University of St Andrews

Evolution of AI

Saurav Jha

sj84@st-andrews.ac.uk

December 18, 2020

1 Abstract

With the rapid emergence of Artificial Intelligence (AI) technologies getting inextricable with everyday life of individuals, economies, societies, and communities, it has never been a better time to look back at the inception of the field and learn from the ups and downs along the way. This essay aims to cast light on the history of AI along with the evolutionary principles guiding its progress. Being connected deep with mankind's expectations of procreating its reflection, the field of AI has evolved greatly within the past decades and can now be thought of having a life of its own. The advancing industrial aspects of AI in the recent years have exerted a huge pressure on the growth and development of the most storage and computationally efficient algorithms. That said, the Darwinian laws of evolution can be applied to keep a check on the growth of modern AI techniques that share surprisingly common ancestry with one another.

2 Introduction

As with all life forms on Earth, the Darwinian evolutions holds true for technological world wherein the variations of a new technology with superior performance are selected to pass on their features to future designs. The evolution here, however, differs from biological evolution in the sense that no new technology is created from scratch but are instead built from the combination of existing ones. This phenomenon of *combinatorial evolution* - as coined by W. Brian Arthur[1] - thus takes place only once a technology exists, and occupies a predominant position for machines unlike in biology where incremental changes through variation and selection are common.

As suggested by Max Tegmark (Tegmark, 2017), the development cycle of all life forms on the planet can be summed up in three different stages: Life 1.0 referring to the organisms such as bacteria that rely entirely upon biological forces for the evolution of their hardware (the anatomy) and their software (the senses or the information preserved in the DNA that helps replicate the anatomy), Life 2.0 alluding to the species such as ours whose hardware have evolved biologically but software is largely designed, and Life 3.0 implying to the potential upgrade to artificially intelligent forms that will no longer be restricted by biological evolution for the design of its hardware and software. While life 2.0 allowed our species to dominate the Earth within a period of few hundred millennia, the potentiality of life 3.0, aka AI is hard to imagine given its ability to impart intelligence into lifeless forms. As a technology, AI however, can be thought of being guided by the same principle of combinatorial evolution.

3 Birth of AI

The period of 1936-56 can be traced as the golden years leading to the birth of AI. In 1936, Alan Turing brought the concept of Turing machine - a first of its kind of universal function approximator in the field of computer programs. Inspired by this, Turing laid the definition for AI by proposing an imitation game between such an all-encompassing machine and a human, namely the Turing Test (Turing, 1950). It suggested that a machine imitating the sentient behavior of a human (for e.g., process natural language, learn and recall from communication, and deliver ideas to a human) can be reasonably said to have gained the potential of thinking.

[1] https://www.computerweekly.com/feature/Darwinism-theory-of-evolution-applied-to-technology

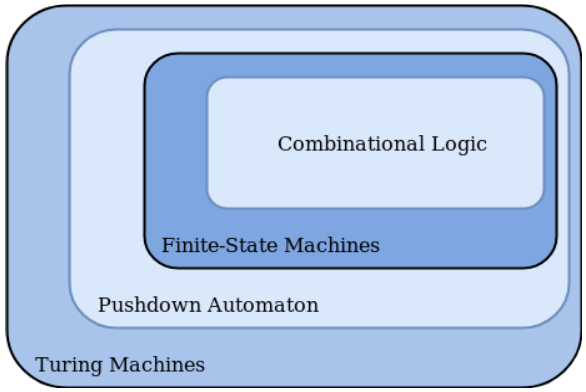


Figure 1. Evolution of self-operating machines: an adaptation from Wikipedia.

An year later, Marvin Minsky from the Princeton University built the first neural machine called SNARC - a randomly connected network of 40 Hebb synapses that employed probability knobs for remembering mathematical functions. Further, the term 'artificial intelligence' was coined at a 1956 workshop in Dartmouth which aimed to present a computer program called Logic Theorist[2] - a software designed to solve mathematical theorems through heuristic programming.

This relatively short timeline can hence be regarded as the event of abiogenesis in the world of computer science wherein the first self-operating systems started branching out of the inanimate primordial soup of hard-coded computer programs. The recently born field caught the attention of several young researchers in the academia and set off a long and steady race of creating systems that could imitate human intelligence. In other words, these fundamental concepts would set up the premise for the upcoming combinatorial evolution in the field of AI.

4 The near wipe-out event for AI

With the inception of AI marking a new era in the field of computer science, it nevertheless excited the then prominent researchers of the field to think about possibilities that such machines can bring to humans. Turing, in 1950, believed that by 2000, computers with 128MB of RAM will be advanced enough to give interrogators a 70% chance after five minutes of conversation. Similarly, Herbert Simon of IBM, fascinated by his recent success of the Logic Theorist (1956) and the General Problem Solver (1959), predicted that AI would be advanced enough to defeat world-level chess champions within 10 years. However, a fundamental notion of growth that their claims overlooked was the law of accelerating returns (Kurzweil, 2006) stating the exponential nature of the growth curves for technology instead of linear. Though the basics of AI had been laid off, these early programs were severely limited in incorporating the domain knowledge and computational complexities involved in solving large scale real-world tasks. Or we can say that they still lacked evolutionary features worth of being inherited to build further successful generations of AI systems. Simon, for instance, had little clue that it would take another three decades for his prediction to be true.[3]

By the late 1950s, it was apparent that the limited survival characteristics, and the lack of diversity in the field were making it confront a steady pace of growth. The following years brought further serious blows to the field. Funding for projects such as the 1957 machine translation project by the U.S. National Research Council for translating Russian scientific papers into English got thwarted after years of research showed inadequacy of AI in grasping concepts such as sense disambiguation and semantic representations. Further, the subsequent failure of general combination-based nature of most search-based AI systems, and the lack of proper training algorithms for the then perceptron networks were made AI to be of little use in solving real-world problems with combinatorial explosion of unknown variables.

If the years leading to late 1950s were the ages of birth and establishment of AI (equivalent of Paleozoic Era to the biological life), the events in the following two decades can be regarded as an eradication for projects and funding in the field. The largely exaggerated promises among the scientific field gathered nothing but negative attention and several objections. Some of these were the Lady Lovelace's objection stating that computers can never do anything new since they are built to obey rules; Godel's incompleteness theorem claiming that any formal

[2] http://shelf1.library.cmu.edu/IMLS/MindModels/logictheorymachine.pdf

[3] It was in 1996 that IBM's Deep Blue first defeated Garry Kasparov, the then world chess champion in just one of a six-game match.

theory has some true statements that have no proof and hence, it is impossible to tell whether a computer actually learns to solve a problem (Strong AI) or pretends to do so (Weak AI); the objection from neuroscience and the qualification problem claiming that computers can never simulate the continuous and complex behaviour of brain. Together with questioning the possibility of AI, such criticisms also shaped the utility of the field by continuously adapting to what the representation of the world should be for machines.

The period also witnessed a number of attempts towards simulation of imitation games. ELIZA (1966) was the first of such program to imitate the initial interview with a non-directional psychotherapist. Later, SHRDLU (1968) processed instructions from users in English to move various objects around in the "blocks world". These simulations advocated the behaviourist notion of human mind in favor of Strong AI saying that an information-processing computer can have a mind in exactly the same sense human beings have minds. On the other hand, the biggest rebuttal to the behaviourist notion came from Searle's Chinese room argument which denied the mere act of information processing by a machine to be called "thinking" and hence, showed the insufficiency of Turing test in detecting the presence of consciousness. The argument, however, did not restrict the amount of intelligence displayed by a machine and hence, can not be thought of having an intention to thwart the field's ongoing progress.

5 The unsurpassed recovery

As the blows from aforementioned arguments continued challenging the aspects of Good old-fashioned AI (GOFAI), the upcoming years proved to be critical for defining the relevant goals of the field. Analogous to a mass extinction event pruning specific lineages off the tree of life while stimulating the growth of other branches[4], the resurgence of AI after 1980s saw the end of general-purpose search and the rise of two major methods: expert-driven and data-driven.

5.1 Shift to expert-driven systems

Having observed the limitations of general purpose search algorithms, more AI research focused incorporation of domain knowledge and hence, the era of "expert systems" kicked off in the early 80s. The restriction of such systems to specific domains helped shift the focus of the industry from Artificial General Intelligence (AGI) to artificial intelligence for specific tasks. The most successful of expert systems were the DENDRAL project that inferred the molecular structure of compounds from mass spectrometer readings and the XCON project, a production system which helped DEC save 40 million dollars annually by 1986. Shortly, governments such as Japan began the Fifth generation computer project with an aim to build machines that could carry on reasoning and conversations, translate languages, and interpret pictures; the UK launched an Alvey programme while the US formed its own research consortium called the Microelectronic and Computer Technology Corporation (MCC). It were these efforts that helped AI grow into a billion-dollar industry by 1988.

The rapid growth of industrial value of the field opened the doors to ample resources that soon resulted in advancement of new branches such as multi-layered neural networks through modification of previous ones (i.e., perceptrons) in the tree of life for AI. The credit for the renewed interest in field of neural nets can be traced back to 1982 when John Hopfield replaced the concept of one-way connections between neurons by bidirectional ones. This slight yet powerful modification was the golden characteristic introducing a whole new lineage in the neural net ancestry - a trait that would be rigorously inherited by the upcoming variations of neural networks lasting till today. Later, in 1986, Geoffrey Hinton and his colleagues reintroduced the "backpropagation" algorithm for training of multi-layered perceptron networks by efficiently computing first-order gradients. This was an evolutionary successor to previous techniques requiring extensive storage and computational capacity. Soon after, neural networks started gaining commercial value.

5.2 Shift to data-driven systems

With the rise of the World Wide Web and the availability of cheaper computational resources (driven by Moore's law) in late 1990s, a new trend of studying data and not just the algorithm gained popularity. This led to a series of discoveries of data-driven machine learning algorithms such as Random Forests (Ho, 1998), Recurrent neural networks (Schuster and Paliwal, 1997), Support vector machines (Suykens and Vandewalle, 1999), and other kernel-based methods started, all of which looked at large quantity of data to optimize its parameters. While the results strongly alluded how crucial having enough data was for algorithms to build knowledge representation of the problem in-hand, the vast amount of resources over the internet fueled the rapid diversification of AI into supervised, unsupervised and semi-supervised learning techniques. This expansion can be thought of being largely Darwinian considering the influence of increasingly available resources.

[4] https://evolution.berkeley.edu/evolibrary/article/massextinct_04

6 The Deep Learning boom

While the concept of neural networks and backpropagation had been existent since three decades and GPU-accelerated computing had already sparked the rise of machine learning frameworks, one particular event of 2012 is firmly believed to have ignited the modern age of deep learning models. The AlexNet architecture, an eight-layered convolutional neural network (Krizhevsky et al., 2012) surprisingly achieved a top-5 error of 15.3% in the 2012 ImageNet Large Scale Visual Recognition Challenge[5] and became an overnight sensation in the domain of machine vision thus inspiring several other image classification models (Alom et al., 2018) such as ZFNet (Zeiler and Fergus, 2014), VGGNet (Simonyan and Zisserman, 2014), GoogleNet (Szegedy et al., 2015), ResNet (He et al., 2016), etc.

On the other hand, the field of language processing dealing with temporal sequences also witnessed a similar upsurge after 2010. The two-decade old concept of RNNs was succeeded by several variants that could alleviate the problem of vanishing gradients in longer sequences by learning long-term dependencies while attending only to relevant parts of the sequences (Chorowski et al., 2015). The most accredited of these were the long short-term memory networks (Hochreiter and Schmidhuber, 1997) and gated recurrent units (Chung et al., 2014).

That being said, Darwinism can be seen to have unprecedented effect on the evolution of deep learning models. The recent years have witnessed enormous rise of different networks with each one addressing the discrepancies of its predecessor. For example, the slower computational processing and memory limitations of LSTM[6] have resulted in rapid adoption of Transformers-based (Vaswani et al., 2017) approaches for sequential processing tasks in recent. The fast-pace transformation of deep learning industry is thus among the most prominent examples of Darwinian evolution of AI.

7 Current state of affairs

AI programs have now gained tremendous applications in our day-to-day life. Virtual personal assistants such as Google, Siri, Bixby and Alexa can process instructions from voice and text. Their conversation techniques, while still not enough to fool an interrogator, can be thought to lie mid-way in the spectrum of weak to strong AI.

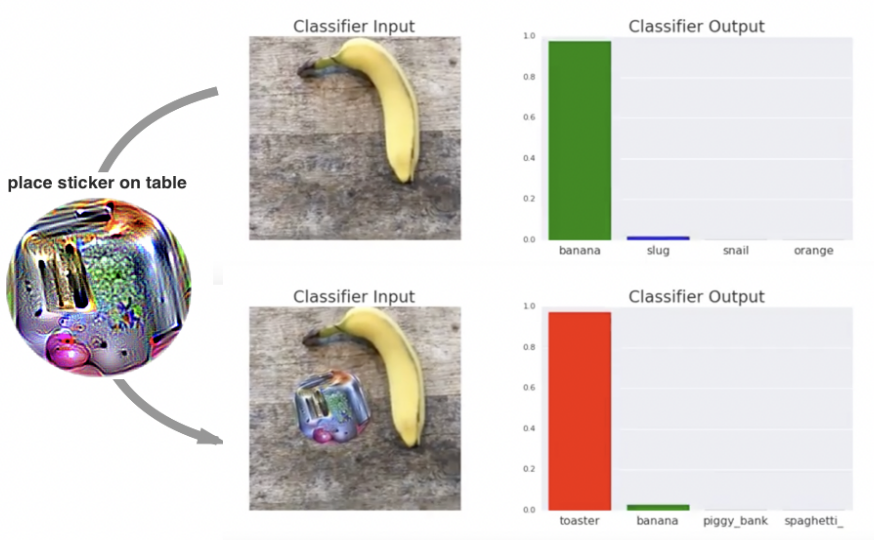


Figure 2. Figure showing a real-world adversarial attack on input of VGG16 model: adapted from (Brown et al., 2018).

Further evidence of lack of strong AI can be derived from the field of machine vision where even the state-of-art image classification models behave miserably when supplied with slightly perturbed inputs. As depicted in figure 2, when a sticker is placed physically next to a banana, the perfectly fine classification model predicts the object as a toaster with 99% confidence.

A similar flaw can be noticed in the domain of natural language processing wherein even the most advanced of AI systems fail to take care of the facts for language modeling tasks. Figure 3 shows the system tries to generate

[5] http://image-net.org/challenges/LSVRC/2012/
[6] https://towardsdatascience.com/the-fall-of-rnn-lstm-2d1594c74ce0

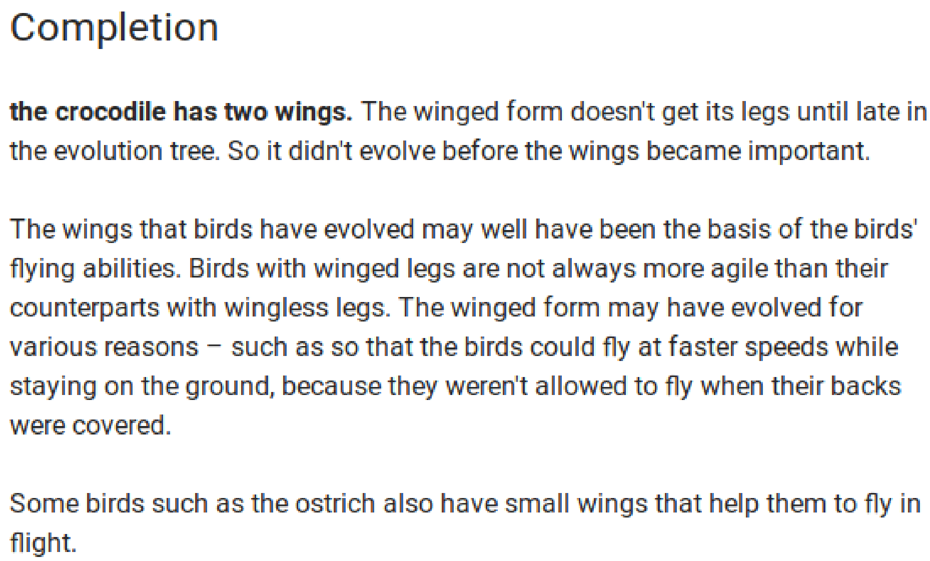


Figure 3. Figure showing the text generated by a Transformer model to continue the sentence - ***"the crocodile has two wings."*** A live demo is accessible at https://talktotransformer.com/.

text for smooth continuation of the input sentence. However, in absence of proper domain knowledge of animal kingdom, it fails to keep a check on common facts about a crocodile's wings. Consequently, we can argue that the intentionality (i.e., beliefs and desires) of the state-of-art deep learning models are still far from letting them fulfill Turing's consciousness argument, and seem more like memorization of input patterns.

The above adversaries have sought the attention of several AI research groups in adhering to well-defined rules to develop and evaluate complex AI systems in terms of more accurate metrics rather than a single metric or resemblance to human behaviour. Deepmind, for example, proposed a behaviour suite enlisting seven core capabilities that any AI agent should possess (Osband et al., 2019): generalization, long-term impact assessment of an action, scalability, noise-robustness, learning curve for simple tasks, willingness to explore and recall its experience. AI industry today can hence, be seen to be moving towards a more robust and sustainable future.

8 The Future

With the tremendous expansion of the AI industry, it may be regarded as certain that the upcoming years will witness another shift to a totally unexpected method. A forewarning of this can be derived from the MIT technology review[7] published this year which tracked the words mentioned across 16,625 arXiv papers belonging to the "artificial intelligence" section to come up with the following three-point synopsis:

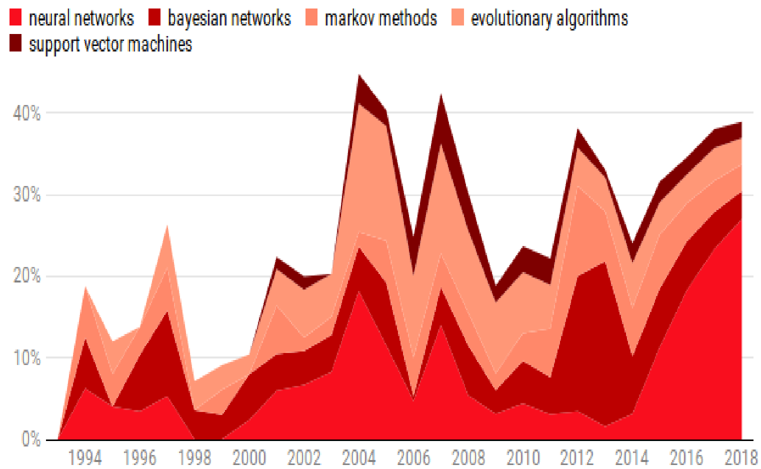
1. Early 2000s drove the transition to data-based systems from hard coded rule and knowledge-based systems.

2. As depicted in figure 4a, a steady competition existed between a variety of machine learning techniques but the pivotal breakthrough of AlexNet on 2012 ImageNet competition (see Section 6) gave deep learning an upper hand.

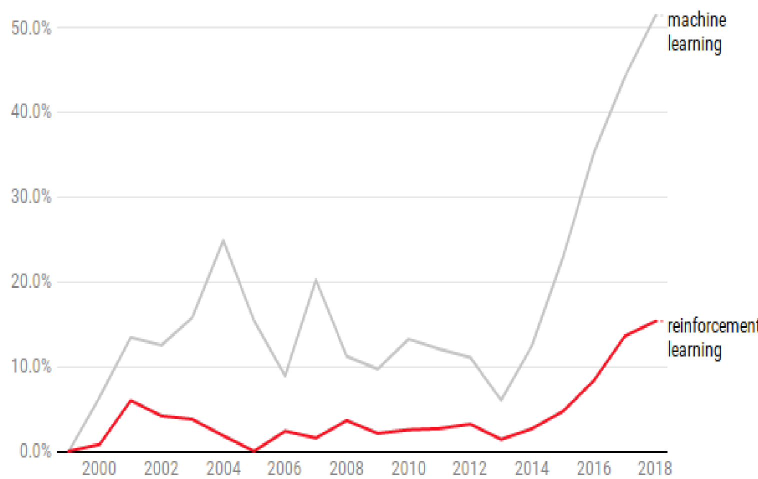
3. The last few years have, however, shown the concept of reinforcement learning (RL) - a branch of machine learning that mimics animal behavior by taking actions in an environment that would maximize its reward - taking over existing learning approaches (see figure 4b).

Given the recent success of RL on tasks such as playing AlphaGo (Chang et al., 2016) and effective robotic locomotion (Levine et al., 2016), it can be anticipated that these methods might be the next big move in the field of AI thus taking over deep learning in the next decade.

[7] https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/



(a) Percentage of papers that mention ML method



(b) Rising momentum of reinforcement learning

Figure 4. Results of Karen Hao's 2019 analysis of 16,625 arXiv papers showing the trend of AI over past 25 years.

9 Conclusion

This essay throws light on the factual history of AI and the evolutionary principles guiding them. Going through several examples in different era of AI, it can be observed that natural selection has indeed been a major driving force for the continuous expansion of the field. From the inception of the domain to the recent success of deep learning, the algorithms and principles guiding AI have been consistently shaped by the criticisms from within the field as well as from non-AI methods. We thus discuss the development of AI through four different phases leading up to its current state and the role of the internet and computing resources in helping it become one of the fastest growing technologies. We further debate how the state-of-art models in machine vision and language processing still lack the essence of strong AI. As with any other technology, AI follows the laws of combinatorial evolution, and hence its progress has been dependent on the quality of traits that are carried out to the next generation algorithms. A good part of the essay explores these traits common to various AI methods and how the inheritance of useful features with iterative build-ups led to new branches in the tree of life for AI. Adhering to the Darwinian rules of evolution, we also talk about the possible direction where AI might be heading. The swift transformation of the field from being research-focused to being used in large-scale industry production has once again attracted high-end speculations. It would thus be thrilling to witness whether the field manages to keep up to the expectations of humankind or falls again into the trap of the laws of accelerating returns. As Stuart Geman once said - "the world is compositional or God exists" - it has always been and forever shall be exciting to see what compositional nature of our world does AI help us uncover.

References

Md. Zahangir Alom, Tarek M. Taha, Christopher Yakopcic, Stefan Westberg, Mahmudul Hasan, Brian C. Van Esesn, Abdul A. S. Awwal, and Vijayan K. Asari. The history began from alexnet: A comprehensive survey on deep learning approaches. *ArXiv*, abs/1803.01164, 2018.

Tom B. Brown, Dandelion Mané, Aurko Roy, Martín Abadi, and Justin Gilmer. Adversarial patch. *ArXiv*, abs/1712.09665, 2018.

Hyeong Soo Chang, Michael C Fu, Jiaqiao Hu, and Steven I Marcus. Google deep mind's alphago. *OR/MS Today*, 43(5):24–29, 2016.

Jan Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio. Attention-based models for speech recognition. In *NIPS*, 2015.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20:832–844, 1998.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

Raymond Kurzweil. The singularity is near: When humans transcend biology. 2006.

Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 17(1):1334–1373, 2016.

Ian Osband, Yotam Doron, Matteo Hessel, John Aslanides, Eren Sezener, Andre Saraiva, Katrina McKinney, Tor Lattimore, Csaba Szepezvári, Satinder Singh, Benjamin Van Roy, Richard S. Sutton, David Silver, and Hado van Hasselt. Behaviour suite for reinforcement learning. *ArXiv*, abs/1908.03568, 2019.

Mike Schuster and Kuldip K. Paliwal. Bidirectional recurrent neural networks. *IEEE Trans. Signal Processing*, 45: 2673–2681, 1997.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

Max Tegmark. Life 3.0: Being human in the age of artificial intelligence. 2017.

Alan M. Turing. Computing machinery and intelligence. 1950.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, 2017.

Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.