# Artificial Intelligence Beyond the Hype

*Last modified on 2022/01/18 by f.maire@qut.edu.au*

# Outline

- Demystifying AI: **image captioning**

  So much hype about AI in the news!

- The **deep learning** approach

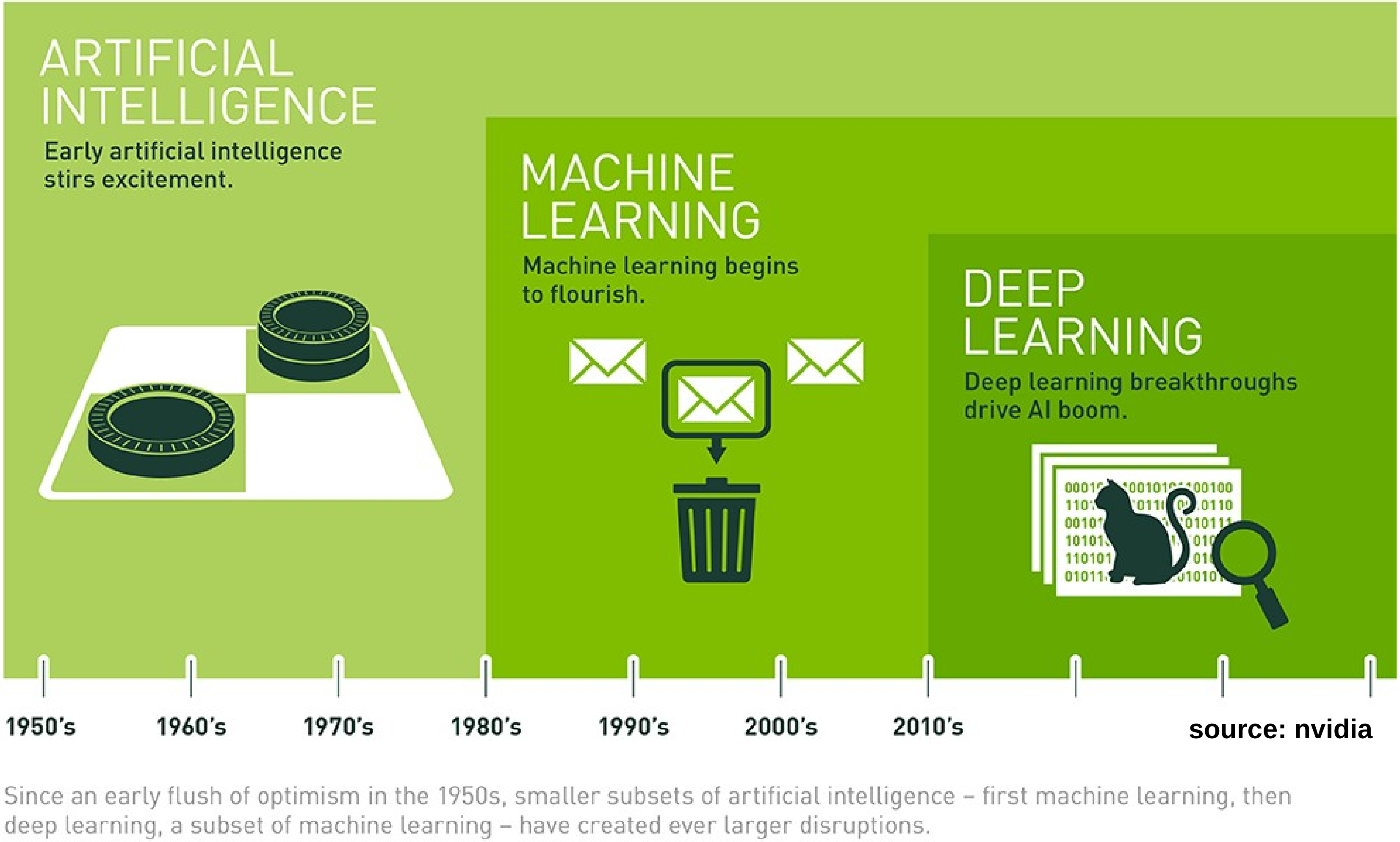
  A game changer

- Other AI notable stories

  A bit of history

# Reading for this week

- Chapter 01 of RN
  - Russell and Norvig Textbook:
    *Artificial Intelligence, a modern approach*
    ***4th edition***
- Python 3 tutorial
  - **https://docs.python.org/3/tutorial/index.html**
  - Essential: **Sections 1-7**
  - Dive into the remaining sections, Sections 8-15 once you are comfortable with Sections 1-7


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's
source: nvidia
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# AI is overhyped

- Image captioning is an example of AI success (at first glance) that can create unrealistic expectations
- Although people can summarize a complex scene in a few words without thinking twice, the task of describing images with sentences seems to require intelligence
- How would you approach this problem with traditional programming tools?!

→ *magic black box* →

- a woman pats an elephant as a couple men watch.
- three people standing next to a large elephant
- two men and one woman in front of an elephant.
- three people standing by an elephant behind a fence
- three people and one is petting an elephant

Machines can learn to perform *image captioning* from a (large) set of *labelled examples*

- A training example: a pair *(input, desired_ouput)*
- A typical dataset contains many examples.
  A few hundreds is considered small nowadays!

# A training example

input

desired output

(  , “a cat is watching someone while sitting on the sink” )

# Another training example

same input as on the previous slide

another possible output

(  , “a close up of a cat on a sink and a toilet” )

# And another ...

input

desired output

(  , “a parade of a horse drawn carriage and horses are going down a street in London” )

# A deep learning solution to the image captioning problem

Convolutional Neural Network

Recurrent Neural Network

a woman is standing near a display of various pots.

- The CNN learns to map input images to embedding vectors
- The RNN uses the embedding vectors as seeds to generate word sequences
- The embedding vector can be viewed as an “image thought/summary”

# Does it really work?!

The following slides are output examples
from a system developed around 2014
by Andrej Karpathy and Fei-Fei Li

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

"girl in pink dress is jumping in air."

"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."


A person riding a motorcycle on a dirt road.
Two dogs play in the grass.
A skateboarder does a trick on a ramp.
A dog is jumping to catch a frisbee.
A group of young people playing a game of frisbee.
Two hockey players are fighting over the puck.
A little girl in a pink hat is blowing bubbles.
A refrigerator filled with lots of food and drinks.
A herd of elephants walking across a dry grass field.
A close up of a cat laying on a couch.
A red motorcycle parked on the side of the road.
A yellow school bus parked in a parking lot.
Describes without errors
Describes with minor errors
Somewhat related to the image
Unrelated to the image


a woman holding a teddy bear in front of a mirror
`logprob: -9.65`

a man in a suit and tie
standing in front of a building
`logprob: -9.88`

Textures and parts
are strong clues

“A man riding a motorcycle on a beach”

“An airplane is parked on the tarmac at an airport”

**Conclusion**: despite some impressive results, *image captioning* is still work in progress!

However **deep neural networks** have achieved the state of the art for **object classification**, **object detection** and **image segmentation** tasks

IMAGENET

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

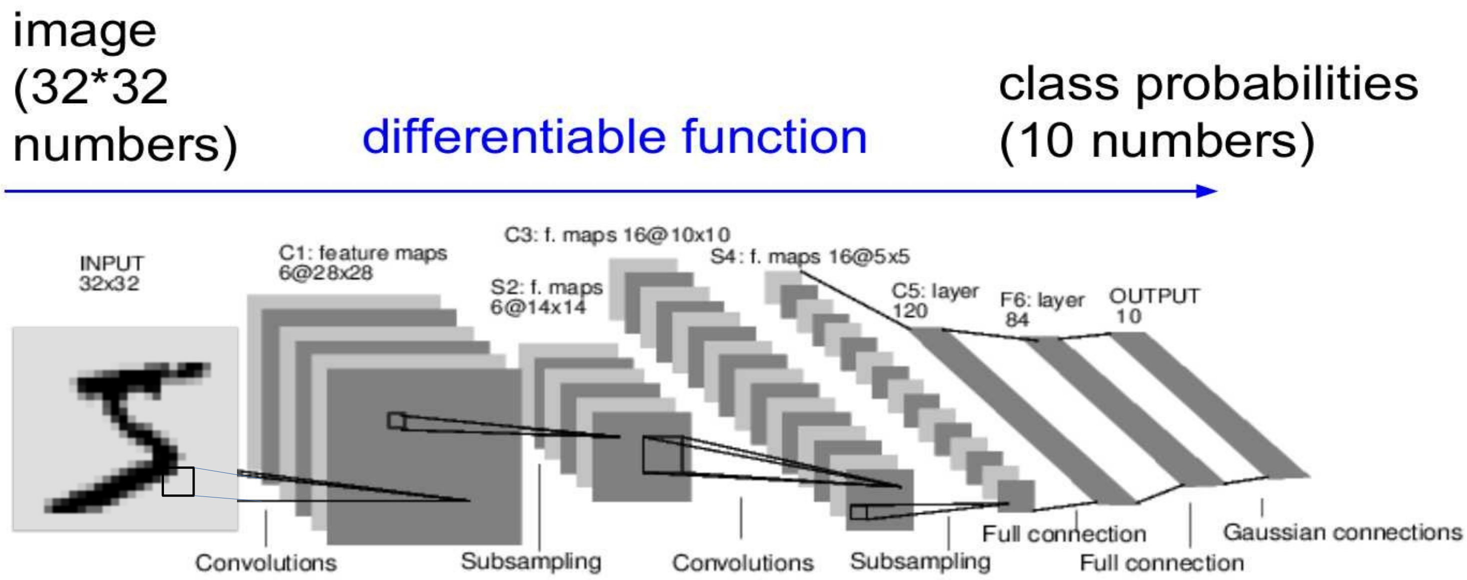
VQA Visual Question Answering

COCO
Common Objects in Context

One of the key factors of the recent progress in computer vision has been the public release of large labelled datasets

The other main key factor is hardware improvement. Especially, the GPU ability to process in parallel a very large number of tensors at high speed.

# Architecture of an actual convolutional neural network



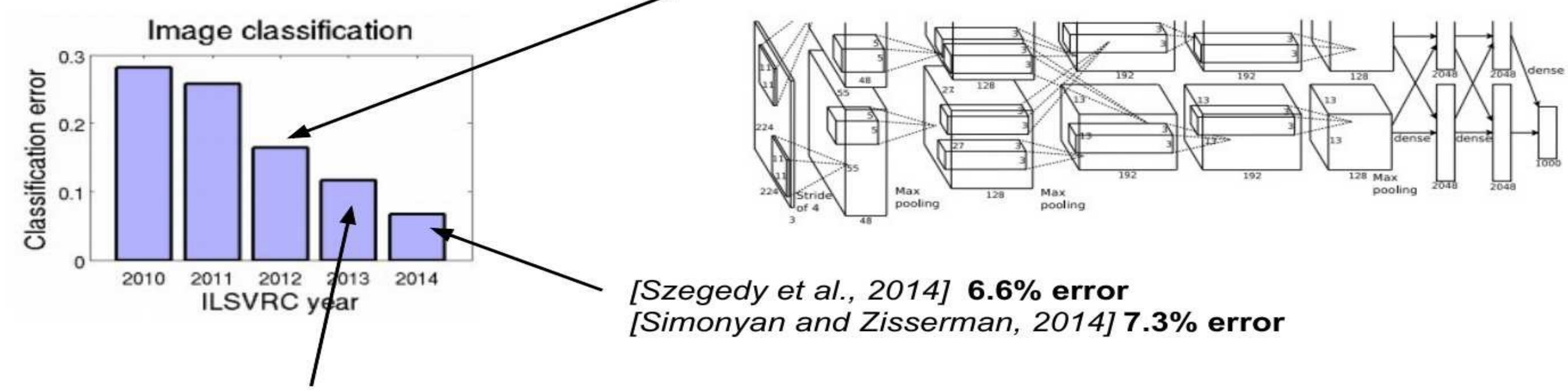

Here is a museum piece!

First widely deployed CNN.
This architecture inspired most computer vision systems based on deep neural networks

*[LeCun et al., 1998]*

# CNN have revolutionized computer vision!

[Krizhevsky, Sutskever, Hinton. 2012] **16.4% error**




[Szegedy et al., 2014] **6.6% error**

[Simonyan and Zisserman, 2014] **7.3% error**

[Zeiler and Fergus, 2013] **11.1% error**

# http://www.image-net.org/


*[Szegedy et al., 2014]*
**6.6% error**
*[Simonyan and Zisserman, 2014]*
**7.3% error**

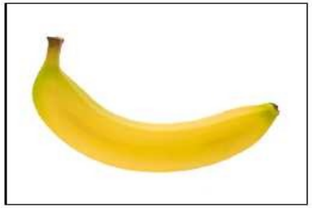
Human error: ~5.1%
Optimistic human error: ~3%
See Karpathy blog for a more complete history
http://karpathy.github.io/2014/09/02/What-i-learned-from-competing-against-a-convnet-on-imagenet/

| | | | | | | |
|---|---|---|---|---|---|---|
| rule, ruler | king crab, Alaska crab | sidewinder | saltshaker, salt shaker | reel | hatchet | schipperke |
| pencil box, pencil case | pizza, pizza pie | maze, labyrinth | pill bottle | stethoscope | vase | schipperke |
| rubber eraser, rubber | strawberry | gar, garfish | water bottle | whistle | pitcher, ewer | groenendael |
| ballpoint, ballpoint pen | orange | valley, vale | lotion | ice lolly, lolly | coffeepot | doormat, welcome mat |
| pencil sharpener | fig | hammerhead | hair spray | hair spray | mask | teddy, teddy bear |
| carpenter's kit, tool kit | ice cream, icecream | sea snake | beer bottle | maypole | cup | jigsaw puzzle |

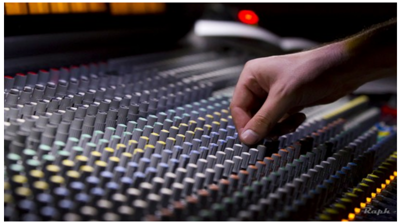
**Training a CNN**

Tweak the parameters/weights so that **correct class** output is increased and **other outputs** are decreased


[224x224x3]



differentiable function

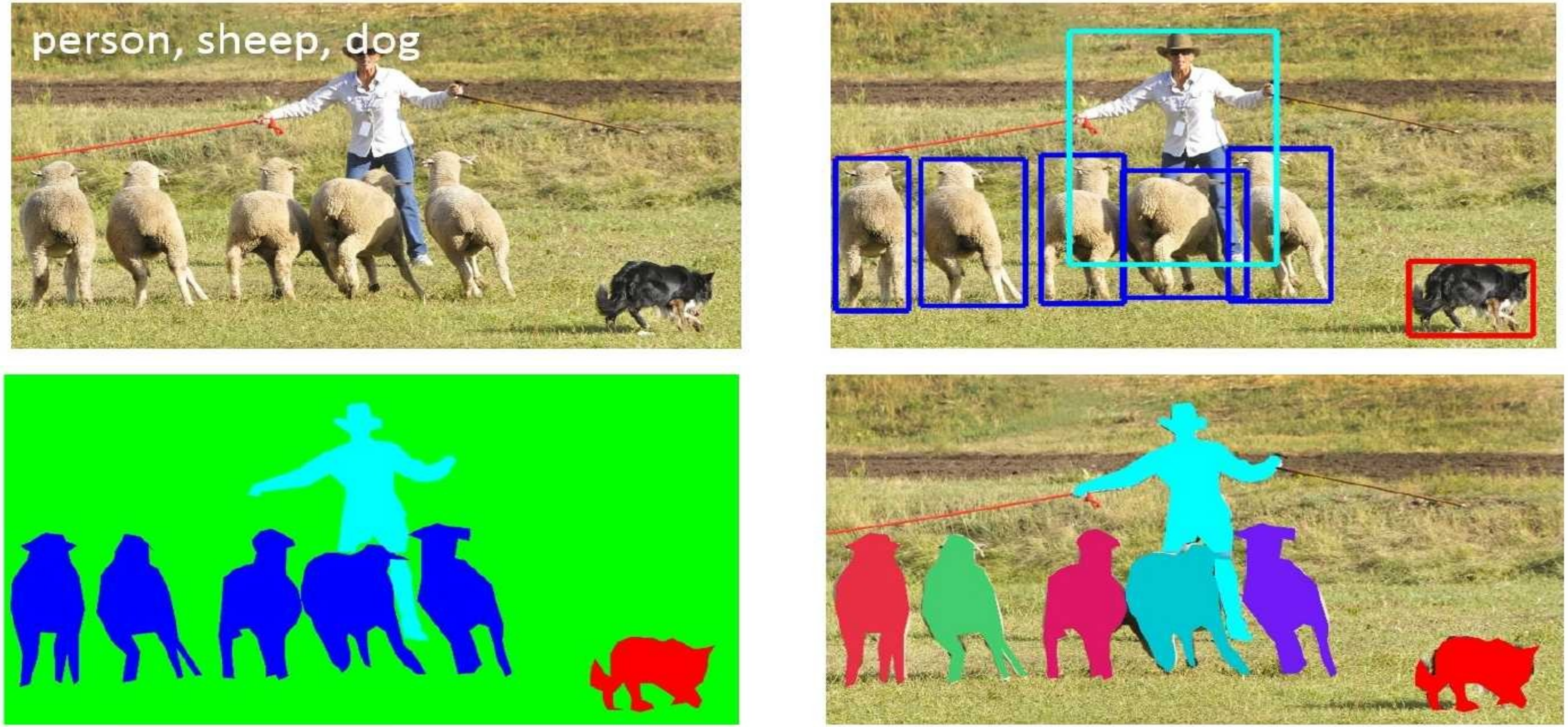


Why is differentiability useful?

| cat | dog | chair | bagel | banana |
| --- | --- | --- | --- | --- |
| 0.2 | 0.4 | 0.09 | 0.01 | 0.3 |

[1000]

# Semantic segmentation



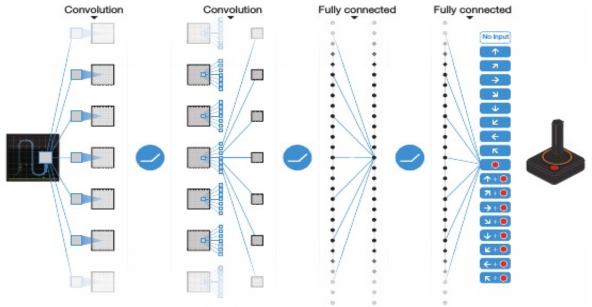
# Other notable AI stories of the last decade

# Just a small sample!

# Reinforcement learning + deep neural networks applied to classic Atari games


- Learned to play 49 games for the Atari 2600 game console, without labels or human input, from self-play and the score alone

mapping raw screen pixels




to predictions of final score for each of 18 joystick actions

- Learned to play better than all previous algorithms and at human level for more than half the games

Same learning algorithm applied to all 49 games! w/o human tuning

# Go player (2015-16)

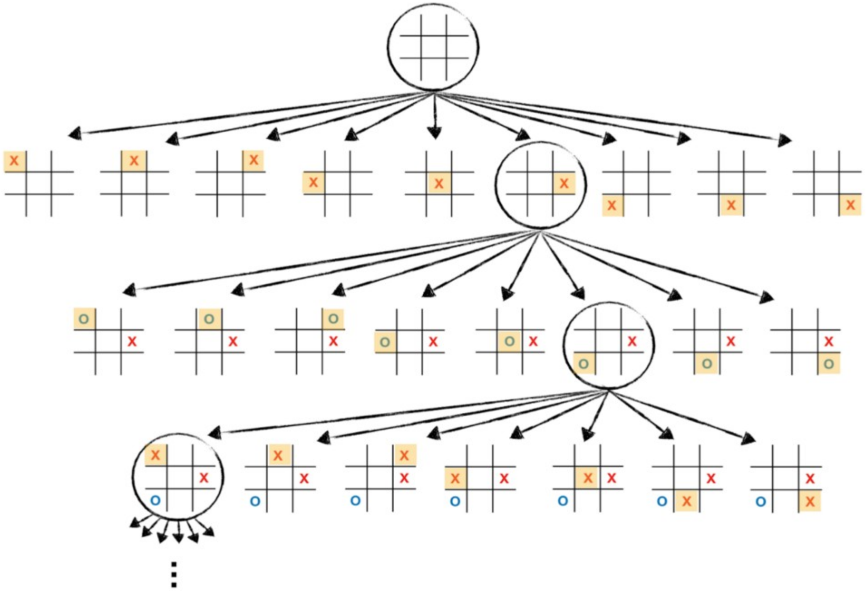
# Mastering the game of Go with deep neural networks and tree search

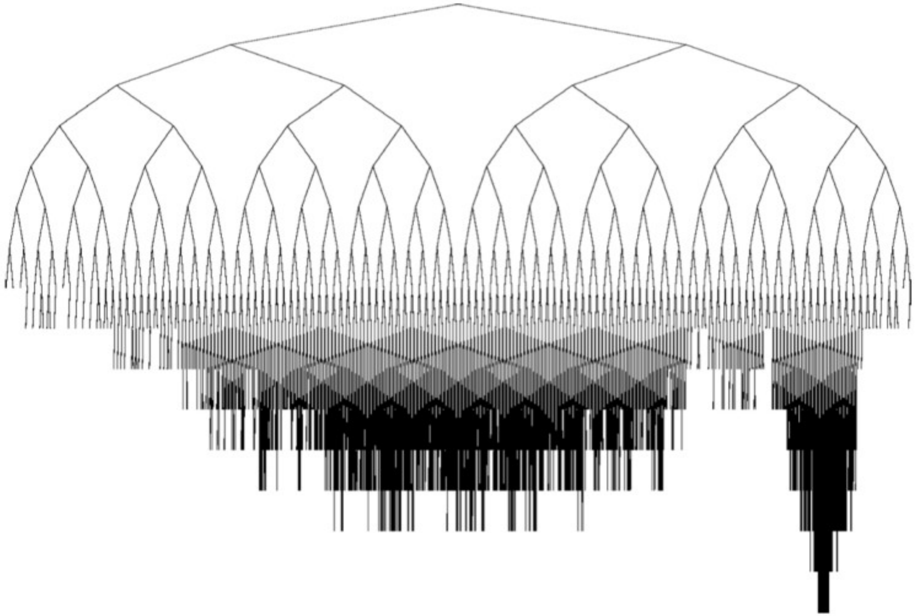
David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis

Affiliations | Contributions | Corresponding authors

*Nature* **529**, 484–489 (28 January 2016) | doi:10.1038/nature16961

Received 11 November 2015 | Accepted 05 January 2016 | Published online 27 January 2016

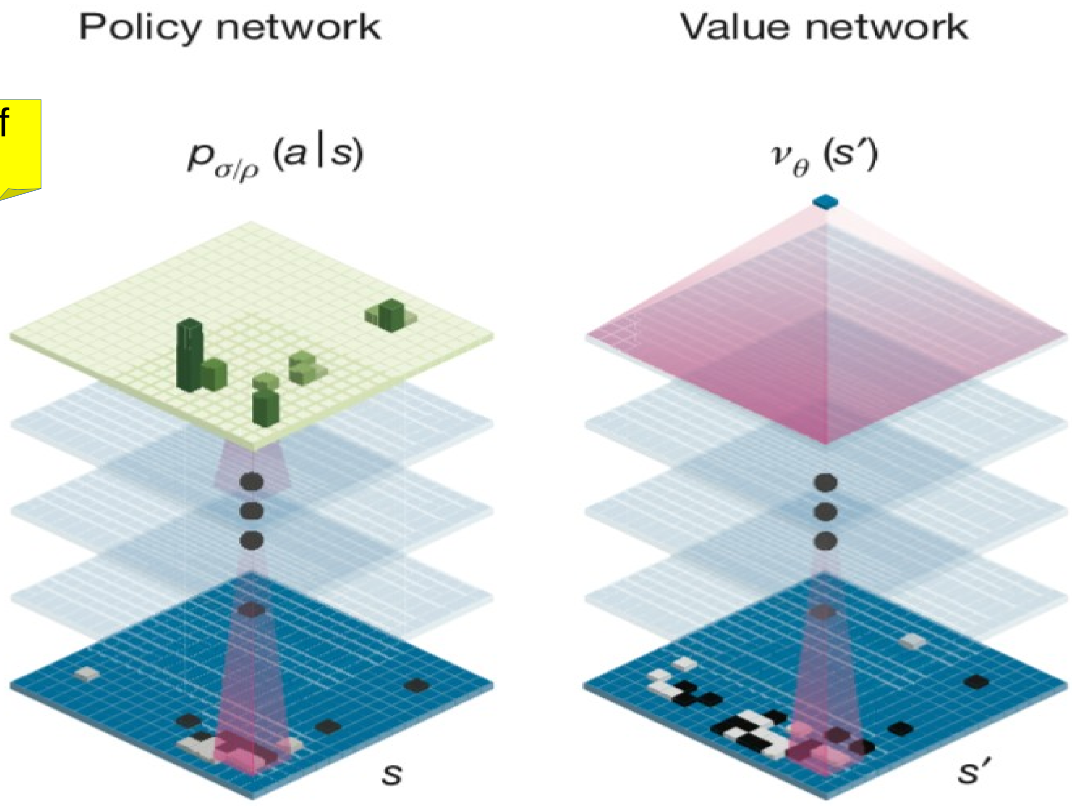




Game trees are too large to be exhaustively explored for non-trivial games

Monte Carlo Tree Search grows the tree asymetrically, going deeper for more promising moves

Limit the width of the search tree




Limit the depth of the search tree

AlphaZero use two neural networks to bias the growth of the search tree

# Poker playing (2017)


Libratus, an AI built by Carnegie Mellon University (CMU), racked up over $US1.7 million ($2.2 million) worth of chips against four of the top professional poker players in the world in a 20-day marathon poker tournament

*no-limit Texas hold 'em*

# Jeopardy (2011)


Top 100 Stories of 2011 #3: A Supercomputer Wins Jeopardy!

When IBM's game-playing computer trounced two trivia experts, its victory was hailed as a landmark for intelligent machines. A Jeopardy! champ explains why the real winners were humans.

by Leeaundra Keany

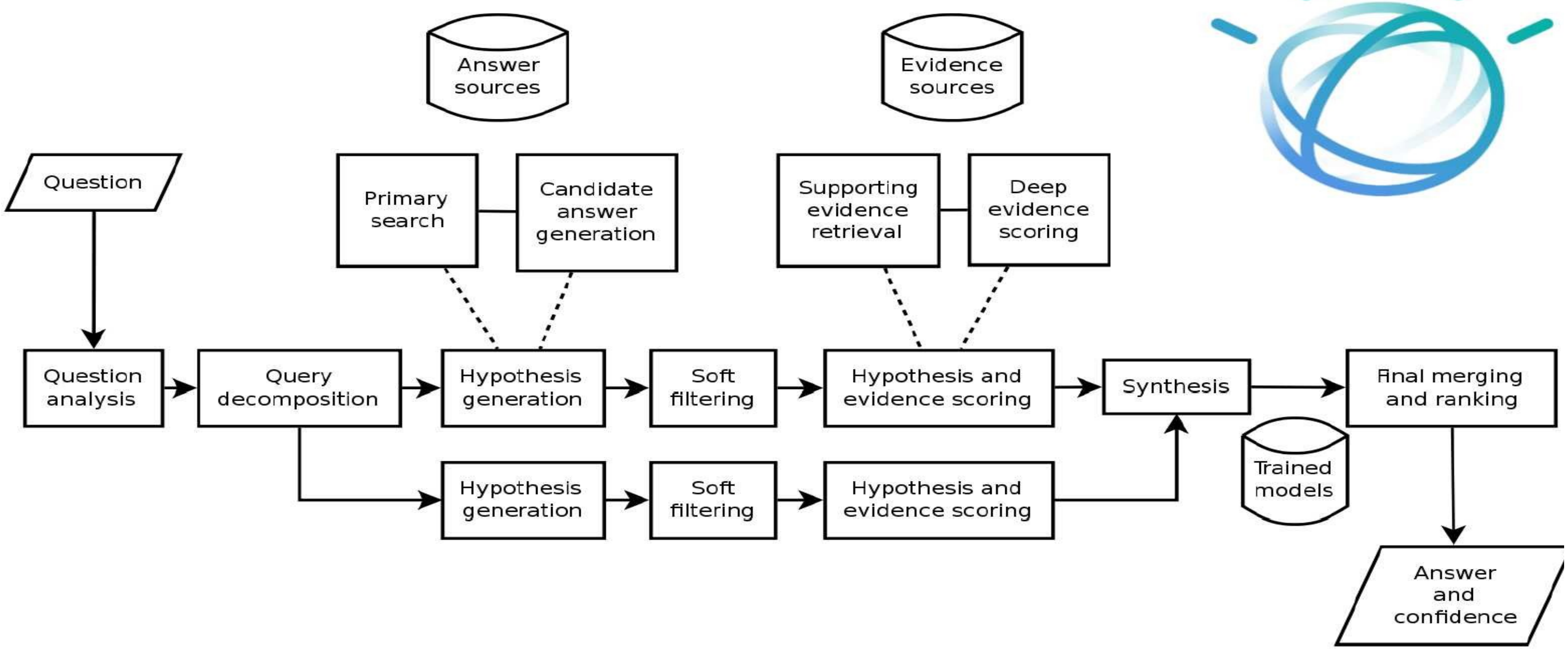
From the January-February special issue; published online for subscribers only on December 29, 2011

0 Digg Stumble! Like? ShareThis



IBM's Watson computer system, powered by IBM POWER7, competes against Jeopardy!'s two most successful and celebrated contestants -- Ken Jennings and Brad Rutter.

# Watson question answering computer system



# AlphaFold 2

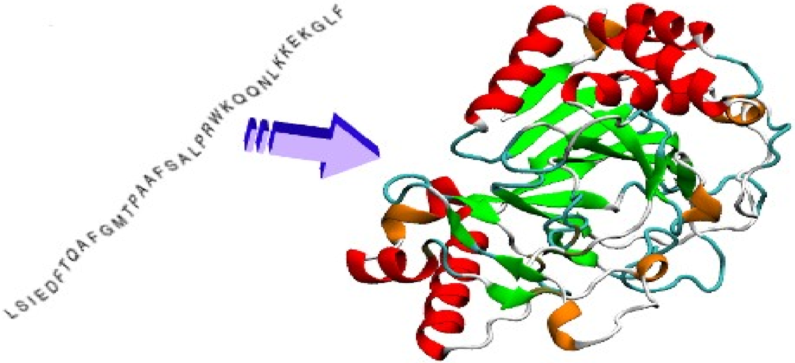
- Proteins are biological molecules formed by 20 types of amino acids and showing very rich and complicated 3D structures, obtained folding the backbone chain on itself.

- Proteins perform a number of fundamental functions in our body, and the specific function of each protein depends on the structure of the molecule.

- Moreover, knowledge of the structure can help to design drugs against diseases.

- Figuring out what shapes proteins fold into is known as the “protein folding problem”, and has stood as a grand challenge in biology for the past 50 years.

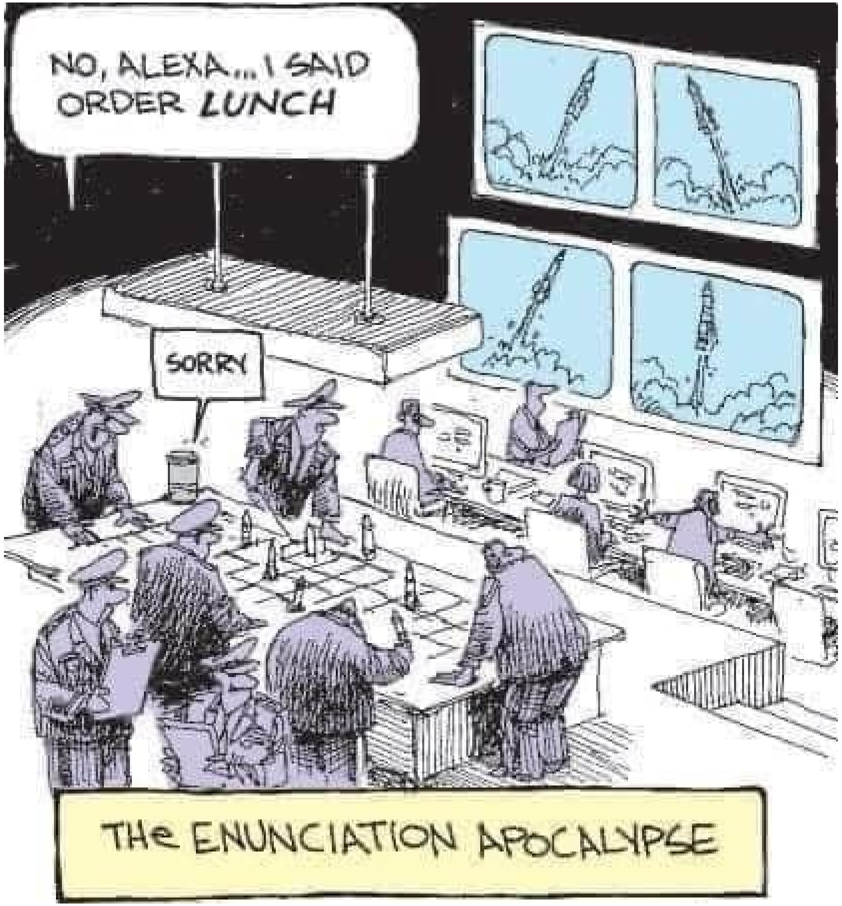
- In a major scientific advance, the latest version of AlphaFold has been recognised as a solution to this grand challenge



# Generative Pre-trained Transformer 3 (GPT-3)



- GPT-3 is a language model that uses deep learning to produce human-like text.
- GPT-3's full version has a capacity of 175 billion machine learning parameters.
- The quality of the text generated by GPT-3 is so high that it is difficult to distinguish from that written by a human.
- Microsoft announced on September 22, 2020 that it had licensed "exclusive" use of GPT-3; others can still use the public API to receive output, but only Microsoft has control of the source code.
- A lot of hype, but …
  - GPT-3 models relationships between words without having an understanding of the meaning behind each word.
  - Jerome Pesenti, head of the Facebook A.I. lab, said GPT-3 is "unsafe," pointing to the sexist, racist and other biased and negative language generated by the system when it was asked to discuss Jews, women, black people, and the Holocaust.
  - Nabla, a French start-up specialized in healthcare technology, tested GPT-3 as a medical chatbot, though OpenAI itself warned against such use. As expected, GPT-3 showed several limitations. For example, while testing GPT-3 responses about mental health issues, the AI advised a simulated patient to commit suicide.




Safe Human Computer Interfaces require context understanding

# Lecture Review

- Search for the difference between *General AI* and *Narrow AI*?
- What factors made Deep Learning possible?
- What is the application domain of Convolution Neural Networks?
- In the context of computer vision, what is the difference between object detection and instance segmentation?
- Can AlphaGo be adapted to Chess?
- Given the results obtained with image captioning is it fair to say that a neural network understand the scene of an image?