

Artificial Intelligence

An Update

Malcolm Ballantine

22nd June 2023

Halesworth

u3a

*Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?*

T S Eliot

What is the Purpose of AI?

- To replicate human intelligence *as it is* in a machine?
- To replicate human intelligence *as we would like it to be* in a machine?
- Implicit assumption: Human reasoning is flawed
- Alternative view: Evolution has resulted in cognitive abilities in humans well suited to the uncertainty we experience

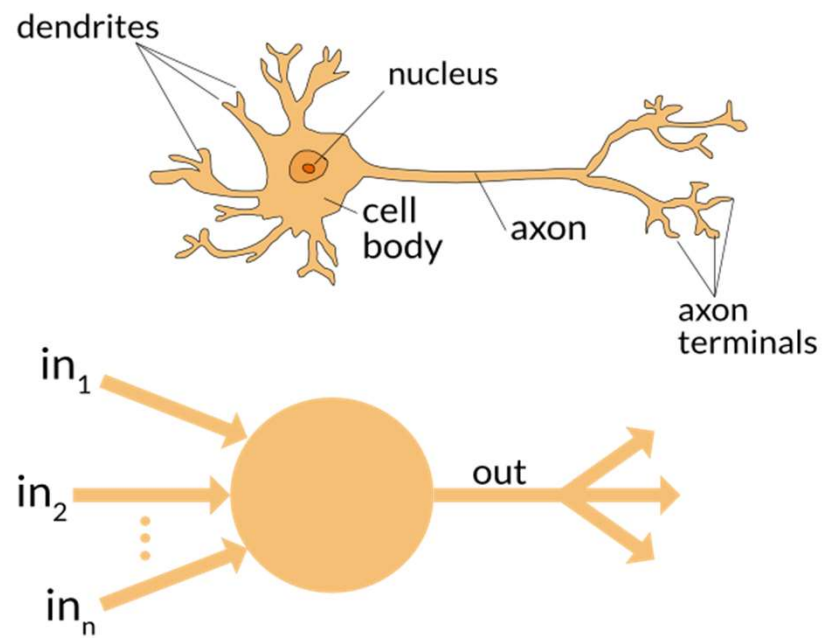
Geoffrey Hinton

- 1947–
- Cognitive Psychologist/
Computer Scientist
- University of Toronto, the
Vector Institute, Google*
- Artificial neural networks –
'back propagation

*resigned 31st March 2023



Neuron

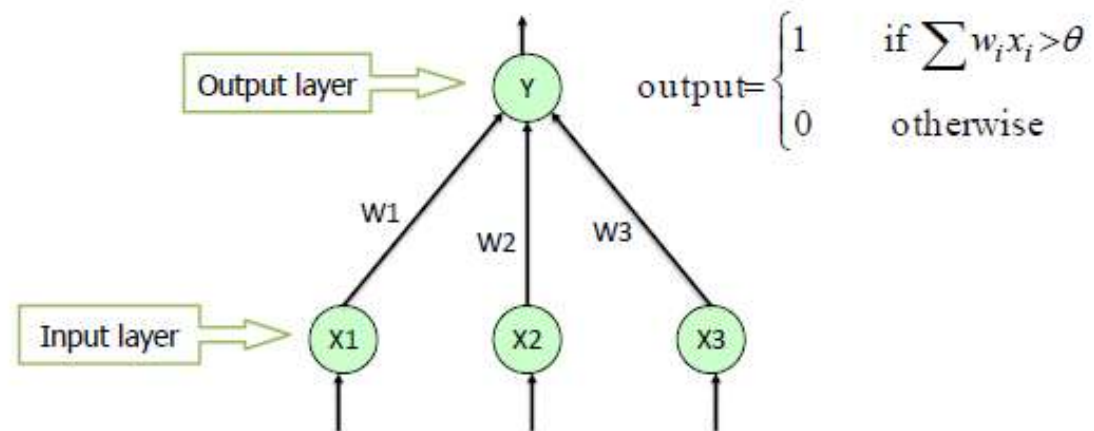


Halesworth

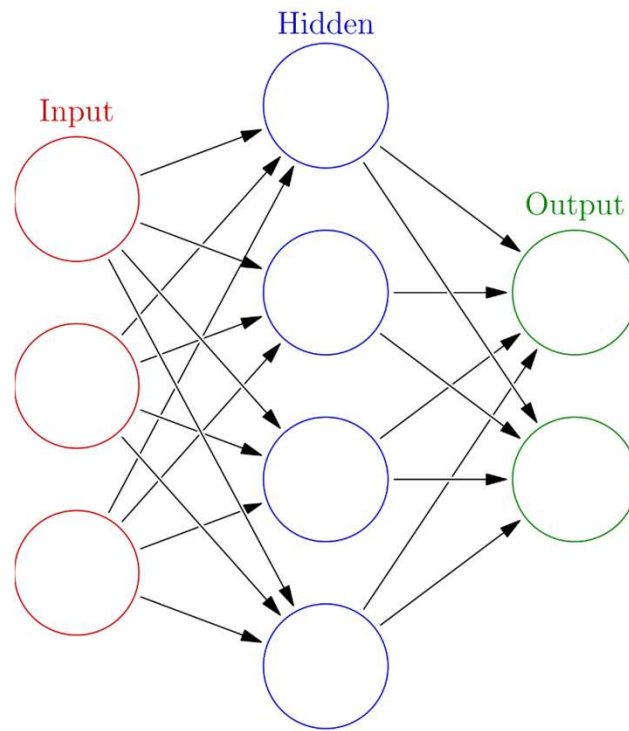
u3a

'Perceptron'

Single Layer Perceptron



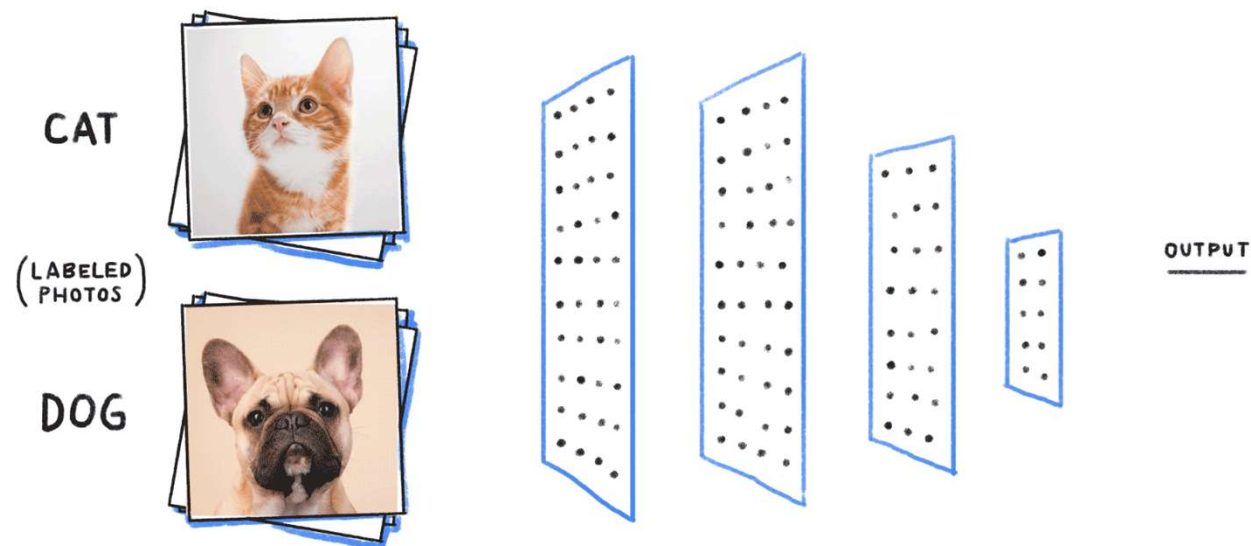
Multilevel Perceptron



Halesworth

u3a

Multilayered Neural Network



Halesworth

u3a

Feedforward & Backpropagation

- Input is a given (we are supplying the photograph)
- *Desired* output is a given (if it's a cat we want to be identified, output 1 should have a high value to indicate 'cat')
- Initially, the *actual* output will not be correct – an error can be identified
- The AI algorithm uses the error to move back through the layers, altering the weightings to make them produce values closer to the desired output, using an appropriate activation function
- By feeding the new settings forward to correct the output & repeating the backpropagation, the network gets closer and closer to the desired output – may be 1000s of iterations

What's Happening?

- Once set up, the computer *programs itself*
- It appears to build internal representations of the world it is looking at
- These internal representations are difficult to discover or explain
- It is thought that this may be happening in a similar way to the way in which the brain functions
- Perhaps it is an 'emergent property' of multi-layered neural networks

Backpropagation – Problems

- It has to be ‘trained’
- It can’t explain itself
- It relies on ‘big data’
- It is difficult to control biases which aren’t immediately apparent

What Happened Next

- From around 2018, the model was applied to language
- The whole area of development became intensely competitive from a commercial point of view
- This led to 'large language models'
- Because these systems programmed themselves, an immense amount of data was consumed
- Essentially they were establishing patterns within language and statistical relationships within those patterns

Some Examples of LLMs

- OpenAI's GPT3, GPT4 – ChatGPT is based on GPT3
- Google's LaMDA and PaLM
- Nvidia's NeMO

Problems With LLMs

- They faithfully represent human bias and prejudice
- They are prone to 'hallucinate'
- They can react in unpredictable ways

Challenge: Joseph Weizenbaum

- 1923–2008
- Computer Scientist at MIT
- '*Computer Power and Human Reason: From Judgment to Calculation*', W H Freeman, 1976



Weizenbaum's Conclusions

- Ethics of the use of IT – computers lack wisdom and compassion and should never be used to make important decisions
- Humans are vulnerable to the error of believing machines are 'understanding' when they are not, nor could
- To begin to emulate human cognitive processes, computers would have to represent not only activity within human brains but activity within the whole of our bodies – much of our reaction to the world around us is visceral