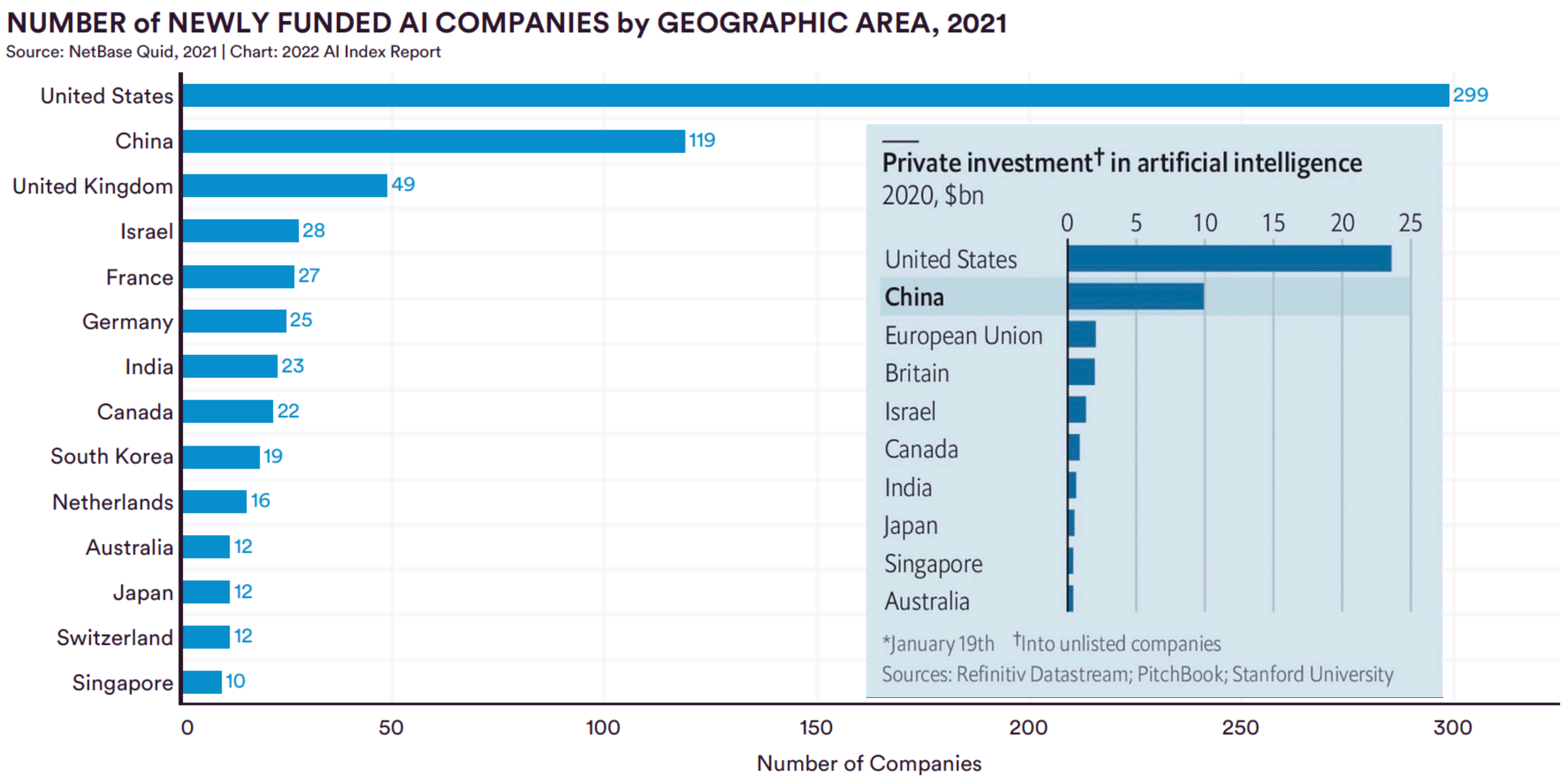


Figure 1.1.22


NUMBER of NEWLY FUNDED AI COMPANIES by GEOGRAPHIC AREA, 2021
Source: NetBase Quid, 2021 | Chart: 2022 AI Index Report
United States 299
China 119
United Kingdom 49
Israel 28
France 27
Germany 25
India 23
Canada 22
South Korea 19
Netherlands 16
Australia 12
Japan 12
Switzerland 12
Singapore 10
0
50
100
150
200
250
300
Number of Companies
Private investment† in artificial intelligence
2020, $bn
0
5
10
15
20
25
United States
China
European Union
Britain
Israel
Canada
India
Japan
Singapore
Australia
*January 19th
†Into unlisted companies
Sources: Refinitiv Datastream; PitchBook; Stanford University

GLOBAL CORPORATE INVESTMENT in AI by INVESTMENT ACTIVITY, 2013–21

Source: NetBase Quid, 2021 | Chart: 2022 AI Index Report

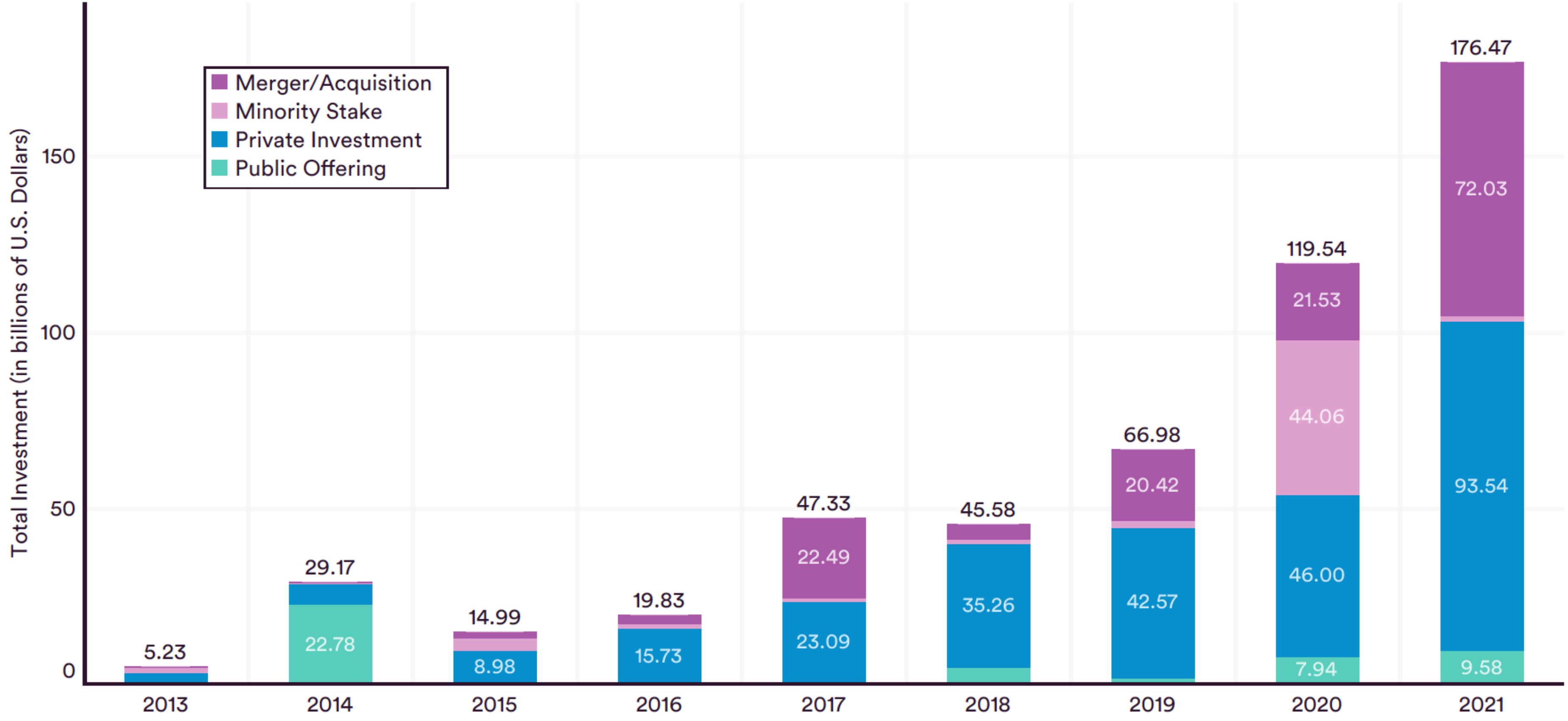


Figure 4.2.1

NUMBER of MENTIONS of AI in LEGISLATIVE PROCEEDINGS in SELECT COUNTRIES, 2021

Source: AI Index, 2021 | Chart: 2022 AI Index Report

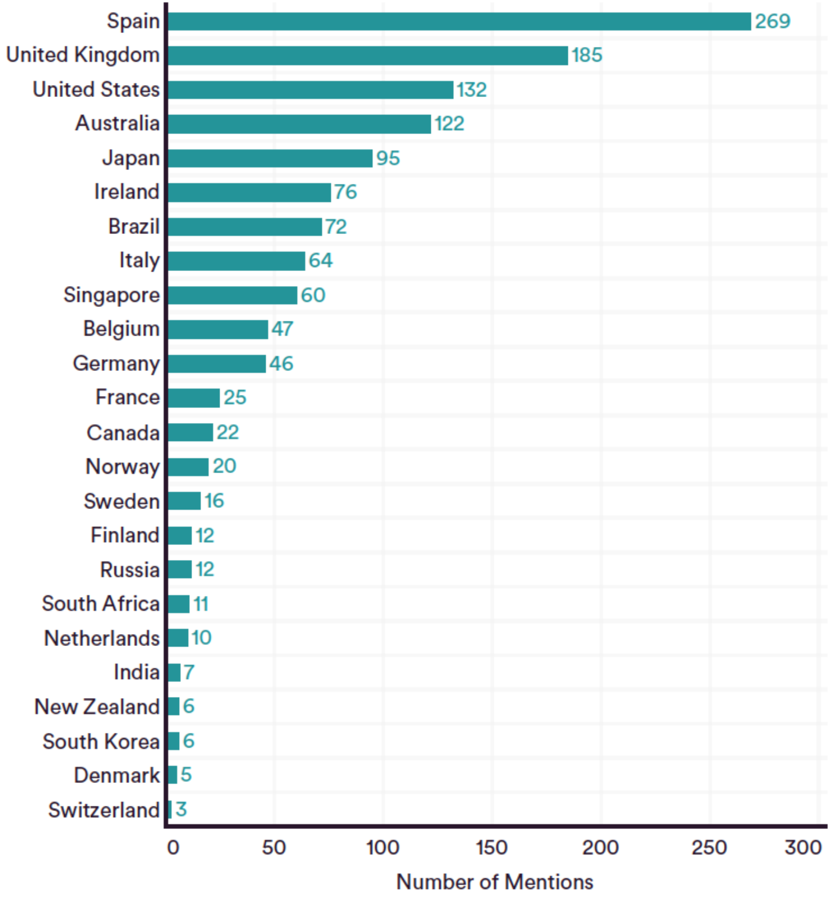


Figure 5.1.10a

NUMBER of MENTIONS of AI in LEGISLATIVE PROCEEDINGS in SELECT COUNTRIES, 2016–2021 (SUM)

Source: AI Index, 2021 | Chart: 2022 AI Index Report

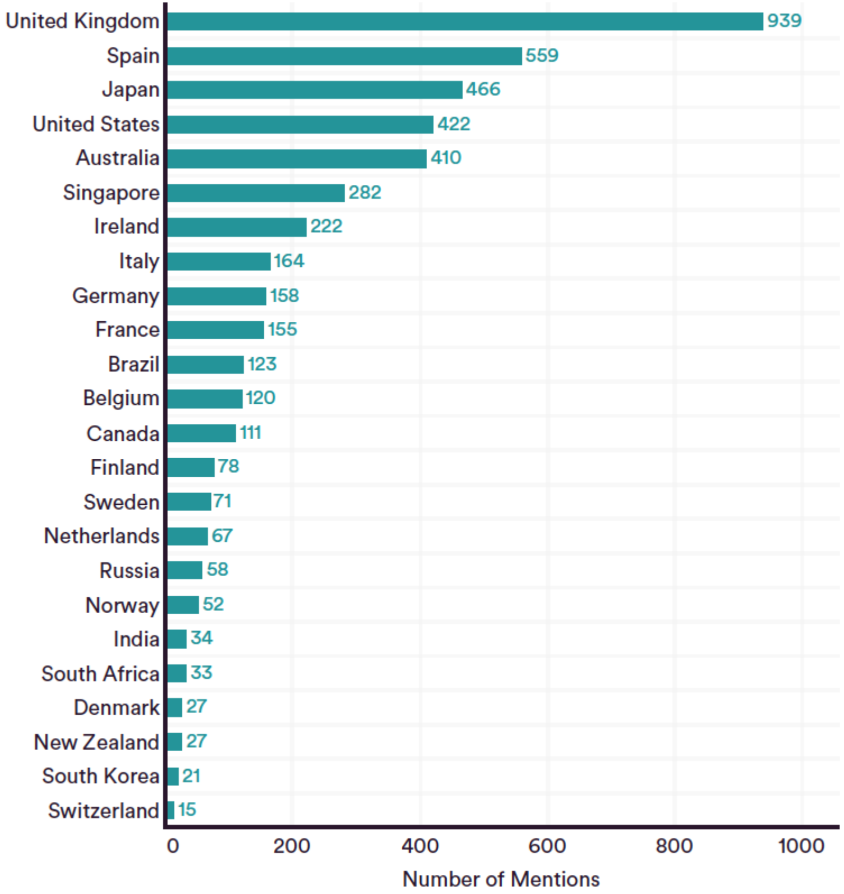
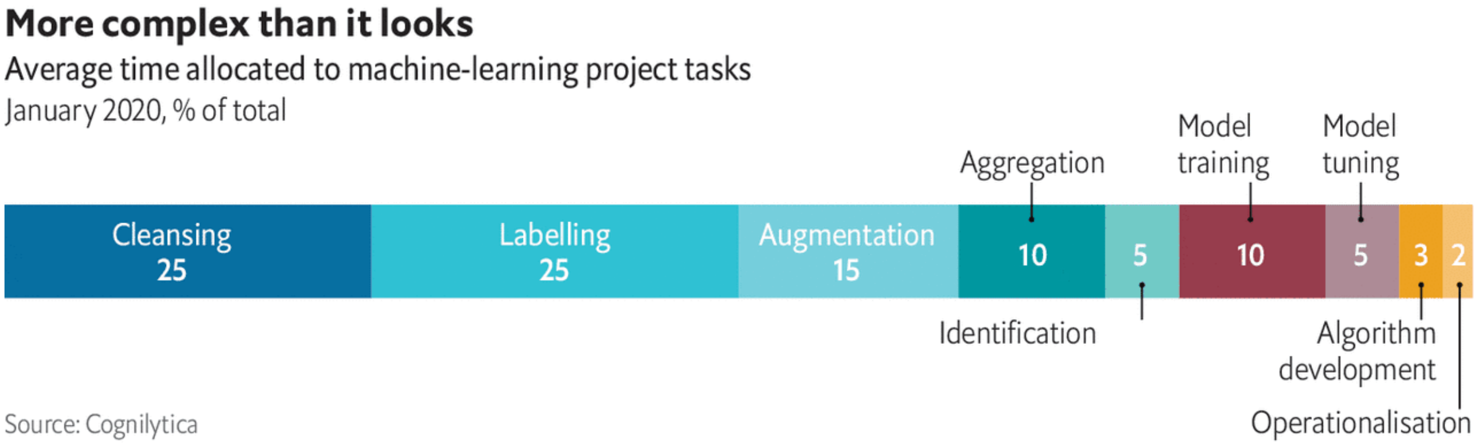


Figure 5.1.10b

ML Projects remain Challenging



- Data issues are one of the most common sticking-points in real AI projects
- Up to 80% of project time is spent on the data and < 20% on ML models!

https://www.economist.com/technology-quarterly/2020/06/11/for-ai-data-are-harder-to-come-by-than-you-think

Datasets Often Contain Inherent Bias

- In 2019 America's National Institute of Standards and Technology tested nearly **200 facial recognition algorithms** and found that many were significantly **less accurate at identifying black faces** than white ones.

- The problem may reflect a preponderance of white faces in their training data. A study from IBM, published in 2019 found that **over 80% of faces** in three widely used training sets **had light skin**.

https://eandt.theiet.org/content/articles/2019/06/microsoft-quietly-takes-down-huge-facial-recognition-database/

Training of Large ML Models is Energy Intensive

- In 2019 researchers at Uni. Amherst estimated that training one version a big language model, could **cost as much as $3m**.

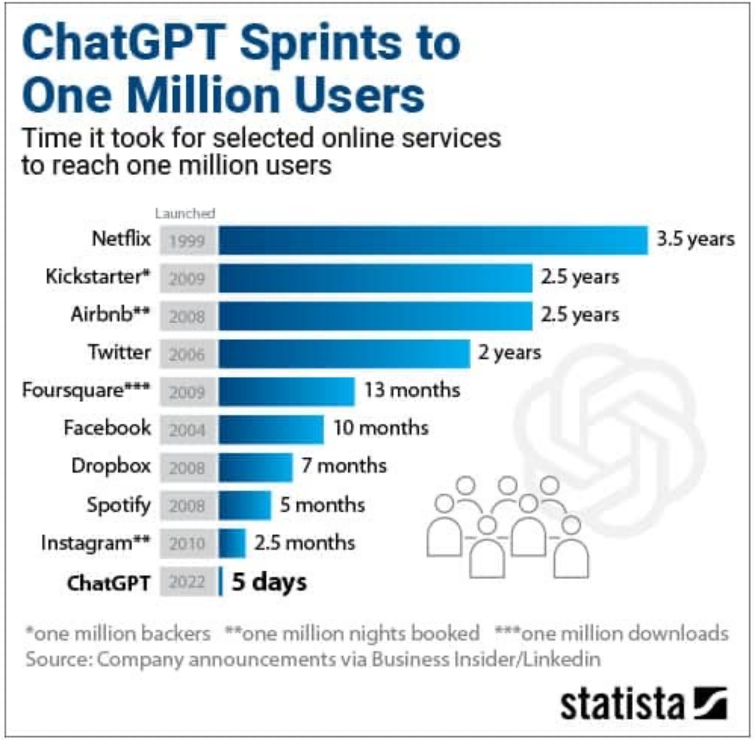
- In 2020 an algorithm capable of learning, through trial and error, how to manipulate the pieces of a Rubik's Cube using a robotic hand consumed about **2.8 gigawatt-hours of electricity** in its training

- Facebook's head of AI, says that one round of training for the biggest models can cost "millions of dollars" in **electricity consumption**.

- Researchers estimate that the training of GPT-3, which ChatGPT is partly based on, consumed 1.3 GWh, and led to emissions of more than **550 tons of carbon dioxide equivalent**

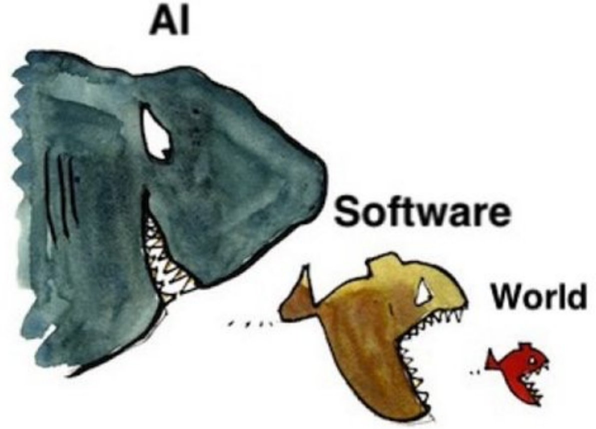
https://www.wired.com/story/ai-great-things-burn-planet/
https://www.wired.com/story/the-generative-ai-search-race-has-a-dirty-secret/

We are at a Turning Point in the History of AI/ML



Bojan Tunguz @tunguz — Machine Learning at Nvidia. Kaggle Quadruple Grandmaster.

This is insane. This is what I've been alluding to for months now. This is an epochal transformative technology that will soon touch - and radically transform - ALL knowledge work. If most of your work involves sitting in front of a computer, you will be disrupted very, very soon.



Brings risks as well as opportunities

President Biden @POTUS · Jun 21

I sat down with experts at the intersection of technology and society who provided a range of perspectives on AI's enormous promise and risks.

In seizing this moment of technological change, we also need to manage the risks.

My Administration is on it.

European Commission @EU_Commission · Sep 14

Mitigating the **risk** of extinction from **AI** should be a global priority.

And Europe should lead the way, building a new global **AI** framework built on three pillars: guardrails, governance and guiding innovation ↓

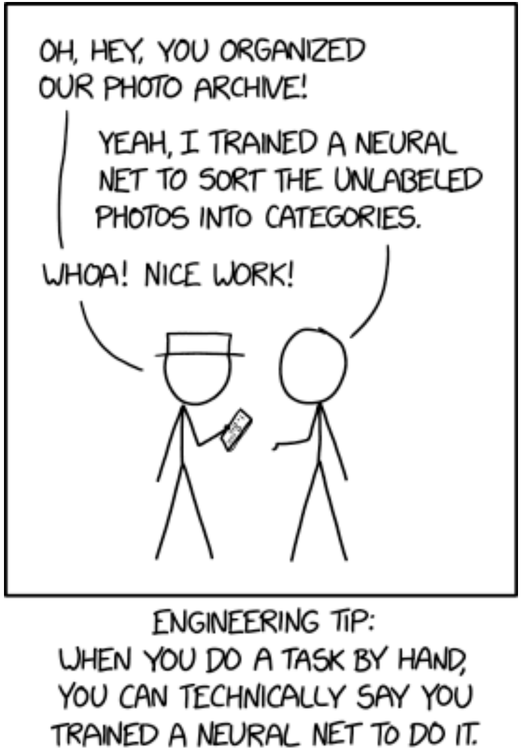
Recap of Today's Lecture

A. Machine Learning in the Context of Artificial Intelligence (AI)

B. The Connectionists: Artificial Neural Networks (ANN's), Machine Learning (ML) and Deep Learning (NN)

C. Three Pillars of Machine Learning: Data, Algorithms & Processing Power

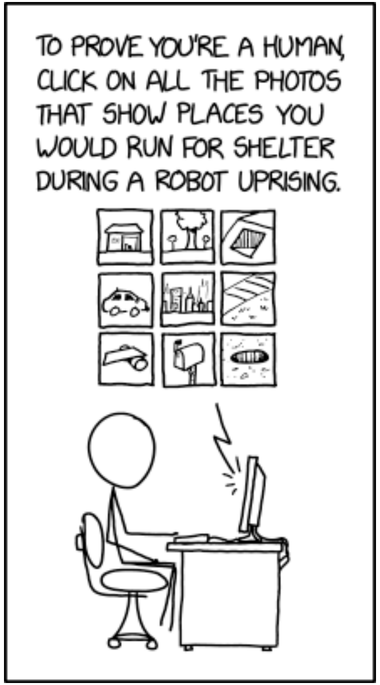
D. Machine Learning Today: Progress, State-of-the-Art & Challenges



https://xkcd.com/2173/

Questions to Think About

- What’s the difference between **AI** and Machine Learning (ML)? What’s the difference between ML and **Deep Learning** (DL)? We had **Neural Nets** in the 1980s, but not DL – why not?
- The average **power consumption** of a typical adult is 100 Watts and the **brain consumes** 20% of this making the **power** of the **brain** 20W, about the same as a light-bulb. How do you think we could get modern machine learning algorithms to work at this level of efficiency?
- Deep Learning (DL) can learn from large amounts of data to **automate cognitive tasks** normally requiring a human. For example we can train a DL algorithm to screen CVs for selecting job candidates. **Can you think of any negative consequences of this**?



https://xkcd.com/2228/

Thank you! Any Questions?

Dr. Donnacha Daly

Head of Bachelor Program in Artificial Intelligence & Machine Learning
Member of the Algorithmic Business (ABIZ) Research Lab

Lucerne University of Applied Sciences and Arts
School of Computer Science & Information Technology

donnacha.daly@hslu.ch | hslu.ch/bachelor-ai | www.abiz.ch