

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/333039347>

Analyzing the Prospect of an Approaching AI Winter

Thesis · May 2019

DOI: 10.13140/RG.2.2.10932.91524

CITATIONS

7

READS

11,887

1 author:

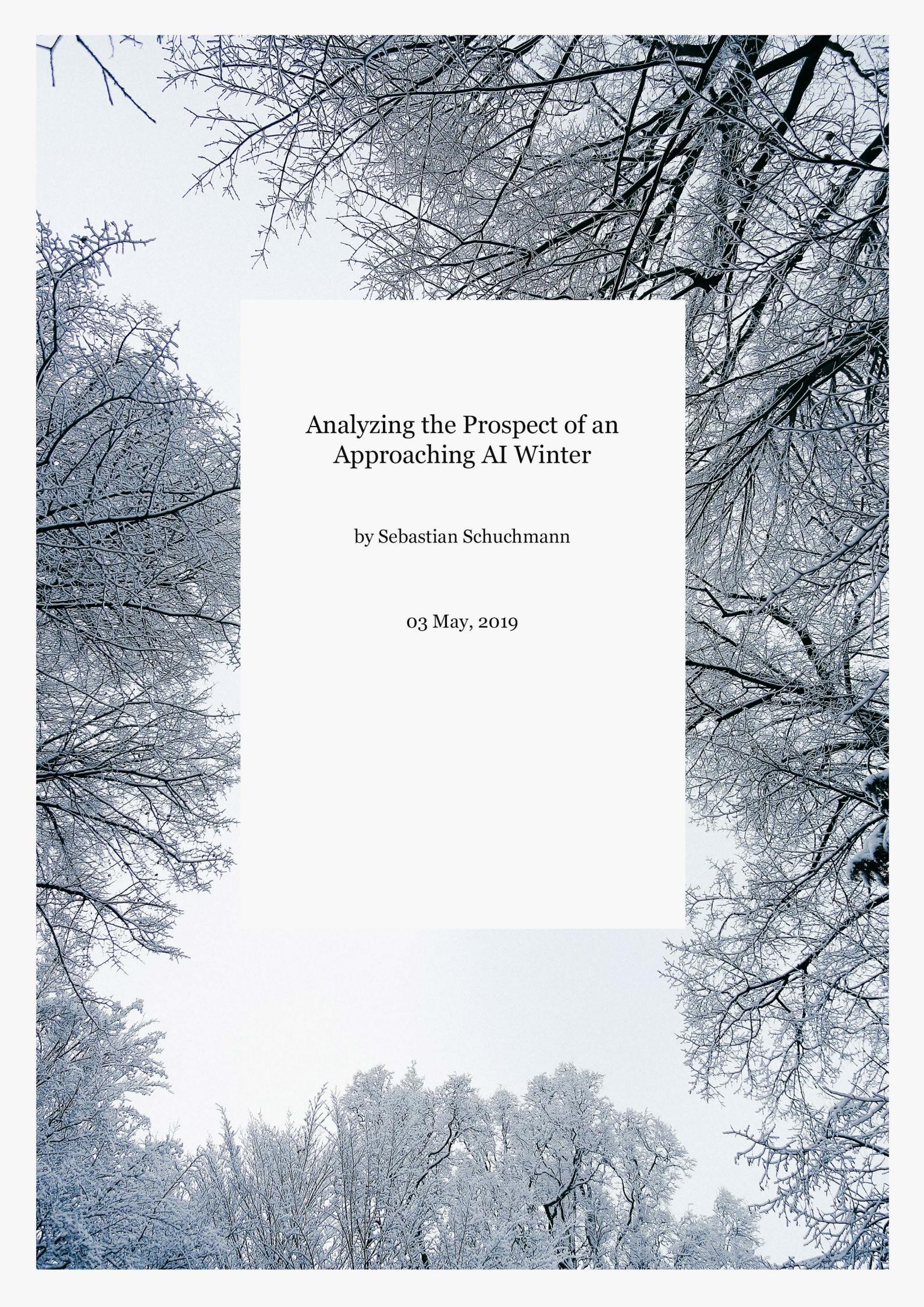


Sebastian Schuchmann

-

1 PUBLICATION 7 CITATIONS

SEE PROFILE



Analyzing the Prospect of an Approaching AI Winter

by Sebastian Schuchmann

03 May, 2019

Contents

1	Introduction	5
1.1	Motivation	5
1.2	Defining Artificial Intelligence	5
1.3	Defining AI Winter	6
1.4	State of the Art Review	7
1.5	Methodology	7
2	History of AI Winters	7
2.1	Events leading to the first AI Winter	8
2.1.1	Beginnings of the AI Field in the 1950s	8
2.1.2	The Quiet Decade	10
2.1.3	Fall of Connectionism	11
2.1.4	The Lighthill report and its Consequences	11
2.1.5	The first AI Winter	12
2.2	Events Leading to the Second AI Winter	12
2.2.1	AI Boom Times	12
2.2.2	Fears of an upcoming Winter	13
2.2.3	The second AI Winter	13
2.3	AI Spring and Modern History	14
2.3.1	The Deep Learning Revolution	15
3	Analysis: What has Led to the AI Winters?	16

4 Probability of an Approaching AI Winter	18
4.1 Expectations and Promises	18
4.1.1 Hype and the Hype Cycle	18
4.1.2 Comparison to Expert Opinion	20
4.2 Investment and Funding	21
4.3 Evaluating Progress	22
4.4 The Criticism of Deep Learning	23
4.4.1 Limitations of deep learning	23
4.4.2 Deep Learning is brittle	24
4.4.3 The Dominance of Supervised Learning	28
5 Conclusion	29

Abstract

This paper examines the probability of an approaching AI winter by comparing historical patterns to the modern era. First, the historical literature is analyzed and the central causes leading to an AI winter period are distilled and discussed:

1. Expectations and promises compared to the actual results;
2. Funding from governments and industries;
3. Technological limitations.

Second, the present is analyzed from the perspective of these causes.

Regarding the expectations and promises, the analysis indicates that a Hype Cycle pattern can be found in current conference attendees numbers similar to that in the AI winter in the 1980s; consequently pointing to a potential decline in attendees and interest. Furthermore, while optimistic claims about the progress of AI are commonly made by futurists such as Kurzweil predicting machines will reach human-level intelligence by 2029, a meta-analysis of AI predictions has shown their unreliability, including those of experts.

Nonetheless, government and industry funding is healthy and the “AI Race” mentality, which is similar to the 20th century Space Race, lessens the prospect of reduced funding. Nonetheless, forecasts for the market in 2025 still show an enormous variance.

Finally, there are substantial critiques of the current deep learning approaches. However, the experts appear to be aware of the critique and have started building solutions.

To conclude, there is a real possibility of an approaching AI winter. Still, the current funding provides the field with a solid foot on the ground even as public sentiment is starting to shift. Furthermore, numerous companies have integrated AI solutions throughout their products.

1 Introduction

This paper addresses the question of whether the field of Artificial Intelligence (AI) is approaching another AI winter or not.

1.1 Motivation

Both industries and governments alike have invested significantly in the AI field, with many AI-related startups established in the last 5 years [1]. If another AI winter were to come about many people could lose their jobs, and many startups might have to shut down, as has happened before [2, p. 327]. Moreover, the economic difference between an approaching winter period or ongoing success is estimated to be at least tens of billions of dollars by 2025, according to McKinsey & Company[3, p. 6].

This paper does not aim to discuss whether progress in AI is to be desired or not. Instead, the purpose of the discussions and results presented herein is to inform the reader of how likely progress in AI research is.

1.2 Defining Artificial Intelligence

“Viewed narrowly, there seem to be almost as many definitions of intelligence as there were experts asked to define it.” [4]

R. J. Sternberg

Many experts define Artificial Intelligence by referencing human intelligence [5][6][7]. Unfortunately, there is no unified definition of what constitutes human intelligence or intelligence in general, which in turn leads to there not being a generally agreed upon definition for AI either.

Some claim that the definition of AI is always shifting so that problems already solved by AI are not regarded as AI anymore. [8, p.12]. In a like manner, modern definitions often expect AI to be a system that can learn from data [9]. However, this would prevent the inclusion of expert systems created manually by surveying experts in a field. These systems work on an “if-then”-rule basis. This would be problematic for the purposes of this paper since these systems were at the forefront of AI in the

1980s and characterize a lot of its history. As this paper deals with AI in a historical context, a definition has to be found that includes modern AI systems as well as everything that researchers called AI at their time.

This thesis uses the definition given by Nilsson:

“Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.” [2, p. xiii]

This definition is purposely broad to enable the inclusion of everything that has been called AI. In this context, the specifics of what constitutes intelligence is less important as long as the researchers define their activities as devoted to making machines intelligent.

1.3 Defining AI Winter

The term AI winter was first used in 1984 at the AAAI (Association for the Advancement of Artificial Intelligence, before 2005: American Association for Artificial Intelligence) conference to warn of an impending AI winter [10]. Following the AAAI’s usage, in this work AI winter is defined as:

A period of declining enthusiasm after times of success in the field of AI.

Depending on the source, the field has experienced either one or two winters to date. Howe talks about the AI winter of 1973 lasting a decade[11], while Nilsson only regards the years after the mid- to late 1980s as such [2, p. 327]. Even though the term was first used in 1984, it often gets applied retroactively to the times after 1974 up until the early 1980s, as similar patterns occurred.

For this thesis, it is useful to look at both time frames to gain additional insights. The approximate time frames for the AI winters considered in this thesis are:

- First winter: 1973-1979
- Second winter: 1988-1994

The events that led to the winters and the rationale for the approximate time frames are discussed in the chapter “2. History of AI Winters”.

1.4 State of the Art Review

To formulate a novel approach for analyzing the prospect of an approaching AI winter the existing literature on the topic is reviewed. Attempts at predicting an approaching AI winter have been made plentiful though mostly in non-scientific literature with some of them gaining considerable attention [12]. These articles are often very opinionated and spend little time discussing different sides of the arguments. Additionally, the data backing up the proposed claims is often weak or non-existent. Experts in the field like Hinton[13], LeCun[14], Chollet[15], and Marcus[16] have all expressed critique on current approaches in the field.

1.5 Methodology

This work tries to analyze the problem with a more structured approach compared to previous attempts. Before discussing the present, a clear understanding of the previous AI winters has to be attained. Non-scientific literature on the matter discusses various subjects without clearly attributing relevancy beforehand. In contrast, this work analyzes the central causes leading to AI winters by examining the historical literature. Followed by a discussion of expert views of the present time with a narrow frame of focus.

2 History of AI Winters

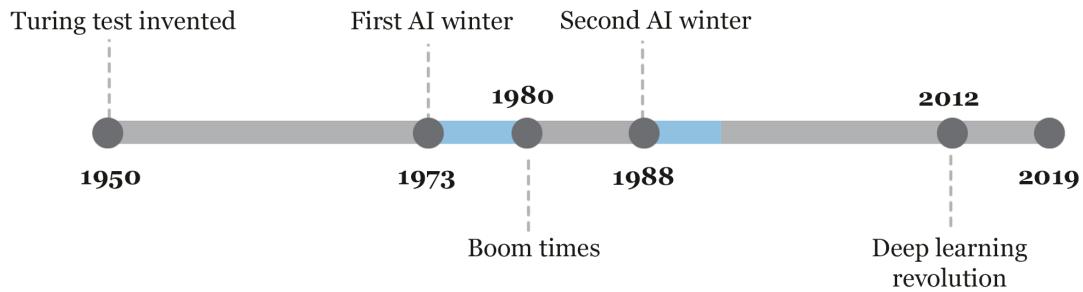


Figure 1: Timeline of the AI winters

AI has a long history. One can argue it even started long before the term was first

coined; mostly in stories and later in actual mechanical devices called automata [2]. This chapter only covers events relevant to the periods of AI winters without being too exhaustive in hope to extract knowledge that can be applied today. Afterwards, the most recent history is told to contrast it with the times leading to AI winter periods.

2.1 Events leading to the first AI Winter

To aid understanding the phenomenon of AI Winters, the events leading up to them are examined.

2.1.1 Beginnings of the AI Field in the 1950s

Many early ideas about thinking machines appeared in the late 1940s to '50s by people like Turing or Von Neumann [17]. Turing tried to frame the questions of "Can machines think?" differently and created the imitation game, now famously called the Turing Test [18].

In 1955, Arthur Samuel wrote a program that could play checkers very well. A year later, it even appeared on television. It used a combination of a tree search with heuristics and learned weights. Samuel handcrafted the heuristics inspired by a book from checkers experts. He used a learning algorithm he called "temporal-difference learning" where the weights are adjusted using the "error" between the score initially calculated and the score after the search was completed [19].

In 1954, one of the first experiments in machine translation was executed. It used a 250-word dictionary for translation combined with syntactical analysis. Translations between English and Russian were demonstrated [20]. *The New York Times* commented:

"This admittedly will amount to a crude word-for-word translation ... but will nevertheless be extremely valuable, the designers say, for such purposes as scientists translations of foreign technical papers in which vocabulary is far more of a problem than syntax [2]."

By then, the designers thought that most of the work was done with only some small errors to fix. Hutchin noted that it was the most far-reaching coverage that machine

translation has ever received [20]. For this reason, it generated tremendous hype and made it easier to obtain funding for the following work.

AI research gained much funding from U.S. Defense Establishments (ONR and ARPA, later called DARPA) in the hope that these technologies would be useful for the U.S. Navy. At the time, there was a substantial amount of enthusiasm and optimism regarding the state of AI. The field of machine translation was especially important during the cold war, as the government had a big interest in automatic translation from Russian to English. [21]

Early experiments served as an inspiration to create the Dartmouth Summer Project in 1956, where the term AI was coined. The summer project was held under the motif that “Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it” [7]. Researchers from many different fields were invited and a lot of different ideas, papers, and concepts were put forward. Though progress was made some were disappointed, for instance, McCarthy said in response to the workshop “[the] main reason the Workshop did not live up to my expectations is that AI is harder than we thought.”

In 1957, Rosenblatt invented perceptrons, a type of neural network where binary neural units are connected via adjustable weights [22]. He was inspired by the work of neuroscience in the 1940s, which led him to create a crude replication of the neurons in the brain [23].

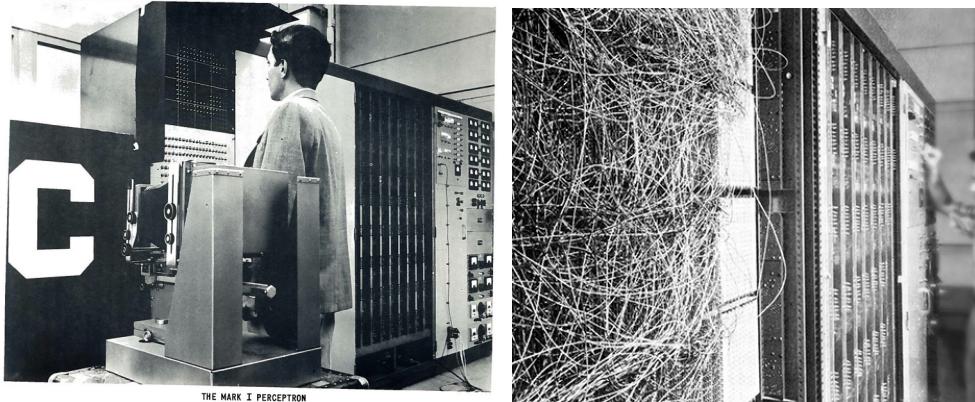


Figure 2: The Mark 1 Perceptron [24]

He tried many different layouts and learning algorithms. One type of perceptron was called series-coupled, which would in today terms refer to the standard feedforward

layout of a neural network where data flows from input to output. A prominent layout was what he called alpha perceptron, a three-layer series-coupled network where the three layers, in this case, included the input and output layer. Computers at the time would have been too slow to run the perceptron, so Rosenblatt built a special purpose machine with adjustable resistors (potentiometers) controlled by little motors. The apparatus was able to learn to classify different images of shapes or letters. *The New York Times* reported on the perceptron: “The Navy revealed the embryo of an electronic computer that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence [25].”

In the same year (1957), Simon summarized the current progress on AI like this:

“It is not my aim to surprise or shock you—but the simplest way I can summarize is to say that there are now in the world machines that think, that learn, and that create. Moreover, their ability to do these things is going to increase rapidly until—in a visible future—the range of problems they can handle will be coextensive with the range to which the human mind has been applied.” [26, Chap. 4]

2.1.2 The Quiet Decade

After the increases in funding and enthusiasm for machine translation in the 1950s and early 1960s, progress stalled. Hutchins called the period of 1967 to 1976 the quiet decade of machine translation [27, Chap. 5]. Bar-Hillel said that machine translation was not feasible. He demonstrated that computers would need too much information about the world for correct translation[28], which he thought was “utterly chimerical and hardly deserves any further discussion.” The Automatic Language Processing Advisory Committee concluded in 1964 that there was no immediate or predictable prospect of useful machine translation [29]. A report in 1966 concluded that “there has been no machine translation of general scientific text, and none is in immediate prospect,” which led to a cut in funding for all academic translation projects [27, Chap. 4].

The disappointments in machine translation were especially detrimental to the field of AI because the U.S. Defense Establishment hoped for usable systems to emerge, but similar patterns were noticed in other fields of AI too. In 1965, Dreyfus paralleled AI with alchemy. He studied the achievements of several fields at the time and concluded: “An overall pattern is taking shape: an early, dramatic success based on

the easy performance of simple tasks, or low-quality work on complex tasks, and then diminishing returns, disenchantment, and, in some cases, pessimism.” [30]

The quiet decade of machine translation was one of the most significant events that gave rise to the first AI winter. Another was the fall of the connectionist movement, which is discussed in the next section.

2.1.3 Fall of Connectionism

In 1969, Minsky and Papert’s book *Perceptrons* was published. It was a harsh critique on Rosenblatt’s perceptrons. Minsky and Papert proved that perceptrons could only be trained to solve linear separable problems [31]. For instance, one of the most dooming examples of non-linear separable problems is the exclusive OR (XOR). For the network to successfully solve the XOR problem, its output can only be true if one of its inputs is true, but not both. This represented a big hit for the connectionists, who believe AI could be best achieved by mimicking the brain. Minsky and Papert knew that multiple layers would be able to solve this problem, but there was no algorithm to train such a network. It took 17 years until such an algorithm, now known as backpropagation, was devised [32]. Only later on, was it discovered that the backpropagation algorithm had been discovered before. Indeed, it turned out that the backpropagation algorithm had been invented before *Perceptrons* was even published [33].

2.1.4 The Lighthill report and its Consequences

The Lighthill report (published in 1973) was an evaluation of the current state of AI at that time written for the British Science Research Council. The report came to the conclusion that the promises of AI researchers were exaggerated: “in no part of the field have discoveries made so far produced the major impact that was then promised.”[34] Though it pointed out that the most disappointing area of research had been machine translation, “... where enormous sums have been spent with very little useful result...” James Lighthill, the author of the report, thought that the failure to defeat the “combinatorial explosion” was at the heart of the issue. The term combinatorial explosion he refers to is a well-known problem in search spaces, like trees, where the number of nodes increases exponentially when going down the tree. For example, in a game like chess, Shannon demonstrated that the number of possible games increases from 20 with the first move to 400 with the second move,

and by the 5th move there are already 4,865,609 possibilities [35], thus representing a combinatorial explosion.

There was a lot of criticism of the Lighthill report at the time and even a debate filmed for the BBC that unfortunately never got televised. [36] Comparisons to other fields of science were made and comments made that one should not expect results that fast [37]. Though the comments were arguably correct, the report had an effect. After the report, the UK government cut funding for all but two universities involved in research in this field, and it started a wave that swept throughout Europe and even had an impact on the U.S. [38] [11] [2, p. 204]

2.1.5 The first AI Winter

Several circumstances combined to create the first AI winter. In the beginning, enthusiasm grew quickly about the potential of this new field with high amounts of optimistic press coverage. Then, disappointments in machine translation created a quiet era. Followed by Minsky and Pappert putting forward obstacles that impeded the progress of perceptrons. Finally, resulting in the instruction to create a realistic evaluation of the field, the Lighthill report came. With the arrival of the Lighthill report, the first winter started around 1973. The report had an effect on funding thus research on AI became difficult. [2, p. 204] DARPA (Defense Advanced Research Projects Agency) started funding more applied AI projects and less fundamental work. The AI winter lasted for a few years, but in the early 1980s, the field of AI experienced another high.

2.2 Events Leading to the Second AI Winter

2.2.1 AI Boom Times

After the effects of the first AI winter had begun to decline, a new era of AI began to start. This time a lot more effort was focused on creating commercial products. Additionally, large conferences, like AAAI started in the early 1980s and experienced a rapid increase in tickets sold. The general industry and government officials alike started showing a renewed interest in AI technology.

At the heart of the commercialization of AI were expert systems. These systems were handcrafted by surveying experts and creating “if-then” rule sets accordingly. This

method is called the “top-down” approach to AI with many believing that expert knowledge was the best way to create AI. These systems were implemented in fields like financial planning, medical diagnosis, geological exploration, and microelectronic circuit design.[39]

The magazine *Business Week* joined the hype and published the headline “AI: It’s Here” in 1984. Similarly, many companies made extraordinary claims like: “We’ve built a better brain” and declared that “[I]t is now possible to program human knowledge and experience into a computer … Artificial intelligence has finally come of age.” [39]

2.2.2 Fears of an upcoming Winter

As the hype regarding AI increased, researchers feared that the field might not deliver the expected results. At a panel called “The Dark Ages of AI - Can we avoid or survive them?” at the 1984 AAAI conference scientists discussed if an upcoming AI winter could be prevented. [10]

“This unease is due to the worry that perhaps expectations about AI are too high, and that this will eventually result in disaster. I think it is important that we take steps to make sure the AI winter doesn’t happen [...].”

The fear was that funding would once again dry up, when unrealistic expectations could not be fulfilled. This fear proved to be correct.

2.2.3 The second AI Winter

In the following years, the claims of what AI systems were capable of slowly had to face reality. The expert systems at the center of the revolution faced many issues. In 1984, John McCarthy criticized expert systems because they lacked common sense and knowledge about their own limitations [40]. He described the expert system MYCIN built to assist physicians. He then laid out a situation where a patient has *Cholerae Vibrio* in his intestines. When asked, the systems prescribed two weeks of tetracycline. This would most likely kill off all the bacteria, but by then the patient would already be dead. Additionally, many tasks were too complicated for engineers

to design rules around them manually. Systems for vision and speech contained too many edge cases[41].

Schwarz, Director of DARPA ISTO (Defense Advanced Research Projects Agency/Information Science and Technology Office) from 1987 to 1989 concluded that AI research has always had “... very limited success in particular areas, followed immediately by failure to reach the broader goal at which these initial successes seem at first to hint...”. This led to a decrease in funding in AI research [2].

The general interest in AI declined as the expectations could not be met. At this time, many AI companies closed their doors. The AAAI conference that attracted over 6000 visitors in 1986 quickly decreased to just 2000 by 1991 [42]. Similarly, a decrease in AI-related articles starting in 1987 and reaching its lowest point in 1995 can be observed in *The New York Times*[43].

2.3 AI Spring and Modern History

As the AI winter emerged in the late 1980s and continued to the early 1990s, there was a paradigm shift in the field of AI. Expert systems that dominated the field gave way to more machine-learning-based methods. In the 2000s, many different machine-learning-based methods, like support vector machines, Bayesian networks, Markov-based methods, evolutionary algorithms, and neural networks were used. [41]

In 2005, the DARPA Grand Challenge was launched with a 2 million dollar prize. The challenge was to build an autonomous vehicle that could navigate a 132 mile course in the desert of Nevada. A team from Stanford University under the leadership of Sebastian Thrun won the challenge. *The New York Times* commented:

“Until recently, progress in artificial intelligence lagged so far behind computing technology that some in the field talked about an “AI winter”, after commercial and government funding evaporated in the mid-1980s. Now there’s talk about an AI spring among researchers like Sebastian Thrun, the director of the Stanford lab.”[44]

Percentage of papers that mention each method

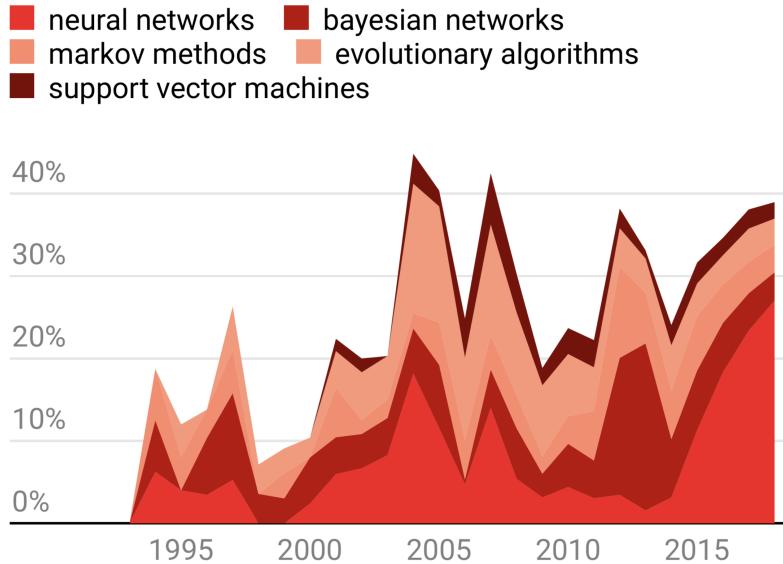


Figure 3: Data from 16,625 papers from arXiv. [41]

2.3.1 The Deep Learning Revolution

Around 2012, the field of machine learning had several breakthroughs. The yearly occurring Image-Net competition was won by a neural network for the first time, and by a wide margin [45]. Similarly, impressive results were achieved in different domains, like speech recognition [46]. This sparked a revolution, and machine learning using deep-learning was at the forefront of this new period of AI [47, p. 440] [48, p. 129]. Deep learning refers to the use of neural networks with two or more layers [47, p. 436].

While the algorithm used to train most of these networks, backpropagation, was popularized in 1986 [32], the data and computing requirements for it to work well were not achieved until much later [47, p. 438]. To elaborate, the amount of data created every day by the Internet, smartphones, and IoT-devices enables deep learning techniques to flourish [49, p. x]. In 2002, the amount of data created per year was estimated to be about 5 exabytes [50], while today the global data sphere is estimated to be 33 zettabytes large, and the International Data Corporation (IDC) predicts it to grow to 175 zettabytes by 2025, which means an estimated increase of 10 zettabytes between 2019 and 2020 [51]. Coupled with the fact, Banke and Brill

famously demonstrated in 2001 that large quantities of data trump the individual differences of algorithms [52], an idea later revisited by Norvig et al. in their paper “The Unreasonable Effectiveness of Data” [53]. At the same time, graphics processing units (GPUs), which can process data in parallel, were becoming widely available, and techniques to train on them effectively were being developed [54, p. 1]. Furthermore, computer power has increased manifold as described by Moore’s law [55]. Together, the increased computing capabilities and data access have reached a critical mass and allowed deep neural networks to outperform classical approaches in many different fields. Breakthroughs in computer vision [56][57], speech recognition [58], game-playing [59], and many more [33, p. 25] have been achieved.

3 Analysis: What has Led to the AI Winters?

In this section, the central causes of the AI winters are extracted from the above discussion of previous winters.

First, a recurring pattern that can be observed are that promises that kindled initial excitement, but which later turned out to be inflated have been the leading cause for the AI winters. For instance, government funding was cut during both AI winters after honest assessments of the results compared to the promises. The progress was overestimated because AI initially led to significant improvements in various fields, very quickly. This suggested most of the work was done with only some minor problems to solve. However, as it later turned out these problems were not so minor in the first place. The Lighthill report, a primary contributor to the first AI winter stated: “in no part of the field have discoveries made so far produced the major impact that was then promised.” Similarly, the 1984 panel at AAAI expressed: “This unease is due to the worry that perhaps expectations about AI are too high [...].”

Second, the cut in funding had a major impact on research in both AI winters. In the first AI winter, the Lighthill report led to a cut of funding for all but two universities in the U.K. and further led to cuts in Europe and the U.S. In the second AI winter, funding from DARPA was reduced. Moreover, the commercial failure of many AI-related startups in the late 1980s marked the second AI winter.

Third, technological limitations, like the perceptron experienced in the 1960s, inhibited progress. The perceptron, which was first expected to soon “be conscious of its existence” could not solve the XOR-problem at that time. Similarly, limitations were faced with expert systems in the 1980s. They could not solve fundamental problems

like vision or speech and lacked common sense.

Consequently, in evaluating the likelihood of another AI winter, the following aspects should be examined closely:

1. Expectations and promises compared to the actual results;
2. Funding from governments and industries;
3. Technological limitations.

Many technologies exhibit similar patterns to those mentioned above. To further narrow the focus, it is necessary to figure out how AI deviates from other technologies. Though similar in some regard, AI appears to be very susceptible to inflated estimations and technological limitations. Some reasons why AI differs from other technologies are:

1. Intelligence is highly multidimensional:

At some point, AI researchers believed that by solving chess, the riddle of intelligence would be solved. This turned out to be wrong, because intelligence involves more than the dimension of conscious, strategic thinking. [60]. Chess is only a tiny particle in the cosmos of intelligence [61]. Researchers gave it such a central position because it turns out to be hard for humans, which leads to reason number two.

2. Moravec's Paradox

Chess, which requires higher level thinking, is a very new skill in our evolutionary history, which might be the reasons why it is relatively difficult for humans and therefore associated with Intelligence. Vision, on the other hand, is old and mainly subconscious, which leads people to believe it is easy, but there is no reason to assume it is not as hard or even more difficult than higher level thinking. This is Moravec's Paradox, and one can argue AI researchers have fallen prey to this statement by underestimating the processes we do subconsciously, like sensorimotor skills or common sense [62].

3. Hype and fear associated with achieving human-level intelligence

As I. Jordan pointed out, the hype and fear surrounding machines that are capable of achieving intelligence easily leads to exaggerations and creates media attention less common in other fields [63].

With these reasons in mind, the possibility of a coming AI winter can be analyzed with the appropriate framing.

4 Probability of an Approaching AI Winter

Subsequently, the possibility of an upcoming AI winter is assessed. The current landscape of artificial intelligence and its public reception is studied. Furthermore, the present and the historical pre-winter times are compared regarding the key areas extracted beforehand. As a recap, these areas are:

1. Expectations and promises compared to the actual results;
2. Funding from governments and industries;
3. Technological limitations.

4.1 Expectations and Promises

Many public figures are voicing claims that are reminiscent of those of early AI researchers in the 1950s. By doing this, the former group creates excitement for future progress, or hype. Kurzweil, for instance, is famous for not only predicting the singularity, a time when artificial superintelligence will be ubiquitous, will occur by 2045 [64, p. 135-136], but also that AI will exceed human intelligence by 2029 [65]. In a similar manner, Scott is predicting that “there is no reason and no way that a human mind can keep up with an artificial Intelligent machine by 2035 [66].” Additionally, Ng views AI as the new electricity [67].

Statements of this kind set high expectations for AI and spark hype. Consequently, the phenomenon of hype and how it relates to the present state of AI is investigated.

4.1.1 Hype and the Hype Cycle

A tool often used when looking at hype is Gartner’s Hype Cycle [68]. It has practical applications that let us make predictions easily, but its validity is not scientifically established. First of all, it is not a tool developed for a scientific approach. It is a

stylized graph made for business decisions. That said, attempts to empirically validate the Hype Cycle for different technologies have been made. It can be concluded that the Hype Cycle exists, but that its specific patterns vary a lot [69].

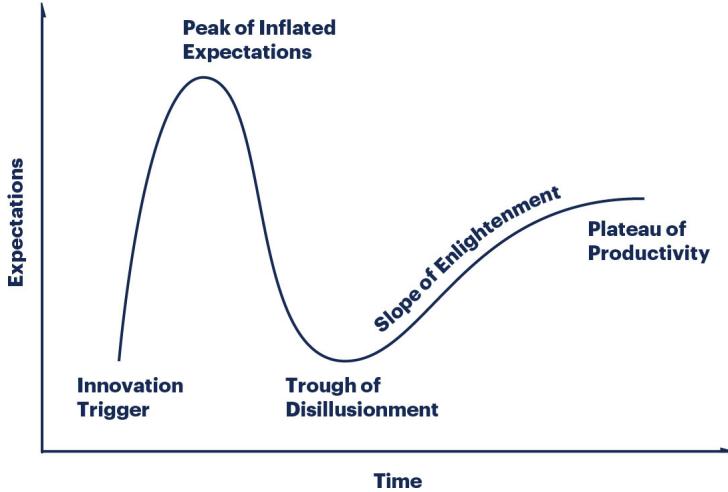


Figure 4: Gartner’s Hype Cycle. [68]

The key phases of the cycle are the peak, where interest and excitement are at their highest, and the trough of disillusionment, where the initial expectation cannot be met. Here, interest in the field is at its lowest. Then, the field slowly recovers and reaches the plateau of productivity.

As Menzies demonstrates, the Hype Cycle is well represented in the AAAI conference attendee numbers in the 1980s [42]. First, the conference started with a rapid increase in ticket sales leading to a peak, and then those numbers quickly dropped down. Currently, conference attendee numbers for conferences like NIPS reach or even exceed the peak of AAAI in the 1980s, and they are quickly gaining in size [1].

Similar patterns for interest in the field can be observed in venture capital funding for AI startups, job openings, and earning calls mentions [1]. Researchers of hype point out that the quantity of coverage is important, but that it has to be supported by qualitative sentiments. Sentiment analysis in media articles shows that AI-related articles became 1.5 times more positive from 2016 to 2018. Especially in the period from January 2016 to July 2016, the sentiment shifted [1]. This improvement could be correlated with the public release of Alpha Go in January 2016, and its victory against world champion Lee Sedol in March.

Following the trend of the Hype Cycle, this could lead to another trough of disillusionment.

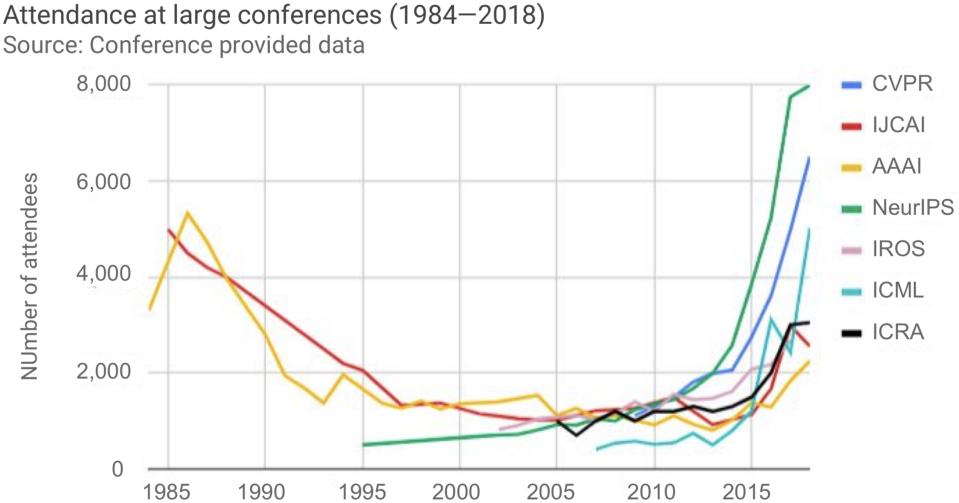


Figure 5: [1]

sionment with ticket sales, funding, and job openings quickly plummeting. However, AI is a very broad term describing many technologies. This further complicates the matter, as each technology under the umbrella term AI can have its own Hype Cycle and the interactions of Hype Cycles both with each other and to AI, in general, remain unclear.

Going further, a more in-depth look into these claims is made, evaluating if the quick rise in AI interest is just the consequence of exaggerated promises or if the claims stand on firm ground.

4.1.2 Comparison to Expert Opinion

Now, the statements and promises made by public figures are compared to a survey of leading AI researchers. In 2017, a survey of 352 machine learning researchers, who published at leading conferences, was conducted. This survey forecasts high-level machine intelligence to happen within 45 years at a 50% chance and at a 10% chance within the next nine years. However, full automation of labor was predicted much later, with a 50% probability for it to happen within the next 122 years [70].

This study presents results far from the predictions of futurists like Kurzweil. Further, a meta-study on AI predictions has found some evidence that most predictions of high-level machine intelligence are around 20 years in the future no matter when the

prediction is made [71, p. 1]. In essence, this points to the unreliability of future predictions of AI. Moreover, every prediction of high-level machine intelligence has to be viewed with a grain of salt.

In summary, a Hype Cycle pattern is present in the current AI landscape, leading to a potential decline in interest soon. Furthermore, optimistic predictions are made by public figures, but empirical evidence questions their validity.

Nevertheless, statements like those from Ng, who views AI as the new electricity, refer more to the current state of the industry. Accordingly, the industry and governments funding is examined next.

4.2 Investment and Funding

Funding has always had a significant role in AI research. As Hendler points out, cuts in government funding are only felt years later, since existing research programs continue [38]. Thus, time passes until the lack of new research programs becomes evident. This means that a reduction in funding would need to be in place presently to be perceived in the years to come.

In April 2018, EU members agreed to cooperate on AI research. A communication on AI was issued that dedicated 1.7 billion dollars of funding for AI research between 2018 and 2020 [72]. Then, in June 2018, the European Commission proposed the creation of the Digital Europe funding program with a focus in five key areas and total funding of 9.2 billion euros, of which 2.5 billion is dedicated to AI research [73].

In March 2018, the U.S. Administration stated the goal of ensuring that the United States “remains the global leader in AI[74].” Later, in September 2018, DARPA announced a two billion dollar campaign to fund the next wave of AI technologies [75]. In direct opposition, China has declared the goal of leading the world in AI by 2030. Consequently, several Chinese AI initiatives have been launched. These conflicting statements have motivated many to adopt the term, “AI Race,” to refer to the battle of leadership in the AI realm between the U.S. and China [76]. It is similar to the Space Race in the 20th century between the U.S. and the Soviet Union with the countries fighting for dominance in space travel. Back then, the race sparked much funding and research [77]. Likewise, the “AI Race” mentality could make any reduction in funding unlikely in the next years. This is a strong point against an upcoming AI winter, as previous winters were accompanied by a decline of government funding.

Another key point is the growing AI industry. Past AI researchers have been very reliant on government funding but, according to McKinsey & Company, non-tech companies spent between \$26 billion and \$39 billion on AI in 2016, and tech companies spent between \$20 billion and \$30 billion on AI.

The market forecasts for 2025, on the other hand, have an enormous variance ranging from \$644 Million to \$126 billion [3, p. 6]. This disparity demonstrates the economic difference between an upcoming AI winter and another period of prosperity.

To summarize, government funding is very solid, and the “AI Race” mentality makes it likely that this situation will continue. Additionally, the industry is currently thriving. Market forecasts, however, diverge drastically.

To determine which forecasts are more convincing, the progress AI has made in the last years is related to the criticisms of current approaches.

4.3 Evaluating Progress

To view the criticisms of present AI techniques in the appropriate frame, the progress that has been made since 2012 until today (April 2019) is evaluated.

As we have seen before, AI and machine learning have risen in popularity across many measurements [1]. A few key events stand out in the shaping of the landscape. In 2012, a convolutional neural network won the ImageNet competition by a wide margin [45]. This, combined with progress in object detection, changed the field of computer vision completely from handcrafted feature-engineering to learned representations, thereby enabling autonomous cars to become viable in the foreseeable future. Similarly impressive results have been made in the natural-language understanding space. Deep learning has enabled all the popular voice assistants, from Alexa and Siri to Cortana [33].

Reinforcement learning with deep neural networks has had impressive results in game-playing. In 2014, DeepMind used a deep-q-learner to solve 50 different Atari games, without changing the model’s architecture or hyperparameters [78]. This flexibility in tasks was unprecedented, which led to them being acquired by Google soon after and subsequently leading the space of reinforcement learning with achievements like AlphaGo [79], and AlphaStar [80].

Finally, in the last few years, generative adversarial networks (GAN) have achieved

impressive results in generating images of e.g. human faces [81].

In essence, deep learning has had groundbreaking results across many industries.

4.4 The Criticism of Deep Learning

In this chapter, criticisms of deep learning are discussed. As demonstrated, deep learning is at the forefront of progress in the field of AI, which is the reason a skeptical attitude towards the potential of deep learning is also criticism of the prospects of AI in general. It is a similar situation to the 1980s, when expert systems dominated the field and their collapse led to a winter period. If deep learning methods face comparable technological obstacles as their historical counterpart, similar results can be expected.

There are a few categories that have been identified in which most criticisms of deep learning fall: limitations of deep learning, brittleness, and lack of unsupervised learning.

4.4.1 Limitations of deep learning

“Today more people are working on deep learning than ever before – around two orders of magnitude more than in 2014. And the rate of progress as I see it is the slowest in 5 years. Time for something new.”

François Chollet, creator of Keras on Twitter [15]

As this quote is taken from Twitter, its validity is questionable, but it seems to fall in line with similar arguments he has made and it captures the general feeling well. In his book “Deep Learning with Python,” Chollet has a chapter dedicated to the limitations of deep learning, in which he writes: “It [deep learning] will not solve the more fundamental problem that deep learning models are very limited in what they can represent, and that most of the programs that one may wish to learn cannot be expressed as a continuous geometric morphing of a data manifold.” As a thought experiment, he proposes a huge data set containing source code labeled with a description of the program. He argues that a deep learning system would never be able to learn to program in this way, even with unlimited data, because tasks like these require reasoning and there is no learnable mapping from description to source

code. He further elaborates that adding more layers and data make it seem like these limitations are vanishing, but only superficially [82].

He argues that practitioners can easily fall into the trap of believing that models understand the task they undertake. However, when the models are presented with data that differs from data encountered in training data, they can fail in unexpected ways. He argues that these models don't have an embodied experience of reality, and so they can't make sense of their input. This is similar to arguments made in the 1980s by Dreyfus, who argued for the need of embodiment in AI [83]. Unfortunately, a clear understanding of the role of embodiment in AI has not yet been achieved [84]. In a similar manner, this points to fundamental problems not yet solved with deep learning approaches, namely reasoning and common sense.

In short, Chollet warns deep learning practitioners about inflating the capabilities of deep learning, as fundamental problems remain.

4.4.2 Deep Learning is brittle

The common terminology used to describe deep learning models is brittle. There are several examples of why such a description would be accurate, including adversarial attacks, lack of ability to generalize, and lack of data. A detailed discussion of these flaws and eventual prevention mechanisms is pursued.

1. Adversarial attacks: It has been demonstrated that deep learning algorithms can be susceptible to attacks via adversarial examples. Adversarials use data modified in a non-recognizable way for humans to affect the behavior of deep learning models drastically. There are multiple methods for creating adversarial examples. In one technique, noise is added to the image by another learning algorithm in order to affect the classification, without being visible [85].

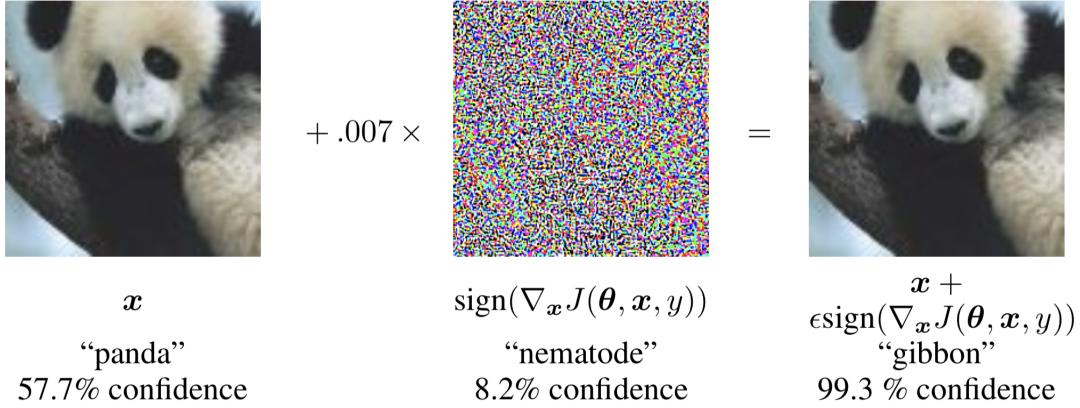


Figure 6: A demonstration of fast adversarial example generation applied to GoogLeNet on ImageNet [85].

With this technique, it is possible for an image to be changed in such a way that a specified classification can be achieved, even if it is very different like from the original classification (like „panda“ and „gibbon“ which humans can easily distinguish). When the method of adversarial attack is known, it can be possible to defend against it by augmenting the training set with adversarial examples [85]. To clarify, defending against a specific adversarial attack can be possible, but protecting against adversarials, in general, is hard. Nonetheless, there are successful methods recently developed that show promise in the issue. A formal way to defend against general adversarials has been used by bounding the output space of the model. Techniques like interval bound propagation have state-of-the-art accuracy in different popular image sets [86].

Alcorn et al. point out that extreme misclassification also happens when familiar objects appear in strange poses [87]. Examples like these reveal that deep learning models, understanding of objects can be quite naive.



Figure 7: The Google Inception-v3 classifier fails to recognize familiar objects in strange poses [87].

Furthermore, adversarial attacks demonstrate an underlying problem that is more profound - the lack of explainability. Due to the black box nature of deep learning models, predicting what the network is doing is hard. These adversarial attacks show that a model may have found an optimal way to classify an object in the training data, but it may still fail to capture the vastness of the real world.

That said, there has been much work in improving the interpretability of models, mostly in the vision space through methods like semantic dictionaries [88], saliency maps, [89][90] and activation atlases [91]. These works represent attempts to gain insights into the hidden layers of deep learning models.

2. Lack of ability to generalize: Furthermore, deep learning models have problems generalizing beyond the training data provided. Kansky et al. demonstrated that a model trained on the Atari game Breakout failed when small changes were made to the environment [92]. For example, changing the height of the paddle slightly resulted in very poor performance of the agent. Similar criticism can be applied to any reinforcement learning system.

Cobbe compares the evaluation of reinforcement learning agents with supervised learning and concludes that evaluating an agent in the environment he trained in is like evaluating the performance of a supervised learner with the test set. The difference is that the first case is well accepted and practiced, and the second one is not tolerated in any sense.

To solve this problem, Cobbe, as part of OpenAI, devised a benchmark for generalization to promote work in this area [93]. Additionally, transfer learning in the domain

of reinforcement learning has recently seen impressive results with OpenAI’s Dota agent. They announced that they were able to continue training on the agent despite substantial changes in rules and model size by using a transfer learning technique. Using similar methods, the lack of generalization in agents could be improved [94].

3. Lack of data: As “The Unreasonable Effectiveness of Data” demonstrates, data is essential in deep learning. Moreover, the rise in available data was one of the main contributors to the deep learning revolution. At the same time, not every field has access to vast amounts of data.

That said, there are two ways to tackle this problem: by creating more data or by creating algorithms that require less data. Lake et al. show that humans are able to learn visual concepts from just a few examples [95]. Recent approaches in one-shot



Figure 8: Test yourself on one shot learning. From the example boxed in red, can you find the others in the array? On the left is a Segway and on the right is the first character of the Bengali alphabet. [95]

or few-shot learning, where an algorithm is only presented with one or a few data points e.g., one image of a given category, have made substantial improvements [96]. At the same time, transfer learning approaches have improved immensely. By using a model pre-trained on a large data set as a basis, it is possible to significantly reduce training time on new data sets [97].

To summarize, deep learning models are fairly described as brittle. That said, researchers are working on promising solutions to this problem.

4.4.3 The Dominance of Supervised Learning

Most achievements realized by deep learning have been through supervised or reinforcement learning. However, as LeCun points out, humans mostly learn in an unsupervised manner by observing the environment [98] [14]. Additionally, approximate estimates predict that around 95 percent of data is unstructured [99]. Moreover, labeling is a time-consuming and expensive process, but labels only contain very little information about each data point. This is why LeCun believes the field has to shift more to unsupervised learning.

A particular type of unsupervised learning, sometimes called self-supervised learning, has gained traction in the last couple of years [100]. Self-supervised learning procedures exploit some property of the training data to create a supervision signal. In a video clip, for example, all frames are sequential, and researchers exploit this property by letting the model predict the next frame of the clip, which can easily be evaluated because the truth is inherent in the data [101]. Similar methods can be used for text or audio signals. Additionally, different characteristics of the data can be used such as rotating an image and predicting the correct angle [102]. The intuition is that in order to turn a rotated image back to its original form, a model needs to learn properties about the world that would also be useful in different tasks like object recognition. This proves to be correct, as this model can achieve great results in classification tasks via transfer learning. When looking at the first layer of the network, the filters are very similar to supervised models and even more varied.

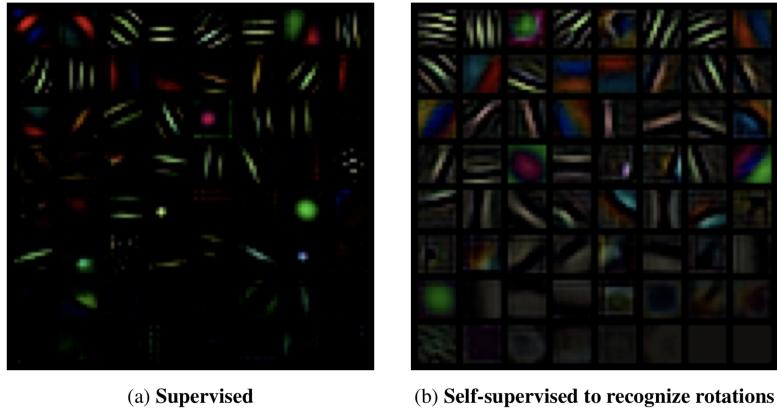


Figure 9: First layer filters learned by a AlexNet model trained on (a) the supervised object recognition task and (b) the self-supervised task of recognizing rotated images [102]

This criticism could be detrimental to deep learning and AI in general, if researchers dismissed it, but it seems not to be the case. OpenAI has presented some promising results, achieved by using unsupervised learning earlier this year, with the GPT-2 transformer language model [103]. This model can generate very human-like texts by using a very large model and vast amount of data from Reddit. They used a type of self-supervised learning by exploiting the sequentiality of text and letting the model predict the next word. Using the same architecture, MuseNet, a model that composes music, has been created recently [104].

Unsupervised learning has the potential to solve significant obstacles in deep learning. Current research evidence suggests optimism regarding the progress of this learning technique.

5 Conclusion

A complex interplay is present between AI researchers, companies, the technology and the perception of AI on many different levels. Therefore, making any prediction is hard. However, there are a few key things we can observe about the field that differ from historical pre-winter times.

In the past, reliance on government funding was very strong and the industry weak. That is far from the case today; many large companies like Google, Facebook, and Alibaba are investing more in AI technologies alone than the AI industry was worth during its boom times in the 1980s. Even more importantly, those companies did not only invest heavily in AI, but they also incorporated it heavily into their products. This provides the field a solid foot on the ground, even when public sentiment starts to shift. Similarly, stability is provided by the “AI Race” mentality, which reduces the risk of a decline in funding from governments.

Equally important are the criticisms regarding deep learning and its limitations. Though most of it is valid, the evidence suggests that researchers are already working on solutions or are aware of the innate limitations of the technique.

Furthermore, unsupervised learning, especially self-supervised learning presents promising opportunities by enabling the use of vast amount of unlabeled data and by saving immense amounts of tedious labor.

That said, the expectations of the field are too high. Predictions about machines

reaching human intelligence are unreliable. Furthermore, a Hype Cycle pattern can be exhibited in current conference attendee numbers, with the field growing quickly on many scales. As Hype Cycle patterns vary, no certain statements or predictions can be made.

Finally, the historical perspective demonstrates the wave-like nature of the field. New technologies are being created every day; a vast amount of them die out; and some are being revived. In this light, it seems adequate to be prepared for current methods to die out, as well as to be on the lookout for some forgotten technology worth reviving.

To summarize: The funding for further AI research appears stable at the moment. However, there are some technological limitations which may, coupled with very high expectations, lead to another AI winter.

“People worry that computers will get too smart and take over the world, but the real problem is that they’re too stupid and they’ve already taken over the world.” [105] **Pedro Domingos**

References

- [1] Y. Shoham, R. Perrault, E. Brynjolfsson, J. C. Openai, J. Manyika, J. C. Niebles, T. Lyons, J. Etchemendy, B. Grossz, and Z. Bauer, “AI Index Report”, pp. 1–94, 2018. [Online]. Available: <http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf>.
- [2] N. J. Nilsson, *The Quest for Artificial Intelligence*. 2010, ISBN: 9780511819346. DOI: [10.1017/cbo9780511819346](https://doi.org/10.1017/cbo9780511819346).
- [3] M. G. Institute, “Artificial Intelligence: the Next Digital Frontier?”, *McKinsey Global Institute*, pp. 1–80, 2017. [Online]. Available: <https://www.mckinsey.com/~media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx>.
- [4] D. W. Massaro and R. L. Gregory, “The Oxford Companion to the Mind”, *The American Journal of Psychology*, 2006, ISSN: 00029556. DOI: [10.2307/1423066](https://doi.org/10.2307/1423066).
- [5] J. E. Laird, R. E. Wray III, R. P. Marinier III, and P. Langley, “Claims and Challenges in Evaluating Human-Level Intelligent Systems”, 2009. DOI: [10.2991/agi.2009.17](https://doi.org/10.2991/agi.2009.17).
- [6] P. D. Thorn, “Nick Bostrom: Superintelligence: Paths, Dangers, Strategies”, *Minds and Machines*, 2015, ISSN: 0924-6495. DOI: [10.1007/s11023-015-9377-7](https://doi.org/10.1007/s11023-015-9377-7).
- [7] J. McCarthy, M. L. Minsky, and C. E. Shannon, “A proposal for the Dartmouth summer research project on artificial intelligence - August 31, 1955”, *Ai Magazine*, vol. 27, no. 4, pp. 12–14, 2006, ISSN: 2371-9621. DOI: [10.1609/aimag.v27i4.1904](https://doi.org/10.1609/aimag.v27i4.1904). [Online]. Available: <https://bit.ly/2GLzIwK>.
- [8] B. J. Grossz, R. Altman, E. Horvitz, A. Mackworth, T. Mitchell, D. Mulligan, and Y. Shoham, “Standing Committee of the One Hundred Year Study of Artificial Intelligence”, 2016.
- [9] A. Kaplan and M. Haenlein, *Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*, 2019. DOI: [10.1016/j.bushor.2018.08.004](https://doi.org/10.1016/j.bushor.2018.08.004).
- [10] D. Mcdermott, M. M. Waldrop, R. Schank, B. Chandrasekaran, and J. Mcdermott, “The Dark Ages of AI: A Panel Discussion at AAAI-84”, *Ai Magazine*, vol. 6, no. 3, pp. 122–134, 1985. [Online]. Available: <https://bit.ly/2GLzIwK>.

- [11] J. Howe, “Artificial Intelligence at Edinburgh University: A perspective”, 2007.
- [12] F. Piekniewski, *AI winter is well on its way*, 2018. [Online]. Available: <https://blog.piekiewski.info/2018/05/28/ai-winter-is-well-on-its-way/>.
- [13] S. LeVine, *Artificial intelligence pioneer says we need to start over*, 2017. [Online]. Available: <https://www.axios.com/artificial-intelligence-pioneer-says-we-need-to-start-over-1513305524-f619efbd-9db0-4947-a9b2-7a4c310a28fe.html>.
- [14] Y. Lecun, “Deep Learning Hardware : Past, Present and Future”, Tech. Rep., 2019, pp. 53–59. [Online]. Available: <https://krzakala.github.io/cargese.io/LeCun.pdf>.
- [15] F. Chollet, *Tweet - François Chollet*, 2018. [Online]. Available: <https://twitter.com/fchollet/status/1039176548719312896?lang=de>.
- [16] G. Marcus, “Deep Learning: A Critical Appraisal”, pp. 1–24, 2018. [Online]. Available: <https://arxiv.org/pdf/1801.00631.pdf>.
- [17] J. Neumann, “General and Logical Theory of Automata”, in *Cerebral Mechanisms in Behavior: The Hixon Symposium*, 1951.
- [18] A. M. Turing, “Computing machinery and intelligence”, in *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, 1950, ISBN: 9781402067105.
- [19] A. L. Samuel, “Some Studies in Machine Learning”, *IBM Journal of Research and Development*, 1959.
- [20] J. Hutchins, “The first public demonstration of machine translation : the Georgetown-IBM system, 7th January 1954”, *MT News International*, no. 8, pp. 15–18, 2006. [Online]. Available: <http://www.hutchinsweb.me.uk/GU-IBM-2005.pdf>.
- [21] J. Hutchins, *Encyclopedia of Language & Linguistics, Second Edition*. Oxford: Elsevier, 2006, vol. 7, pp. 375–383.
- [22] F. Rosenblatt, *The Perceptron a percieving and recognizing automaton*, 1957. [Online]. Available: <https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf>.
- [23] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain”, *Psychological Review*, 1958, ISSN: 0033295X. DOI: [10.1037/h0042519](https://doi.org/10.1037/h0042519).
- [24] F. Rosenblatt, “Perceptrons and the Theory of Brain Mechanics”, *Cornell AERONAUTICAL LAB INC BUFFALO NY.*, vol. VG-1196-G, p. 621, 1961.

- [25] *NEW NAVY DEVICE LEARNS BY DOING - Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser*, Jul. 1958. [Online]. Available: <https://www.nytimes.com/1958/07/08/archives/new-navy-device-learns-by-doing-psychologist-shows-embryo-of.html>.
- [26] P. Norvig, *Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp*. 1992, p. 962, ISBN: 1558601910.
- [27] J. Hutchins, “Machine Translation: A Brief History”, 1995, ISSN: 00978507. DOI: [10.2307/417787](https://doi.org/10.2307/417787). [Online]. Available: <http://hutchinsweb.me.uk/ConcHistoryLangSci-1995.pdf>.
- [28] Y. Bar-Hillel, “The Present Status of Automatic Translation of Languages”, *Advances in Computers*, 1960, ISSN: 00652458. DOI: [10.1016/S0065-2458\(08\)60607-5](https://doi.org/10.1016/S0065-2458(08)60607-5).
- [29] C. Automatic Language Processing Advisory, “Language and machines-computers in translation and linguistics”, *Nas-nrc Publication No. 1416. 1966. National Academy Of Sciences-national Research Council, Washington, D.c. 124 P.*, 1966, ISSN: 13132296.
- [30] H. Dreyfus, “Alchemy and Artificial Intelligence”, 1965, [Online]. Available: <https://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf>.
- [31] M. M. Papert and S., “Perceptrons. An Introduction to Computational Geometry.”, in *Science*, 1969, ISBN: 9780262630221.
- [32] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors”, *Nature*, 1986, ISSN: 00280836. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0).
- [33] J. Schmidhuber, *Deep Learning in neural networks: An overview*, 2015. DOI: [10.1016/j.neunet.2014.09.003](https://doi.org/10.1016/j.neunet.2014.09.003).
- [34] I. Literature and R. Lighthill, “Lighthill Report - Artificial Intelligence: A General Survey”, pp. 1–18, 1972.
- [35] C. E. Shannon, “Programming a Computer for Playing Chess”, *Philosophical Magazine*, 1950. [Online]. Available: <https://vision.unipv.it/IA1/aa2009-2010/ProgrammingaComputerforPlayingChess.pdf>.
- [36] *BBC TV - June 1973 - Lighthill Controversy Debate at the Royal Institution*, 1973. [Online]. Available: <http://www.aiai.ed.ac.uk/events/lighthill1973/>.

- [37] J. McCarthy, "Review of "Artificial Intelligence: A General Survey""", 2000. [Online]. Available: <http://www-formal.stanford.edu/jmc/reviews/lighthill.pdf>.
- [38] J. Hendler, "Avoiding Another AI Winter", *IEEE Intelligent Systems*, 2008, ISSN: 1541-1672. DOI: [10.1109/mis.2008.20](https://doi.org/10.1109/mis.2008.20).
- [39] G. Johnson, "Thinking About Thinking", *APF Reporter*, vol. 8, no. 1, pp. 1–6, 1985. [Online]. Available: <https://web.archive.org/web/20090412093552/https://aliciapatterson.org/APF0801/Johnson/Johnson.html>.
- [40] J. McCarthy, "Some Expert System Need Common Sense", pp. 1–11, 1984. [Online]. Available: <https://ieeexplore.ieee.org/document/1200723>.
- [41] K. Hao, "We analyzed 16,625 papers to figure out where AI is headed next", *MIT Technology Review*, pp. 1–19, 2019. [Online]. Available: <https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next>.
- [42] T. Menzies, "21st-Century AI: Proud, not smug", 2003. [Online]. Available: <http://proquest.umi.com/pqdweb?index=67&did=353949121&SrchMode=1&sid=6&Fmt=2&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1259545262&clientId=23824>.
- [43] E. Fast and E. Horvitz, "Long-Term Trends in the Public Perception of Artificial Intelligence", no. January 1986, pp. 963–969, 2016. [Online]. Available: <http://arxiv.org/abs/1609.04904>.
- [44] J. Markoff, *AI reemerges from a funding desert*, 2005. [Online]. Available: <https://www.nytimes.com/2005/10/13/business/worldbusiness/ai-reemerges-from-a-funding-desert.html>.
- [45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", 2012. [Online]. Available: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
- [46] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors", pp. 1–18, 2012. [Online]. Available: <http://arxiv.org/abs/1207.0580>.
- [47] G. E. Hinton, Y. Lecun, and Y. Bengio, "Deep learning", *Nature*, vol. 13, no. 1, p. 35, 2015, ISSN: 15487105. DOI: [10.1038/nmeth.3707](https://doi.org/10.1038/nmeth.3707).
- [48] T. Sejnowski, *The Deep Learning Revolution*. 2016, ISBN: 9780262038034.
- [49] E. Alpaydin, *Machine Learning - The New AI*. The MIT Press Essential Knowledge series, 2016.

- [50] “How Much Information?”, 2003. [Online]. Available: <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm>.
- [51] D. Reinsel, J. Gantz, and J. Rydning, “The Digitization of the World From Edge to Core Mankind is on a quest to digitize the world”, no. November, 2018.
- [52] M. Banko and E. Brill, “Scaling to very very large corpora for natural language disambiguation”, 2001. DOI: [10.3115/1073012.1073017](https://doi.org/10.3115/1073012.1073017). [Online]. Available: <https://www.aclweb.org/anthology/P01-1005>.
- [53] A. Halevy, P. Norvig, and F. Pereira, “The Unreasonable Effectiveness of Data”, *IEEE Intelligent Systems*, 2009, ISSN: 1541-1672. DOI: [10.1109/mis.2009.36](https://doi.org/10.1109/mis.2009.36).
- [54] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, “Revisiting Unreasonable Effectiveness of Data in Deep Learning Era”, *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 843–852, 2017, ISSN: 15505499. DOI: [10.1109/ICCV.2017.97](https://doi.org/10.1109/ICCV.2017.97).
- [55] G. E. Moore, “Cramming More Components Onto Integrated Circuits, Electronics, April 19, 1965”, *Electronics*, 1965, ISSN: 0018-9219. DOI: [10.1109/JPROC.1998.658762](https://doi.org/10.1109/JPROC.1998.658762).
- [56] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, 2015. [Online]. Available: <http://arxiv.org/abs/1506.02640>.
- [57] C. Szegedy, A. Toshev, and D. Erhan, “Deep Neural Networks for Object Detection”, in *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 2013, ISBN: 978-1-4799-5118-5. DOI: [10.1109/CVPR.2014.276](https://doi.org/10.1109/CVPR.2014.276).
- [58] A. Graves, A. R. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks”, in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2013, ISBN: 9781479903566. DOI: [10.1109/ICASSP.2013.6638947](https://doi.org/10.1109/ICASSP.2013.6638947).
- [59] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing Atari with Deep Reinforcement Learning”, pp. 1–9, 2013. [Online]. Available: <http://arxiv.org/abs/1312.5602>.
- [60] N. Ensmenger, “Is chess the drosophila of artificial intelligence? a social history of an algorithm”, *Social Studies of Science*, vol. 42, no. 1, pp. 5–30, 2012, ISSN: 03063127. DOI: [10.1177/0306312711424596](https://doi.org/10.1177/0306312711424596).
- [61] P. Norvig, “Artificial intelligence: Early ambitions”, *New Scientist*, vol. 216, no. 2889, pp. ii–iii, 2012, ISSN: 02624079. DOI: [10.1016/s0262-4079\(12\)62783-3](https://doi.org/10.1016/s0262-4079(12)62783-3).

- [62] F. Tuck and H. Moravec, *Mind Children: The Future of Robot and Human Intelligence*. 1988, pp. 15–16. doi: [10.2307/1575314](https://doi.org/10.2307/1575314).
- [63] M. Jordan, *Artificial Intelligence — The Revolution Hasn't Happened Yet*, 2018. [Online]. Available: <https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7>.
- [64] R. Kurzweil, *The Singularity is Near - When Human Trascends Biology*. 2005, ISBN: 0143037889.
- [65] ——, *How to create a mind: the secret of human thought revealed*. 2012. doi: [10.5860/choice.50-6167](https://doi.org/10.5860/choice.50-6167). [Online]. Available: <https://bit.ly/2GtcYS9>.
- [66] A. Brown, *11 Quotes About AI That'll Make You Think*, 2019. [Online]. Available: <https://www.techopedia.com/11-quotes-about-ai-thatll-make-you-think/2/33718>.
- [67] S. Lynch, *Andrew Ng: Why AI Is the New Electricity*, 2017. [Online]. Available: <https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity>.
- [68] A. Linden and J. Fenn, “Understanding Gartner’s Hype Cycles Understanding Gartner’s Hype Cycles Understanding Gartner’s Hype Cycles”, no. May, 2003. [Online]. Available: <https://www.bus.umich.edu/KresgePublic/Journals/Gartner/research/115200/115274/115274.pdf>.
- [69] H. Van Lente, C. Spitters, and A. Peine, “Comparing technological hype cycles: Towards a theory”, *Technological Forecasting and Social Change*, 2013, ISSN: 00401625. doi: [10.1016/j.techfore.2012.12.004](https://doi.org/10.1016/j.techfore.2012.12.004).
- [70] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans, “When Will AI Exceed Human Performance ? Evidence from AI Experts”, pp. 1–21, 2017. [Online]. Available: <https://arxiv.org/pdf/1705.08807.pdf>.
- [71] S. Armstrong and K. Sotala, “How We’re Predicting AI – or Failing to”, pp. 11–29, 2014. [Online]. Available: <https://intelligence.org/files/PredictingAI.pdf>.
- [72] E. Commission, “Artificial Intelligence for Europe”, Tech. Rep., 2018. [Online]. Available: <https://ec.europa.eu/transparency/regdoc/rep/1/2018/EN/COM-2018-237-F1-EN-MAIN-PART-1.PDF>.
- [73] E. Commission, T. E. Commission, and D. Europe, “EU budget : Commission proposes 9.2 billion investment in first ever digital programme”, Tech. Rep. June, 2018. [Online]. Available: http://europa.eu/rapid/press-release_IP-18-4043_en.pdf.

- [74] The White House Office of Science and Technology Policy, “Summary of 2018 White House Summit on Artificial Intelligence”, 2018. [Online]. Available: <https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf>.
- [75] DARPA, *DARPA Announces \$2 Billion Campaign to Develop Next Wave of AI Technologies*, 2018. [Online]. Available: <https://www.darpa.mil/news-events/2018-09-07>.
- [76] S. Cave, D. Seán, and S. Óhéigeartaigh, “An AI Race for Strategic Advantage: Rhetoric and Risks”, *Proceedings of the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*, 2018.
- [77] W. P. Barry and A. A. Siddiqi, “Challenge to Apollo: The Soviet Union and the Space Race, 1945-1974”, *The Journal of Military History*, 2006, ISSN: 08993718. DOI: [10.2307/2677603](https://doi.org/10.2307/2677603).
- [78] X. Guo, H. Lee, X. Wang, and R. L. Lewis, “Deep learning for real-time Atari game play using offline Monte Carlo tree search planning”, *Nips*, 2014. [Online]. Available: <https://web.eecs.umich.edu/~baveja/Papers/UCTtoCNNsAtariGames-FinalVersion.pdf>.
- [79] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, “Mastering the game of Go with deep neural networks and tree search.”, *Nature*, vol. 529, no. 7587, pp. 484–489, 2016, ISSN: 1476-4687. DOI: [10.1038/nature16961](https://doi.org/10.1038/nature16961). [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26819042>.
- [80] *AlphaStar: Mastering the Real-Time Strategy Game StarCraft II*, 2019. [Online]. Available: <https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/>.
- [81] T. Karras, “Progressive Growing of Gans for improved Quality, Stability, and Variation”, pp. 1–26, 2018. [Online]. Available: https://research.nvidia.com/sites/default/files/pubs/2017-10_Progressive-Growing-of-karras2018iclr-paper.pdf.
- [82] F. Chollet, *The limitations of deep learning*, 2017. [Online]. Available: <https://blog.keras.io/the-limitations-of-deep-learning.html>.
- [83] H. Dreyfus, *What Computers Can't Do*. Harper-Collins, 1979.
- [84] B. Goertzel, *Post-Embodied AI*, 2004. [Online]. Available: http://www.goertzel.org/papers/PostEmbodiedAI_June7.htm.

- [85] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and Harnessing Adversarial Examples”, 2015. [Online]. Available: <https://arxiv.org/abs/1412.6572>.
- [86] S. Gowal, K. Dvijotham, R. Stanforth, R. Bunel, C. Qin, J. Uesato, R. Arandjelovic, T. Mann, and P. Kohli, “On the Effectiveness of Interval Bound Propagation for Training Verifiably Robust Models”, 2018. [Online]. Available: <http://arxiv.org/abs/1810.12715>.
- [87] M. A. Alcorn, Q. Li, Z. Gong, C. Wang, L. Mai, W.-S. Ku, and A. Nguyen, “Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects”, 2018. [Online]. Available: <http://arxiv.org/abs/1811.11553>.
- [88] C. Olah, A. Satyanarayan, I. Johnson, S. Carter, L. Schubert, K. Ye, and A. Mordvintsev, “The Building Blocks of Interpretability”, *Distill*, 2018. doi: [10.23915/distill.00010](https://doi.org/10.23915/distill.00010).
- [89] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps”, pp. 1–8, 2013. [Online]. Available: <http://arxiv.org/abs/1312.6034>.
- [90] P.-J. Kindermans, S. Hooker, J. Adebayo, M. Alber, K. T. Schütt, S. Dähne, D. Erhan, and B. Kim, “The (Un)reliability of saliency methods”, pp. 1–12, 2017. [Online]. Available: <http://arxiv.org/abs/1711.00867>.
- [91] S. Carter, Z. Armstrong, L. Schubert, I. Johnson, and C. Olah, “Activation Atlas”, *Distill*, 2019. doi: [10.23915/distill.00015](https://doi.org/10.23915/distill.00015).
- [92] K. Kansky, T. Silver, D. A. Mély, M. Eldawy, M. Lázaro-Gredilla, X. Lou, N. Dorfman, S. Sidor, S. Phoenix, and D. George, “Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics”, 2017. [Online]. Available: <http://arxiv.org/abs/1706.04317>.
- [93] K. Cobbe, O. Klimov, C. Hesse, T. Kim, and J. Schulman, “Quantifying Generalization in Reinforcement Learning”, 2018. [Online]. Available: <http://arxiv.org/abs/1812.02341>.
- [94] OpenAI, *How to Train Your OpenAI Five*, 2019. [Online]. Available: <https://openai.com/blog/how-to-train-your-openai-five/>.
- [95] B. M. Lake, R. Salakhutdinov, J. Gross, and J. B. Tenenbaum, “One shot learning of simple visual concepts”, 2011, ISSN: 0031-3831. [Online]. Available: <http://palm.mindmodeling.org/cogsci2011/papers/0601/paper0601.pdf>.

- [96] F. Sung, Y. Yang, and L. Zhang, “Learning to Compare : Relation Network for Few-Shot Learning”, 2018. [Online]. Available: <https://arxiv.org/abs/1711.06025>.
- [97] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning”, *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, ISSN: 0278-0062. DOI: [10.1109/TMI.2016.2528162](https://doi.org/10.1109/TMI.2016.2528162).
- [98] *Self-Supervised Learning*, Yann LeCun, Facebook AI Research, 2018. [Online]. Available: <https://news.dartmouth.edu/events/event?event=53673#.XMXOTC1XbKI>.
- [99] A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics”, *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015, ISSN: 02684012. DOI: [10.1016/j.ijinfomgt.2014.10.007](https://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007). [Online]. Available: <http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007>.
- [100] A. Zisserman, C. Doersch, I. Misra, A. Owens, C. Vondrick, and R. Zhang, “Self-Supervised Learning Slides”, 2018. [Online]. Available: <https://project.inria.fr/paiss/files/2018/07/zisserman-self-supervised.pdf>.
- [101] R. Child, S. Gray, A. Radford, and I. Sutskever, “Generating Long Sequences with Sparse Transformers”, 2019. [Online]. Available: <http://arxiv.org/abs/1904.10509>.
- [102] S. Gidaris, P. Singh, and N. Komodakis, “Unsupervised Representation Learning by Predicting Image Rotations”, no. 2016, pp. 1–16, 2018. [Online]. Available: <http://arxiv.org/abs/1803.07728>.
- [103] A. Radrof and J. Wu, “Language Models are Unsupervised Multitask Learners”, 2018. [Online]. Available: https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.
- [104] C. Payne, *MuseNet*, 2019. [Online]. Available: <https://openai.com/blog/musenet/>.
- [105] J. Langston, *A Q & A with Pedro Domingos: Author of ‘The Master Algorithm’*. [Online]. Available: <https://www.washington.edu/news/2015/09/17/a-q-a-with-pedro-domingos-author-of-the-master-algorithm/>.

Many thanks for your attention!