On the feasibility of Intelligent AI

To what extent can mathematical models simulate human intelligence?

Word Count: 3719

February 13, 2024

Contents

Abstract

After the release of the popular NLP model, ChatGPT, the attention on AI has brought upon the possibility of AI superseding humans. This naturally prompts the question: how *intelligent* is AI right now? The answer can be found by analyzing the capabilities of AI. Currently, AI is only able to replicate specific functions of intelligence, rather than simulating it. This leads one to the conclusion that mathematical models cannot accurately replicate human intelligence, because mathematicians do not understand the full scope of brain function that makes intelligence possible.

Introduction

Artificial Intelligence (AI) is a blanket term that refers to the ability to replicate the functions of intelligent beings. Its usage has become more normalized after or during COVID, as people adjusted to a ”reduced human element” (Collins et al., 2021, p.1). A simple way AI has made its way into society is at the ATM, where the machine can recognize handwriting on cheques, depositing them faster. From this example, it is implied that AI has been able to replicate how humans read. This prompts the question: can AI replicate more human functions? This paper explores this idea with the research question:

To what extent can mathematical models simulate human intelligence?

What is human intelligence?

First, it is necessary to establish what *intelligence* and *human intelligence* means in relation to this paper. This is rather difficult, since philosophers have differing viewpoints on intelligence. For example, Spearman defines intelligence as “the tendency of all human abilities to be positively correlated”, Weschler describes it as the “aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with the environment”, and Pyle argues that intelligence is discipline dependent (Asodun, 2017). Due to the ambiguity in what intelligence is, a broad definition is used in this paper. Namely, that “intelligence can be defined as a general mental ability for reasoning, problem solving, and learning.” (Colom et al., 2010). This paper distinguishes human intelligence as a subset of intelligence, where the characteristics or mechanisms are those that are thought to occur in humans. Of the aspects of human intelligence, such as emotion or the ability to learn, simulating intelligence refers to replicating some aspect of it.

Approach

Arguing the extent to which mathematics can simulate intelligence necessitates an understanding of what knowledge frameworks makes humans intelligent. Unfortunately, different philosophers have different ideas on human intelligence, so a philosopher needs to be picked. The chosen philosopher is Marvin Minsky (1927-2016), an MIT professor who completed research in physics, neurophysiology, psychology and a Ph.D. in Mathematics (Dennis, 2023). Minsky was chosen because of background in mathematics, which is the second subject of this paper, and has written profusely on the mechanisms for intelligence in his books Society of Mind (1985) and The Emotion Machine (2006) as well as Perceptrons: An Introduction to Computational Geometry (1969) where he discusses a mathematical model for simulating intelligence (Dennis, 2023). Minsky believes in the "society of mind" where intelligence stems from several unintelligent organisms, called agents, that collectively achieve intelligence.

The Society of Mind

In Society of Mind, Minsky contemplates what the mind is and how it operates. Essentially, he attempts to create a model of the human mind, a model that can explain:

> How do we recognize objects and scenes? How do we use words and language? How do we achieve goals? How do we learn new concepts and skills? How do we understand things? What are feelings and emotions? How does 'common sense' work? (Singh, 2003, para. 4)

Minsky argues that there is no simple rule governing the behaviour of the mind. Intelligence comes from the combination of several, unintelligent and simple agents, existing in a "society of mind". It is the combination of societies (groups) of agents that can perform complicated tasks and achieve intelligence.

How do societies of mind work?

Minsky argues that for each specific cognitive process, there exists an agency. An agency is a group of agents that are specialized to perform a specific function, including but not limited to creating predictions, remembering, giving context to, and simplifying data. Minsky suggests the existence of k-lines which are amongst the simplest agents. The purpose of k-lines is to activate some set of agents. The k-line is a simple and powerful way to have the mind problem-solve, create memories, and set goals. He further classifies k-lines into nemes and nomes. Nemes are responsible for representing aspects of the world and nomes control how these aspects are processed in the mind. Minsky suggests models to combine agents to form agencies (remember an agency is just a group of agents).

Emphasis is placed on Minsky's viewpoint that intelligence stems from unintelligent life forms that combine to form something intelligent. An argument can therefore be made that the mathematics behind intelligence are simple and structured. Of course, Minsky's viewpoint on intelligence may be incorrect, but contemplating finding a simple model for intelligence is a starting step in finding a mathematical model that simulates intelligence.

Finding a simple model for intelligence

Determining if there is a simple model for intelligence is necessary to understand the degree to which mathematical models can simulate intelligence. There are subtle hints that mathematicians can describe intelligence as a set of simple, structured rules. In April 2000, there was an experiment conducted where the signal transmitted from a ferret's eyes to its visual cortex was rerouted to go to their auditory cortex (the part of the brain usually used for hearing) (Sharma et al., 2000). The visual cortex and auditory cortex have different structures; the visual cortex contains orientation columns which are essentially slabs of neurons that only activate for light travelling in a particular direction. As light hitting a ferret's eyes moves, different orientation columns are activated. When the signal was rerouted to the auditory cortex, scientists found that it changed. Orientation columns formed in the auditory cortex, and the ferrets were able to respond to light stimuli, indicating they could see (Nielsen, 2015). Ferrets are biologically similar to humans so these results can be transferred to humans (Ferret — Ari.info, n.d.). This result implies commonalities between different parts of the human brain. But does mathematics exist to describe this set of rules governing intelligence? Unfortunately, there is no clear answer to this question because scientists do not have an adequate grasp on how the brain works. It is agreed that there is a link between the brain and intelligence and research is being done to understand brain function. Current research discusses how quantum entanglement inside the brain may arise in intelligence (Jedlicka, 2017). But these mathematical models are not yet understood which suggests that the limiting factor of mathematical models that can simulate intelligence is not necessarily underdeveloped mathematics, but an underdeveloped understanding of how intelligence works.

That is not to say mathematical models cannot simulate any brain function: upon inspection of the human brain structure, it appears that neural networks are made up of interconnected neurons which alone are not intelligent, but when combined form an intelligent entity - the human brain (Explained: Neural Networks, 2017). A mathematical model that simulates human neural networks is a promising first step to at least replicate functions of intelligence.

Neural Networks: a mathematical model for learning

One aspect of human intelligence that can be simulated is the ability to learn. Specifically, intelligent species can make predictions of future events based on what they have experienced previously. For example, a human can identify a handwritten number they have never seen before because they have seen enough numbers to recognize that the new number is a different form of the same number.

What are artificial neural networks?

One way that mathematical models have been able to predict handwritten digits is by recreating neural networks. Artificial neural networks (ANN) are loosely based on how neurons are arranged in human brains (Hardesty, 2017). The ANN consists of several interconnected neurons arranged in vertical columns called layers. The first layer is called the input layer, the last layer is called the output layer, and all layers in between are called hidden layers. The input layer uses input values to compute some output which is *fed forward* as input for all connected artificial neurons. The output of the final layer is (typically) the neuron that outputs the highest value. For an ANN that recognizes handwritten digits, the output layer consists of 10 neurons, corresponding to the digits {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}.

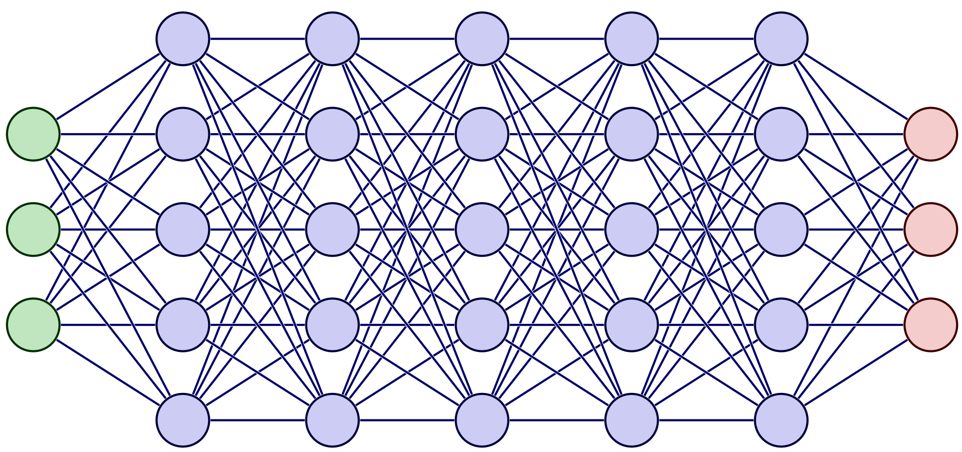

Figure 1: A pictorial representation of a neural network where each node (green, blue, and red circles) is connected to every node from the previous layer. Image generated from code by (Neutelings, 2023).

Not all connections between neurons are equal because not each neuron is as significant in determining the output. This is quite like how intelligent entities make decisions. For example, a human may consider purchasing a new phone. Some considerations may be the price of the phone, if their current phone is broken, and the time to receive the phone. The most important consideration in the purchase decision is whether their phone is broken or not, and the least important consideration is the phone shipping time! Each connection has a strength called a weight. At each neuron, the inputs are summed up and an activation function is applied to it. An activation function is a continuous function applied to the output of each neuron that ensures it outputs a value between 0 and 1. Initially, the connections are assigned random weights, and the output of the neural network is determined. Since these weights were assigned arbitrarily, they need to be adjusted so that the neural network outputs the correct value, as the initial value is likely incorrect. The process of adjusting these weights is done to minimize the error in the output and is called training. Finding out how to minimize the error in the function is an example of optimization; at its heart, neural networks seem to be simply an application of statistics. Understanding the mathematical underpinnings of neural networks helps explore this argument.

The mathematics behind neural networks

Output of a neuron

A helpful analogy to conceptualize the need for mathematics in neural networks is to imagine each node in Figure as a knob. These knobs can be turned to change the output of the neural network, and by turning each knob the correct amount, internal parameters of the neural network (weights and biases) are tuned to generate the intended output.

At each node there are n inputs where x_i denotes the ith input and w_i denotes its corresponding weight. An additional parameter is introduced, called the bias term (b), which is a value added to the weighted sum that can be tweaked to optimize the performance of a neural network by shifting the graph of the activation function (What Is the Necessity of Bias in Neural Networks?, n.d.). The output of a a single neuron is therefore:

$$\sum_{i}^{n} x_i w_i + b \text{ or } \mathbf{w} \cdot \mathbf{x} + b$$

where $\mathbf{w}$ and $\mathbf{x}$ are vectors whose components are weights and inputs respectively.

The activation function

Several neurons are arranged into three groups of columns called layers: (i) the input layer, (ii) the hidden layers and (iii) the output layer. As an example, suppose a neural network is trained to recognize handwritten digits. An input image with a handwritten digit is provided; the input layer consists of some number of neurons which read the brightness value of each pixel in the input image. The hidden layer is where data from the input layer is processed, and the neuron with the largest output in the output layer is the neural network's prediction. To recognize a handwritten digit, the output layer would consist of 10 neurons, corresponding to the 10 possible digits that can be identified.

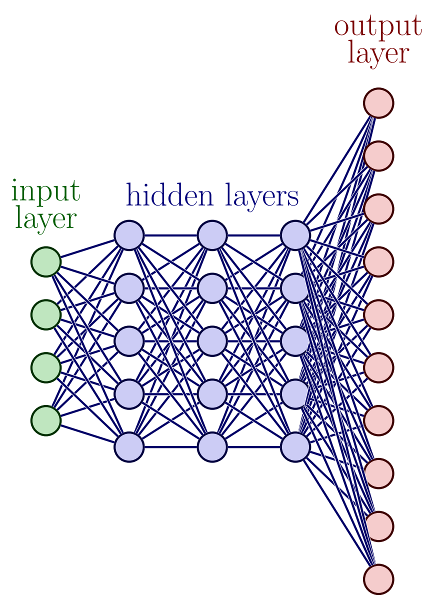

Figure 2: Input, hidden, and output layers of a neural network. Image generated from code by (Neutelings, 2023).

The neuron in the output layer that fires is some combination of the weights and biases of previous layers. Since the output of every neuron is a linear function, the output of the neural network is some linear combination of the input neurons, which limits the complexity of patterns a neural network can predict. An activation function can be applied to the output of each neuron which helps the neural network predict more complicated patterns (Nielson, 2015).

Common activation functions include the sigmoid, ReLU, tanh, and softplus functions (What Are Neural Networks? — IBM, n.d.) Take the sigmoid function, for example:

$$\sigma(z) = \frac{1}{1 + \exp(-z)}$$

where $z = \mathbf{w} \cdot \mathbf{x} + b$ - the output of a neuron. Another advantage of an activation function is that the output is always $0 < \sigma(z) < 1$ which constrains how much changing the output of one neuron has on the output of the neural network (What Are Neural Networks?, IBM, n.d.).

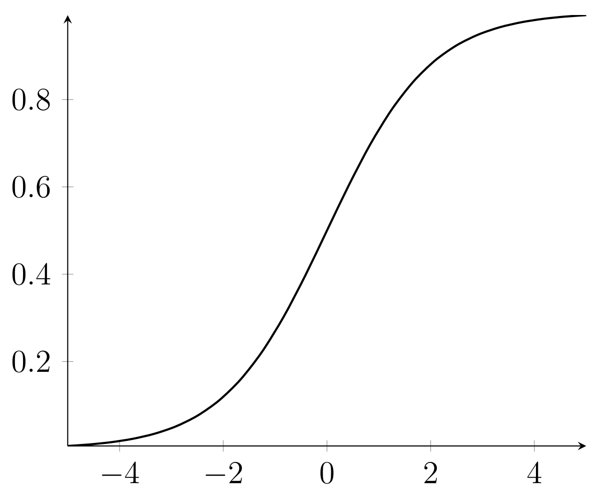

Figure 3: Graph of a sigmoid function; its range is between 0 and 1. Graph generated using pgfplots and TikZ in LaTeX

Gradient Descent

It is necessary to tweak the weights and biases to change the output of the neural network. Changing the the jth weight by Δw_j and the bias by some amount Δb changes the output by some amount Δoutput: $\Delta\text{output} = [{}^{\partial\text{output}}/_{\text{w}_\text{j}}]\Delta\text{w}_\text{j} + [{}^{\partial\text{output}}/_{\text{b}}]\Delta\text{b}$.

The neural network must gauge how much to tweak its weights and biases. This is rather like humans learning when they receive negative feedback for a task. Depending on the breadth of feedback, a human will change its behaviour more drastically to achieve the desired result. This is achieved mathematically using a cost function.

$$C(w,b) = \frac{1}{2n} \sum_x || \; y(x) - a \; ||^2 \tag{1}$$

where w,b are the weight and bias respectively of a neuron, $y(x)$ is the output of a neural network, and a is the expected output of the neural network. The cost function evaluates the square of the distance between the output and the expected result. Note that there must be data on the expected value, and this is called labelled data (*What Is Data Labeling? — IBM*, n.d.)

The goal of the neural network is to minimize this error, and this is done through gradient descent. Gradient descent is analogous to finding the fasting to walk down a hill; an individual will walk opposite to the direction of steepest slope. First, compute the slope of the cost function. Since the cost function depends on two variables, the partial derivatives must be summed (Nielson, 2015).

$$\Delta C(w,b) = \frac{\partial C}{\partial b}\Delta b + \frac{\partial C}{\partial w}\Delta w \tag{2}$$

For convenience in notation later on, let $\Delta d = (\Delta b, \Delta w)^T$ and $\nabla C = \left(\frac{\partial C}{\partial b}, \frac{\partial C}{\partial w}\right)$ then $\Delta C = \nabla C \cdot \Delta d$

The next step is to update the values of b and w. Recall that the goal of gradient descent is to move opposite to the direction of the gradient, so the values of b and w are updated to move in the direction of the negative gradient of the cost function like so:

$$b \to b' = b - \eta \nabla C \text{ and } w \to w' = w - \eta \nabla C$$

Note: computing the gradient of the cost function, ∇C, is computed using a process known as backpropagation (Nielson, 2015) and is omitted here.

After looking at the role of mathematics in computing the output of a neural network, it boils down to just applied statistics: there is some data and need to fit a function to that data. Neural networks are just an elaborate method to approximate complicated functions. It is remarkable that the mathematics behind ANNs have the capability to "learn", so to speak. But even ANNs can only learn in a very crude manner. The easiest limitation to spot is that ANNs only accept numeric input, tasks like NLP (natural language processing) require text to be translated into numbers. This is the primary difference between human intelligence, specifically learning, and machine learning. The human is still more versatile in the breadth of problems it can learn from.

The limitations of neural networks

So far, it does seem that the extent to which ANNs can simulate human intelligence is restricted to decision-making tasks. ANNs cannot simulate, for example, human emotion. Similarly, an ANN is also limited in its ability to tackle new problems; you cannot expect an ANN trained to recognize handwritten digits to accurately recognize handwritten letters. It is necessary to retrain the neural network with training data that includes handwritten letters or train a different neural network specifically for recognizing letters. The idea of training two separate neural networks does seem appropriate considering Minsky's perspective. Each neural network is trained to make a specific decision (i.e. recognize letters, digits, or even letters in different languages) could be considered an agent. And in a society of mind, these agents would be connected by k-lines to form an agency. One would expect that different agencies would correspond to different aspects of intelligence. But this, of course, depends on the breadth of problems that neural networks can solve. Part of what makes human intelligence is its ability to adapt to almost any new stimuli by learning how to act differently. This is especially prevalent in infants who must learn to interact with their surroundings. For example, if a child touches something sharp and feels pain, the child will tune internal weights in neural connections so that they avoid sharp objects. Mathematically speaking, we define some function, called a cost function which corresponds to an undesired response to stimuli (i.e. pain). By adjusting weights in the neural network, one can minimize the cost function and, in that way, adapt to new environments quickly (this is exactly what a neural network does). Furthermore, neural networks are a suitable mathematical representation of intelligence because neural networks can learn anything. By changing internal weights and biases a neural network can minimize the cost function and fit a curve through as many training data points as possible.

Are neural networks a representation of intelligence?

If ANNs are applied statistics, the extent of their intelligence needs to be contemplated. Artificial neural networks are in some capabilities intelligent because they can replicate the human's ability to learn. This begs the question: are neural networks the basis for representing other characteristics of human intelligence? This is a challenging question to answer because, once again, the mechanisms of human intelligence are not yet understood. However, there do seem to be recurring patterns in brain structures, like how ferrets were able to learn to see with their auditory cortex as discussed earlier. This same thinking can be applied to other faucets of intelligence, namely emotion. Perhaps emotion can be described with mathematical models with ANNs. Arguably, ANNs possess emotional intelligence in some capacity. This is seen in the ability of ANNs to recognize patterns in images. Specifically, the ANN could be trained with images containing facial expressions and tasked to classify emotions. This is potentially the beginning of a mathematical model capable of simulating intelligence through predicted emotions. However, the problem arises in the computation power necessary to analyze a large image. An ANN requires numbers as inputs; each pixel can be quantified as three RGB colour values. The difficulty with this is that for large images, say 100×100 pixels, that would require $10,000$ input neurons that process 3-dimensional input (since there are 3 RGB colour values). Increasing the image size becomes computationally intensive and the simple, 2-layer ANN demonstrated cannot be expected to process such a large image (Algorithmic Simplicity, 2022). If there is a way to reduce the number of inputs into an ANN, it may be possible to identify emotions in images. The solution comes from a special type of neural network called a Convolutional Neural Network (CNN) which looks at small, perhaps 10 by 10 pixel patches of the input images. A neural network is applied to every patch. And another neural network is applied to the output of the neural network applied to every patch. If applying a neural network to the output of a patch is done enough iterations, the CNN can identify "special characteristics " of the image, such as the position of grass compared to a horizon (Algorithmic Simplicity, 2023). In this way, a CNN is potentially able to accurately classify emotions.

On the limitations of neural networks

Ultimately, the success of a neural networks (a mathematical model for intelligence) is computer hardware capability. This has been seen previously in the development of neural networks. For example, foundations for the mathematical models used in neural networks were available in Rosenblatt's paper on perceptrons (Rosenblatt, 1958). Turing essentially predicted the advent of machine learning in The Imitation Game paper (Turing, 1950). But ANNs only began extensive use in the early 2000s after computers become more capable. This suggests that perhaps the existing mathematical models can simulate human intelligence entirely. That is, mathematical models can work to essentially create an artificial human indistinguishable from a real human in the way it learns, has emotions, and problem-solves. However, scientists are limited by the capabilities of computer hardware. There was some discussion earlier on how the brain potentially has quantum effects that are responsible for intelligence. It is possible that advances in quantum computers, and with it, mathematical neural network models, are able to better replicate intelligence. Admittedly, computer hardware may not be the only limiting factor in determining a mathematical model that simulates intelligence, but human ignorance as well. Minsky are Papert are largely held accountable for the "AI winter" in the 1980s after they published a book Perceptrons outlining the inherent weaknesses of Rosenblatts idea of perceptrons (Schuchmann, 2019). An issue raised was that perceptrons were unable to solve non-linear separable problems (like XOR) (Schuchmann, 2019). This issue could be rectified by using multiple layers of perceptrons, but there was no known way to train a multi-layer network. It required backpropagation, a mathematical tool to compute the derivative of the cost function, which Minsky and Papert were not aware of; to them backpropagation had not been invented yet.

Conclusion

The mathematical models of today are only able to partially simulate human intelligence. Specifically, using ANNs (artificial neural networks), mathematical models can simulate learning. This is exciting because it allows humans to solve complicated problems that cannot be described explicitly with algorithms.

As well, ANNs follow the theme of philosopher Marvin Minsky's ideas: that intelligence consists of independent, unintelligent beings (neurons) that come together and act intelligently (the neural network).

ANNs have been promising because of their utility in society, like ATMs that read bank cheques. As well, ANNs have been expanded to interpret data from images and text, leading to the rise of services like ChatGPT or Google Lens.

However, there currently exists no simple way to generalize ANNs to all brain functions. For example, there is no clear way to have a mathematical model for emotions, which is another aspect of intelligence. Although convolutional neural networks can perhaps act to identify emotion, it does not necessarily have emotion. The mathematical models of today are only able to simulate components of human intelligence.

Areas for further research

This research paper focused only on neural networks as a model for human intelligence. This focused the discussion on the paper since artificial neural networks are loosely based on the human brain, but there exist other algorithms capable of learning. For example, swarm algorithms optimization and genetic algorithms are all biologically inspired, albeit not characteristic of humans (Darwish, 2018). These have the potential to simulate the human population learning to survive, like with natural selection. Perhaps to entirely replicate human intelligence, it is necessary to abandon Minsky's thought process of intelligence being structured and simple. Perhaps intelligence is an amalgamation of several distinct mathematical models that work in conjunction to simulate intelligence.

References

Asodun, F. (2017). The Problematic Search For the Nature of Intelligence: Philosophical Reactions and Projections. *Journal of Interdisciplinary Studies*, *3*(1), 141–164. https://philarchive.org/archive/FATTPS

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues in Clinical Neuroscience*, *12*(4), 489–501. https://doi.org/10.31887/dcns.2010.12.4/rcolom

Crow, J. F. (2002). Unequal by nature: a geneticist's perspective on human differences. https://www.amacad.org/publication/unequal-nature-geneticists-perspective-human-differences#:~:text=As%20a%20result%2C%20we%20share,2%20percent%20of%20our%20DNA.

Darwish, A. (2018). Bio-inspired computing: Algorithms review, deep analysis, and the scope of applications. *Future Computing and Informatics Journal*, *3*(2), 231–246. https://doi.org/10.1016/j.fcij.2018.06.001

Dennis, M. A. (2023, December). Marvin Minsky — AI pioneer, cognitive scientist MIT professor. https://www.britannica.com/biography/Marvin-Lee-Minsky

Explained: Neural networks. (2017, April). https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414

Ferret — Ari.info. (n.d.). https://www.animalresearch.info/en/designing-research/research-animals/ferret/#:~:text=Ferrets%20are%20closer%20biologically%20and,physiological%20similarity%20to%20human%20lungs.

Hardesty, L. (2017, April). Explained: Neural networks. https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414

Jedlicka, P. (2017). Revisiting the Quantum Brain Hypothesis: Toward Quantum (Neuro)biology? *Frontiers in Molecular Neuroscience*, *10*. https://doi.org/10.3389/fnmol.2017.00366

Neutelings, I. (2023, December). Neural networks. https://tikz.net/neural_networks/

Nielsen, M. A. (2015). Neural networks and deep learning. http://neuralnetworksanddeeplearning.com/

Schuchmann, S. (2019). History of the first AI Winter - Towards Data Science. https://towardsdatascience.com/history-of-the-first-ai-winter-6f8c2186f80b

Sharma, J., Angelucci, A., & Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature*, *404*(6780), 841–847. https://doi.org/10.1038/35009043

Simplicity, A. (2022, October). Why do Convolutional Neural Networks work so well? https://www.youtube.com/watch?v=8iIdWHjleIs

Singh, P. (2003). Examining the society of mind. https://www.jfsowa.com/ikl/Singh03.htm

What are Neural Networks? — IBM. (n.d.). https://www.ibm.com/topics/neural-networks#:~:text=Since%20neural%20networks%20behave%20similarly,output%20of%20the%20neural%20network.

What is Artificial Intelligence (AI) ? — IBM. (n.d.). https://www.ibm.com/topics/artificial-intelligence

What is Data Labeling? — IBM. (n.d.). https://www.ibm.com/topics/data-labeling

What is the necessity of bias in neural networks? (n.d.). https://www.turing.com/kb/necessity-of-bias-in-neural-networks#what-is-bias-in-a-neural-network?