INSTITUTE FOR DEFENSE ANALYSES

Deep Learning: Measure Twice, Cut Once

Robert F. Richbourg

June 2018

Approved for public release; distribution is unlimited.

IDA Document NS D-9138
Log: H 18-000245

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under CRP C8235, "Limitations of Artificial Intelligence for Current Operations." The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Deep Learning: Measure Twice, Cut Once

Robert F. Richbourg
Institute for Defense Analyses
Alexandria, VA
rrichbou@ida.org

ABSTRACT

Many in the industrial and defense communities are expecting current artificial intelligence technologies (deep learning and deep neural networks) to solve a wide array of problems. Others are deeply concerned that adversaries investing heavily in these technologies will produce highly autonomous and adaptive weapons that will overmatch any known defenses. This reaction is not surprising given that deep neural networks and deep learning systems have been remarkably successful at tasks long believed to require high levels of (human) intelligence. These technologies are enjoying great success because of two enabling developments. The availability of large amounts of appropriately labeled training data and the continued growth in sheer computing power permit the decades-old neural network technologies to reach surprising performance levels. These success stories beg answers to questions on the limits of performance and potential. This paper describes artificial intelligence in its historical context of boom and bust cycles. The AI discipline has a 60-year record of heightened expectations fueled by remarkable achievement that were soon followed by disillusionment ("AI Winters") when the technologies failed to generalize to wider application. The paper also develops parallels between the current deep neural network requirements for success and those of previous intelligent technologies that were once inspiring but have now been largely retired. Finally, deep neural network technologies have known limitations that should be publicized along with their success stories to frame and temper expectations. The paper promotes awareness of these limitations to foster a rational appreciation for potential. These artificial intelligence technologies can certainly contribute to advancing automated capabilities, but their contribution is not without limit, so careful planning and preparation should precede action.

ABOUT THE AUTHOR

Robert Richbourg, Ph.D. is a member of the Research Staff at the Institute for Defense Analyses. He is a retired Army officer who holds a BS in Mathematics, and MS and Ph.D. degrees in Computer Science (artificial intelligence). His final 10-year assignment of Army active duty was as an Academy Professor of Computer Science and Director of the Office of Artificial Intelligence Analysis and Evaluation at the United States Military Academy, West Point. He has over 20 years of M&S experience including serving as chair of the I/ITSEC Tutorial Board, the I/ITSEC Simulation Subcommittee, the I/ITSEC Fellows Committee, and multiple SISO leadership positions.

Deep Learning: Measure Twice, Cut Once

Robert F. Richbourg
Institute for Defense Analyses
Alexandria, VA
rrichbou@ida.org

INTRODUCTION

> **"…every aspect of learning or any other feature of intelligence can in principle be so precisely stated that a machine can be made to simulate it."**

Figure 1. Words from the Dartmouth AI Conference Call for Papers [Moor, 2006]

Recent advances in the technologies recognized as parts of the artificial intelligence (AI) discipline have been incredible. IBM's Watson system[1] has beaten recognized champions on *Jeopardy*! and Google's AlphaGo easily defeated a human international Go champion.[2] These achievements have helped to establish great expectations for what can be done with current AI technologies. The Department of Defense is crafting an AI strategy that will seek to utilize AI in a wide range of application areas [Serbu, 2018]. Defense investments are being planned [Knapp, 2018]. The USMC *Project Maven* [Allen, 2017] has been a highly-visible success and there are calls for great expansion [Corrigan, 2017]. The media is talking about a potential "AI Arms Race" [Cohen, 2017] and there are those in the Pentagon's "think tanks" who worry that US adversary's AI achievements will give them unique leverage, unconstrained by legal and ethical concerns so ingrained in US thinking [Stewart, 2017].

This newfound eagerness to capitalize on the potential of AI technologies is only new in the context of recent history. Consider the quote presented in Figure 1. This was written in late 1955 for a conference held in the summer of 1956 that marked the birth of artificial intelligence as a recognized discipline. Even though this quote is more than 60 years old, it would not be out of place in a much more recent announcement. The Dartmouth Conference was held at a time of seminal AI achievements that seemingly foretold continued advances. Examining the historical development of AI reveals that the discipline has seen two major "boom and bust" periods during its first 45 years. During boom periods, there are great expectations for potential achievements. Expectations are fueled by spectacular (for the period) demonstrations of capability. Government programs launch. Investment and activity follow apace. The media react with sensational stories. The bust periods follow when the capabilities that seemed so promising fail to generalize. Government programs cancel. Investment stops. The media go silent or critical. Activity slows.

The following discussion offers a brief and somewhat generalized recount of major AI events during the last 60 years. The purpose is to characterize underlying cause and effect for the boom and bust periods. This historical background provides a context for understanding current developments. It seems clear that AI is enjoying a third boom period; however, there are corollaries to past boom periods, both from social and technological perspectives. The paper concludes with thoughts on how best to avoid a complete repeat of history so that a third bust period does not follow. Perhaps the third time really is the charm?

AI RISING

The years immediately before and after the 1956 Dartmouth conference witnessed developments that still influence the artificial intelligence discipline today. As early as 1949, Arthur Samuel began research in machine learning that eventually enabled his work in teaching computers to play board games. By the late 1950s, he had developed a computer program that could defeat human checkers players [Samuel, 2000]. At the time, that sparked incredible

[1] See https://www.ibm.com/midmarket/us/en/article_Smartercomm5_1209.html
[2] See http://www.nature.com/articles/nature16961

Figure 2. Man versus Machine in Checkers

reaction: a computer defeats a human in a game that most feel requires both intelligence and strategic thinking. Today, we are seeing similar reactions to deep learning algorithms that have defeated human champions at the far more difficult game of Go.

Just before the Dartmouth Conference, Allen Newell developed a computer program that could print an "image" of a map, using the printer's characters (letters, digits, punctuation marks) as symbols. This achievement fundamentally influenced one of Newell's Rand colleagues, Herbert Simon. Simon realized that computers were far more than fast calculators; they were symbol manipulation devices and such devices could be used to simulate decision making and other aspects of human intelligence. Subsequently, Newell and Simon devised a computer program, the *Logic Theorist*, [Stefferud, 1963] that eventually developed logical proofs of 38 of the 52 fundamental theorems in *Principia Mathematica* [Whitehead and Russell, 1963]. In the mid-1950s, a computer program that could independently produce proofs of mathematical theorems appeared to demonstrate that computers could simulate human intelligence! Newell and Simon went on to develop the Physical Symbol System Hypothesis: "A physical symbol system has the necessary and sufficient means for general intelligent action" [Newell and Simon, 1976]. In their view, both a computer and the human mind are such symbol systems. This hypothesis still underpins much of the work in artificial intelligence.

Another fundamental early development focused on enabling computers to communicate using the English language. The ELIZA program was intended to simulate the conversational style of a Rogerian psychotherapist [Weizenbaum, 1966]. ELIZA could accept human conversational input and separate the main words to fit them into predefined response templates using a simple set of rules. This was some of the first work in natural language processing, an area that has progressed impressively and is still important today.

Yet another important development during this period was intended to model the animate brain's neural structure and processing as an attempt to duplicate human-like activity. The Perceptron [Rosenblatt, 1958] could be trained and could learn to recognize characters and other suitably-encoded images.[3] In fact, *The New York Times* reported that the US Navy sponsored Perceptron research to produce "the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence" [Olazaran, 1996]. Again, the period saw computers that demonstrated human-like performance, this time recognizing images and characters. Rosenblatt's Perceptron laid the foundation for the (greatly improved) artificial neural networks that now enable deep learning. The Perceptron researchers did not necessarily ascribe to the Physical Symbol System hypothesis. Instead, they have been described as following the Connectionist theories of human intelligence—the animate brain consists of a huge number of connected neurons and together, they can produce intelligence. This fundamental split between symbolic and connectionist approaches to AI work remains today.

There were other developments as well during this early period. Those cited above are representative, but they cover several predominant themes. Researchers were attempting to enable machines to perform tasks associated with human intelligence. These included perception, natural language communication, formal (symbolic) reasoning, learning, and the strategic inference used in gaming. The early collection of impressive (for the period) achievements provoked large reaction. Government funding (both in the United States and abroad) started to flow into artificial intelligence research. Industry embraced developmental work, using neural networks in signal processing applications as an example. The media became energized and sensationalized many of the developments; machines were reported to be on the brink of achieving human levels of intelligence.[4] Expectations for artificial intelligence were widespread and soaring.

THE FIRST AI Winter

[3] See the video on Perceptron training at https://www.youtube.com/watch?v=7BtLqqJVP9w
[4] See the "Thinking Machine" video at https://www.youtube.com/watch?v=aygSMgK3BEM

When the bloom falls from the rose, little more than thorns remain. The checkers playing program used a form of rote learning. It attempted to record every possible board position and an associated score (likelihood of winning). This approach did not extend to more complicated games; rote learning simply did not scale up to more general cases. Newell and Simon followed their *Logic Theorist* effort with the *General Problem Solver* (GPS). This was an attempt to use their previous ideas to create a much more general purpose computer program. While GPS eventually showed some success, "it could only solve simple problems; and those, less efficiently than special-purpose problem solvers" [Barr, 1981]. It was not a "general" problem solver at all because, much like the checkers effort, the approach did not scale to general-case problems.

The Perceptron effort also ran into difficulties. Even though these systems had seen early success, they had severe limitations. Researchers from MIT, staunch advocates of the symbolic approach, proved that the Perceptron architecture was only capable of solving a simple class of "linearly-separable" problems[5] [Minsky, 1991]. This proven and widely-publicized limitation of the Perceptron nearly ended all research[6] on these and similar architectures for years.

The early successes with language processing, particularly machine translation, led to the establishment of multi-year, well-funded programs. A governmental goal was to translate Russian technical publications into English. After 10 years of effort and approximately $20 million in funding, the Automatic Language Processing Advisory Committee [ALPAC, 1966] reported to the National Academy of Sciences, "we do not have useful machine translation [and] there is no immediate or predictable prospect of useful machine translation." The ALPAC report affected machine translation efforts for the next 20 years[7] by some estimates, but certainly ended government funding for more than 10 years.[8] Other efforts in artificial research were also being questioned. The British Parliament commissioned James Lighthill to assess the general progress of artificial intelligence in the United Kingdom. The Lighthill Report[9] concluded that British artificial intelligence had achieved very little, and what had been achieved was really due to using more traditional disciplines.[10] In the United States, DARPA cut back its support for artificial intelligence research[11] following years of programs that failed to achieve ambitious, but stated, goals. The first AI Winter began in earnest. However, artificial intelligence research did not end; work carried on but at greatly reduced scale and funding.

THE RISE OF THE EXPERT SYSTEM

In the late 1970s, a new artificial intelligence technology, known as expert systems, began to emerge and showed some remarkable progress at automating human expertise. These were symbolic reasoning systems that relied on extracting and representing knowledge from human experts to duplicate their judgements[12] and conclusions in specific problem areas. The symbolic, rule-based nature ("if (X) then (Y)") of these systems also enabled them to explain chains of reasoning. The explanation capability was not only useful for decision makers but for system developers as well. A significant difference from the earlier successes was that expert systems were focused on solving very specific problems and not concerned with the nature of intelligence in general. They relied on a specific set of problem-solving techniques that were empowered by custom-built "knowledge bases" which encoded expert knowledge in a problem domain. Also, they were predominantly aimed at commercial enterprise. As an example, the Digital Equipment Corporation (DEC) was losing some revenue because most of their salesmen were not able to configure complex computer orders correctly. Eventually, DEC built an expert system (XCON) to perform order configuration. Claims

[5] See the longer explanation at https://datasciencelab.wordpress.com/2014/01/10/machine-learning-classics-the-perceptron/

[6] See https://web.csulb.edu/~cwallis/artificialn/History.htm

[7] See https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf

[8] For a general review and critique of the ALPAC report, see www.sts.rpi.edu/public_html/nirens/SergeiPapers/Readings%20in%20Machine%20Translation%20Book%20Chapters/13.pdf

[9] See www.mathrix.org/liquid/archives/the-lighthill-parliament-debate-on-general-purpose-artificial-intelligence

[10] See www.nap.edu/read/6323/chapter/11#213

[11] A concise review is available at https://en.wikipedia.org/wiki/AI_winter

[12] A long video interview with Ed Feigenbaum, the "father" of expert system technology, is available at www.youtube.com/watch?v=Uk9YA1kwZLw

were that XCON saved DEC as much as $25 million per year [Winston, 1986]. Some described these developments as: "expert systems symbolize the new wealth of nations; knowledge is power" [Feigenbaum, 1977].

Again, tremendous optimism ensued[13] and both governments and industry invested heavily.[14] New industry sprang up to facilitate expert system construction and use. In 1985, projections for expert system market share called for increasing from $80 million in 1983 to $3–$12 billion in 1990 and to $50–$120 billion by 2000—at that point, accounting for 20% of the computer industry revenues [Hu, 1987]. The US government responded in 1983 with the DARPA Strategic Computing Initiative[15]: "As a result of a series of advances in artificial intelligence, computer science, and microelectronics, we stand at the threshold of a new generation of computing technology having unprecedented capabilities….For example, instead of fielding simple guided missiles or remotely piloted vehicles, *we might launch completely autonomous land, sea, and air vehicles capable of complex, far-ranging reconnaissance and attack missions."* [emphasis added] Similarly, the Japanese government embarked on a ten-year effort designed to make Japan the global leader in knowledge information processing (applied AI) [Shapiro, 1983].

THE SECOND AI WINTER

By the late 1980s, the bloom was starting to fall from the rose once again. "Like big hairdos and dubious pop stars, the term "artificial intelligence" (AI) was big in the 1980s, vanished in the 1990s" [*Economist*, 2002]. Expert system technology proved difficult to maintain and even less promising to extend to new application areas. The US government curtailed new spending on its ambitious[16] Strategic Computing Initiative. The Japanese government revised its futuristic 5th generation computing project[17] to remove artificial intelligence-based goals. Industry that had grown out of the expert system enthusiasm to construct special purpose and highly profitable computing machinery (e.g., Symbolics Inc., Lisp Machines Inc.), failed because the return on investment for those using them was incredibly poor.[18] Other industries that provided expert system building environments and tools failed or moved into other areas such as object-oriented technology development. The artificial intelligence discipline entered its second winter and the term itself became stigmatized. It became popular to equate AI with "almost implemented" [*Economist*, 2002]. Much as in the earlier case, efforts continued, but far more slowly. As an example, the American Association of Artificial Intelligence (AAAI) conference has long been a flagship event for the community. During the boom period, submitted papers continued to grow, reaching almost 900 for the 1990 conference. By 1997, just over 300 papers were submitted. Researchers started referring to their work by more specific terms such as "machine learning," "neural networks," "decision support systems" or other descriptors not involving "AI."

There is considerable debate about the reason for the expert systems' disappearance. While a few will claim that they were subsumed into standard decision-support technologies, many others feel the problems were more fundamental. Some argue that any expert system was much like an idiot savant,[19] excelling in one tiny niche, but basically disabled in the wider context. Others note that not all forms of expertise can be quantified; there is an intuitive and creative basis [Dreyfus, 1986] that is not expressible in simple rules and facts. Others cite the great difficulty and expense of creating and maintaining the knowledge bases[20] (the set of facts and rules that provided human expertise). As an example, in seven years of use, the XCON system grew to include more than 6,200 individual rules, making any changes to the system incredibly difficult [Soloway, 1987]. All of these reasons relate to the difficulty of creating and maintaining the information necessary to support the expert system: the information was often not expansive enough

---

13 For more detail, see http://web.stanford.edu/group/scip/avsgt/expertsystems/aiexpert.html

14 See https://wiserdaily.wordpress.com/2017/02/07/history-17-artificial-intelligence/

15 See the original DARPA document at www.nitrd.gov/nitrdgroups/images/3/3a/20040929_strategic_computing.pdf

16 A review is available at www.revolvy.com/main/index.php?s=AI%20winter&item_type=topic

17 A concise description is at https://en.wikipedia.org/wiki/Fifth_generation_computer

18 Some history of Symbolics machines is available at https://danluu.com/symbolics-lisp-machines/

19 See Jim Seymour, "Debugging AI Myths," *PC Magazine*, vol. 5, no. 21 (9 Dec 1985): 95–96. https://books.google.com/books?id=XrIceI56DbsC&pg=PA95&lpg=PA95&dq=%22expert+system%22+%22idiot+savant%22&source=bl&ots=XbI76TBRn2&sig=DzNJKWX2e4VK9iWYoH35n6TMy0Q&hl=en&sa=X&ved=0ahUKEwiWo-Ci0aXZAhVytlkKHTt_AOUQ6AEIKjAB#v=onepage&q=%22expert%20system%22%20%22idiot%20savant%22&f=false

20 See http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.11.5232&rep=rep1&type=pdf

(idiot savant niche performance), too difficult to obtain, too difficult to maintain, or not quantifiable. In some ways, relying on and exploiting specific knowledge (data) both enabled experts systems and led to their demise.

THE FIRST TWO ARTIFICIAL INTELLIGENCE ERAS

The earliest days of artificial intelligence really framed the work within the discipline. Many of the application areas remain goals today. These include machine learning, game playing, natural language processing, logical reasoning, perception and sensory processing, and others. The goal of work during this time was to produce computer programs with human-like intelligence, and most of the focus was on using symbolic processing to achieve those goals. Some of the earliest efforts showed promise within constrained domains of application. Interest swelled. Governments started development efforts. The media promoted sensational claims for the future.

In the end, the early efforts failed to generalize and many of the researcher promises could not be fulfilled. Government programs cancelled. Industry moved investment elsewhere. The media noted the failures. Efforts in AI research continued, but at much smaller scale and promoting more conservative goals. But the work did continue, both in the symbolic and connectionist communities.

The second era started based on the early performance of expert systems. A major shift was the focus on the role of information and knowledge. To be sure, many types of algorithms were developed, but the new ingredient was exploiting available knowledge in specific problem domains. Expert systems performed powerfully based on restricting the problem context. Effort was not singularly focused on simulating animate intelligence, but on business areas that offered large potential return on investment. The demonstrated achievements for some specific problems fueled familiar reactions from government, industry, and the media. Expectations also met a familiar fate when difficulties with the expert system technologies came to light. This time, the principle difficulty was not only failure to generalize, but a second issue emerged as well: the difficulty of acquiring, maintaining, and extending the expert system's knowledge bases, the empowering data.

A SHIFT IN EMPHASIS AND THE EMERGENCE OF NEW DEFINITIONS

The publication of the limitations affecting Perceptrons greatly reduced the research and interest in neural networks, but work did continue and slow, steady progress was made. A seminal development in the late 1980s (at the height of the expert system fervor) showed that known limitations could be overcome and that neural networks could indeed be used to solve interesting, non-linear problems (multi-layer networks with backpropagation [Rummelhart, 1986]). This development reignited wide interest in connectionist, neural network concepts about intelligent computer processing. The IEEE organized its first conference on neural networks in 1987 and it attracted 1,800 attendees.[21] This was just the beginning of neural network advancement and the enthusiasm that continued through the second AI Winter is still being felt today.

Figure 3 presents some data on numbers of attendees at three major artificial intelligence conferences. The Association for the Advancement of Artificial Intelligence (AAAI)[22] started holding conferences in 1984. The International Joint Conference on Artificial Intelligence (IJCAI) is an international flagship event for the discipline. The Conference on Neural Information Processing Systems (NIPS) began as an invitation-only meeting in 1987 and has since become the largest conference in artificial intelligence. Figure 3 clearly shows the ascendance of AI technologies (basically expert systems) in the late 1980s, the subsequent AI Winter in the 1990s, and the rise of neural network technologies that continues today.

[21] See www.psych.utoronto.ca/users/reingold/courses/ai/cache/neural4.html
[22] Name changed from American Association for Artificial Intelligence in 2007

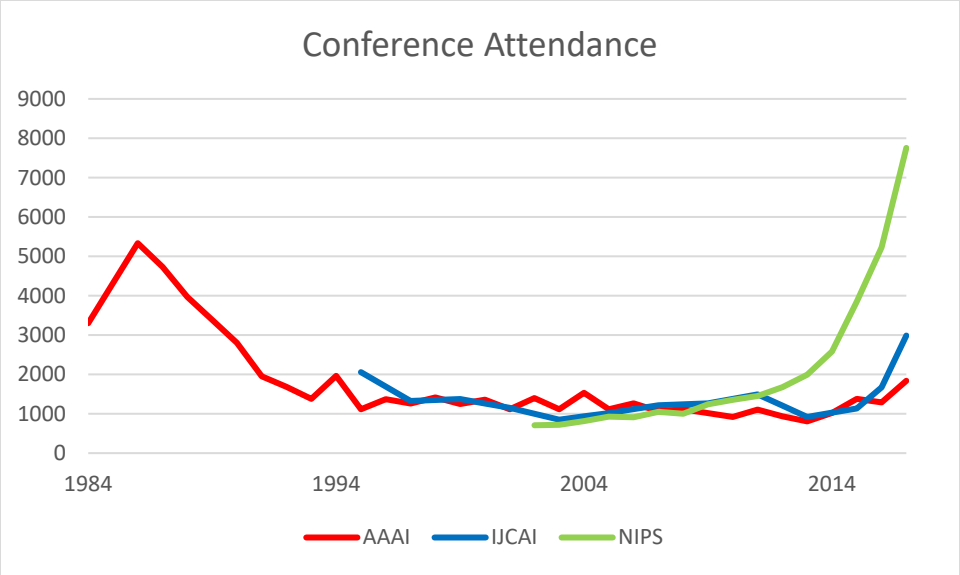

Figure 3. Attendance at Major Artificial Intelligence Conferences [23]

Figure 3 also illustrates the current dominance of neural network technologies within the discipline of artificial intelligence; today, when people talk about artificial intelligence, they are generally referring to machine learning and deep neural networks. This is also sometimes called *narrow AI*, or describing an application that excels at one task, but has limited utility given others. The efforts that were once mainstream AI, going back to the Dartmouth conference idea "that every aspect of learning or any other feature of intelligence can in principle be so precisely stated that a machine can be made to simulate it," are now described as Artificial General Intelligence (AGI). In addition to these, there is the notion of Artificial Super Intelligence, a state where machine intelligence has surpassed human intelligence (think HAL from *2001: A Space Odyssey*). Most people actively working in the discipline of artificial intelligence recognize the distinctions. However, having three "versions" of AI does produce confusion, particularly when much of the media appear to embrace the concept of artificial super intelligence. Any achievement of artificial super intelligence is far off in the future, at best [Brooks, 2017].

THE CURRENT (THIRD) BOOM FOR ARTIFICIAL INTELLIGENCE

Today, we are in a third boom period for artificial intelligence, this time fueled by some spectacular results from deep learning capabilities and architectural improvements in neural network technologies. Figure 4 depicts the rapid growth of venture capital flowing into AI-focused new-start companies.[24] There were 67 new starts in 1992 and more than 7,000 in 2016. Investment fueling growth in artificial intelligence technologies was $1–2 billion in 2010, increased to $5–8 billion just 6 years later,[25] and is projected to be as high as $35 billion in another 7 years.[26]

[23] Data from the AI Index project. See http://cdn.aiindex.org/2017-report.pdf

[24] Data from the AI Index project. See http://cdn.aiindex.org/2017-report.pdf

[25] See the McKinsey report at 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

[26] Multiple projections exist; see an example at www.top500.org/news/market-for-artificial-intelligence-projected-to-hit-36-billion-by-2025/

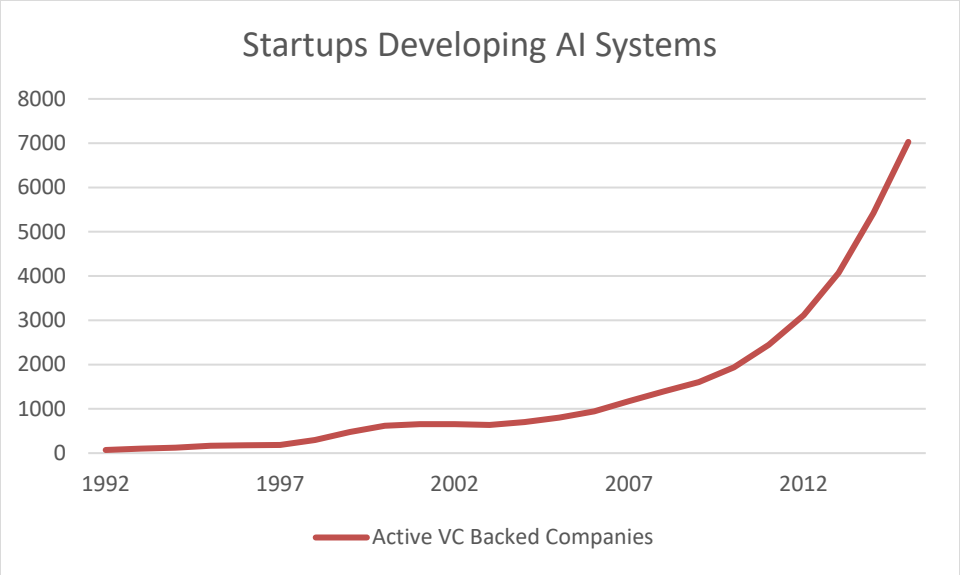

Figure 4. Venture Capital For AI New-Start Companies

Much of the recent success is enabled by two key developments. First, vastly improved and enlarged labeled training data (e.g., ImageNet[27], Project Maven [Allen, 2017]) is available to train neural networks to solve specific problems. Second, advances in computer processing power have allowed enormous networks to be built and used [Potember, 2017]. General reactions to deep learning achievements are similar to those about successes from past booms: Governments are investing heavily in the technologies; industries are rolling out hardware and software designed to make neural networks more powerful, accessible, and usable; media is sensationalizing already sensational achievements like *Watson* winning at *Jeopardy!*[28] or *AlphaGo* defeating an international Go [Silver, 2016] human champion. And again, it is not difficult to find popular media publishing predictions that artificial intelligence will overtake and perhaps destroy human society.[29] However, the history of boom and bust periods for artificial intelligence reminds us that when something seems too good to be true, that might be the case. In addition to the lessons of history, we also need to be aware of chinks in the deep-learning technological armor.

Some limitations have existed since the Perceptron model itself. The early "camps" in artificial intelligence were the symbolic reasoning group and the connectionist group. The former, then as now, believed the best approach was to build machines that reason using formal rules and logical inference. This makes the machine's reasoning processes understandable and explainable, a key factor in gaining trust from any human who might use the machine's recommendations. The connectionist camp did not argue the value of this approach, but believed that creating "biologically plausible" models of animate thought offered a better chance for success. The Perceptron was thought to be a simplified model of neurons in the human brain and much of the connectionist research then and since has been devoted to modeling human mental processes [Rich, 1991]. Modern deep neural networks are incredibly more complex than the original Perceptron and, just as with the human brain, one cannot look inside a neural network to understand how it works. The reasoning ability and knowledge representation of a large, trained network are somehow contained in the behavior of thousands of neurons, perhaps hundreds of layers deep, and all of their intricate interconnections. Today, "no one really knows how the most advanced algorithms do what they do" [Knight, 2017]. This makes explaining the recommendation from a deep learning system difficult. While the DARPA Explainable AI[30] effort is attempting improvements, it is early in that effort.[31] This deficiency is a small matter when the network is used to present advertisements to individuals or recommend music they might enjoy. The attendant risks are entirely

[27] See http://image-net.org/
[28] See www.ibm.com/midmarket/us/en/article_Smartercomm5_1209.html
[29] For a typical media reaction, see www.newsweek.com/stephen-hawking-artificial-intelligence-warning-destroy-civilization-703630
[30] Foe XAI program information, see www.darpa.mil/program/explainable-artificial-intelligence
[31] For examples of current progress, see http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

different when the algorithms are entrusted with piloting autonomous vehicles [Griggs, 2018] as a single example. How would the typical military commander react to a recommendation that could not be explained in familiar terms or recognize the need to begin failure analysis based on a flawed recommendation? This lack of explainable behavior is well known and sometimes referred to as the "black box" problem.[32]

Deep learning algorithm opaqueness also limits the ability to apply verification and validation to these systems, a fairly entrenched requirement for military applications. Testing can be applied, but testing "can be a very effective way to show the presence of bugs, but it is hopelessly inadequate for showing their absence" [Dijkstra, 1972]. A recent NASA study concluded that the AI community appears to have "neglected requirements engineering" and that "For machine learning components to be accepted by regulatory agencies this will have to change." The study was not able to provide a solution to "the complex problem of ML [machine learning] verification." At the end of the study, the author notes that the verification of machine learning algorithms "seems mostly unexplored and full of opportunities" [van Wesel, 2017]. There are other limitations as well.

Google is a leader in the modern deep learning effort. One of its leading artificial intelligence researchers, Francois Chollet, recently made a succinct observation about the deep learning technologies: "Current supervised perception and reinforcement learning algorithms require lots of data, are terrible at planning, and are only doing straightforward pattern recognition."[33] Chollet's words are an important reminder that neural network systems are, at their core, greatly improved pattern recognition systems. Problem solving with a neural network requires the problem to be formulated as a numeric pattern recognition problem, which is often difficult. There are other examples of established and powerful problem-solving technologies where problem representation can be the most important, difficult, and time-consuming requirement. Linear programming, for instance, is an excellent method for solving optimization problems, but a difficulty of using it lies in the art of problem formulation: not every problem is an optimization problem and not every optimization problem can be correctly formulated for the method.

Chollet also cites the imperative for large amounts of data to train the neural networks. You can think about training data as instances of solved problems. As an example, to be useful for training, an image must also include a label that identifies the subject of that image. What about the problem domains in which large amounts of appropriately labeled training data (instances of solved problems) are not available? The expert system knowledge base was a key to its success, and these systems floundered when problem knowledge could not be provided in a useable form. Is there a close corollary between the neural network need for a large training base ("big data") and the expert system's need for specifically encoded problem-solving information (the knowledge base)? In general, the defense industries are poster children for "tiny data." The tendency is to keep secret things out of the public eye. How many images of stealth aircraft were available before those aircraft were used?

Chollet also uses the imprecise "lots" of data when referring to the quantities of data necessary to train a network. The exact requirement for training data quantity is rarely known ahead of time (except that more is better). Neural networks can be over- or under-trained and training usually continues as long as performance improves. Thus, training is an empirical process that is subject to both "underfitting" (poor or insufficient quantity of training data) and "overfitting" (data used for training also allows the learning of noise in the inputs, which may not be present in actual data). Improper training can also result in "accidental behaviors" that have been defined as "unintended and harmful behaviors" that emerge from the machine learning systems.[34] The trial and error approach is used beyond training. As an example, there is no textbook solution linking the type of problem to basic engineering choices such as the network's number of hidden layers, filter use, or specific non-linear compression functions. The lack of an under-pinning theory also contributes to the difficulty verifying and validating deep learning systems. A neural network can assess an image and answer with "at 58 percent confidence, that image is a panda," but, again, it cannot explain how it arrived at that conclusion.

[32] See a more detailed discussion at www.nextplatform.com/2015/09/07/the-black-box-problem-closes-in-on-neural-networks/

[33] See a discussion at www.topbots.com/understanding-limits-deep-learning-artificial-intelligence/

[34] See https://arxiv.org/pdf/1606.06565.pdf

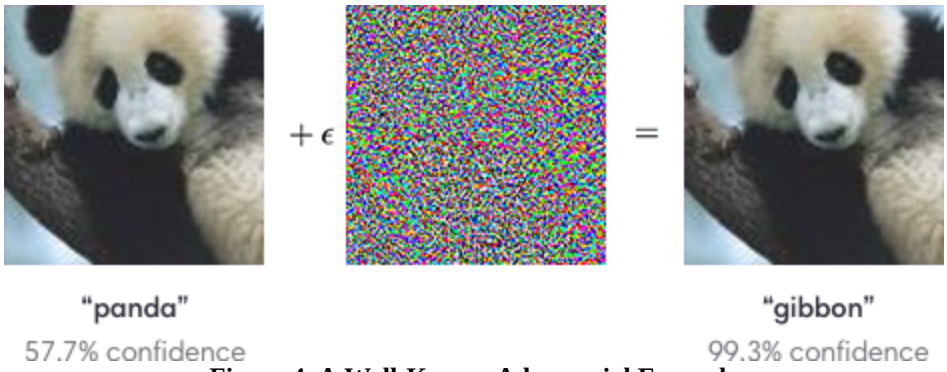

Figure 4. A Well-Known Adversarial Example

An active area of research examines the topic of "adversarial examples" [Goodfellow, 2015] in which minor changes to an input pattern can cause networks to provide wildly different results. A well-known example shows that by changing only 0.04% of the pixel values in an input image, a neural network changes its solution[35] from the correct classification "Panda with 57.7 percent confidence" to an incorrect "Gibbon with 99.3 percent confidence." A 0.04% change would be 400 pixels out of a million. This change goes undetected by the human eye. "A real intelligence doesn't break when you slightly change the requirements of the problem it's trying to solve" [Somers, 2017].

So, neural network technologies have important limitations. One might not come directly to that conclusion if only attending to reports from the popular media. The tendency towards media promoting the sensational started in the 1950s, resumed in the 1980s, and appears in vigorous health today. The result is that everyone must adopt a tempered enthusiasm bounded by healthy skepticism when the popular media publish unproven performance claims. As an example, there have been recent media reports that artificial intelligence research at Facebook resulted in computers independently inventing their own, more efficient, language to communicate with each other ("Facebook AI Creates Its Own Language In Creepy Preview Of Our Potential Future"[36]). The computers were "Bob" and "Alice." Figure 5 provides a small part of their exchange.

Bob: "I can can I I everything else."

Alice: "Balls have zero to me to me to me to me to me to me to me to me to."

Figure 5. Conversation between "Bob" and "Alice"

Claiming this exchange as exemplifying a new, more efficient language seems to be an example of the media grasping for the sensational: "When English wasn't efficient enough, the robots took matters into their own hands."[37] The researchers involved reported[38] that they do not know what the communication means and they do not understand what type of "thinking" goes on inside a neural network to produce this exchange. Given that we don't understand the meaning of the "conversation" or how it emerged from internal reasoning, the exchange between Bob and Alice seems more likely to be a programming error than anything else. In fact, Facebook ultimately changed the software to prevent excursions into language use like the above. It seems as though this event should never have been seen as newsworthy, much less reported as an incredible machine performance holding ominous future potential.

Much in the environment indicates that history is beginning to repeat. Artificial intelligence technology is progressing impressively. Investment is rising apace, if not faster. Venture capital is moving into AI. The media is energized and overstating performance. The DoD is strategizing to build a new Joint Artificial Intelligence Center (JAIC) [Tucker,

[35] See the more complete discussion at https://blog.openai.com/adversarial-example-resear

[36] See the article online at www.forbes.com/sites/tonybradley/2017/07/31/facebook-ai-creates-its-own-language-in-creepy-preview-of-our-potential-future/#60d99447292c

[37] This article is available at www.huffingtonpost.com.au/2017/08/02/facebook-shuts-down-ai-robot-after-it-creates-its-own-language_a_23058978/

[38] See discussion at www.fastcodesign.com/90132632/ai-is-inventing-its-own-perfect-languages-should-we-let-it

2018] and planning to apply resources [Knapp, 2018]. Also in Defense, there are thoughts of expanding the use of artificial intelligence into business reform, intelligence [Serbu, 2018], acquisition [Serbu, Jan 2018], training,[39] and weapon systems [Corrigan, 2017] as a few examples. Media are speculating about the new "AI Arms Race" [Cohen, 2107] and exerting pressure for the United States to respond vigorously to this modern "Sputnik" moment.[40] Will history now repeat full cycle with a third AI Winter?[41] Although there is some debate,[42] a third AI Winter is avoidable given the wide adoption of a reasoned approach that publically and programmatically recognizes both the capabilities and limitations of these technologies and shores them up when necessary. Technologies like modeling and simulation can be layered on top of deep learning capabilities to reduce the likelihood of system error, to sidestep the impact of system opaqueness, and to help explain recommendations. The new Air Force "Data to Decision" effort is using this approach.[43]

In truth, the advent of another AI Winter is not the greatest of all possible concerns. It is important, however, that progress not be slowed. China intends to become the "world leader in AI by 2030" [Churchill, 2018]. There is also some opinion that achieving AI dominance is "…likely to be coupled with a reordering of global power."[44] Russia,[45] Europe,[46] and other countries are also increasing their focus on AI technologies. Some are describing this situation as the "AI Arms Race"[47] and emphasizing the need for the United States to respond. China is advancing quickly, but the United States is the world leader in AI technologies. The path chosen to maintain that distinction could not be more important.

MEASURE TWICE, CUT ONCE

Today's successful artificial intelligence capabilities are making genuine and impressive strides forward, and are likely to continue this way in specific application areas [Potember, 2017]. In other fields, carpenters have learned, probably from unhappy experience, that careful preparation is necessary to avoid wasting resources. That lesson should also have been learned where artificial intelligence is concerned. History shows that incorrectly measuring artificial intelligence potential has led to the frustration of unmet expectations and unrewarded investment, sometimes quite large. Past AI Winters reduced interest, funding and research, thus unnecessarily slowing progress. The winters emerged when capabilities failed to scale as expected or when technology-empowering data proved too difficult to acquire (or both). Neural network technologies have been incredibly valuable in pattern-matching tasks, but it is difficult to see how they might be applied outside of that problem area so, now as before, scaling could well be a decisive issue. Neural networks have significant pre-requisites for use in pattern matching so that, as in the past, availability of data, provided as examples of solved problems, could be an obstacle. The limitations and strengths of deep learning and deep neural networks are common knowledge within the artificial intelligence community of researchers. In 2016, Arati Prabhakar, former director of DARPA, cautioned, "We have to be clear about where we're going to use the technology and where it's not ready for prime time… it's just important to be clear-eyed about what the advances, in for example, machine learning can and can't do."[48] This same appreciation needs to become more widely shared[49] by those who seek to apply "AI" to solve problems, particularly those impacting national defense. Aligning reasonable expectations with known capabilities is a key to continued, rapid progress [Richbourg, 2018].

[39] As an example, see the BAA available at http://cdmrp.army.mil/funding/pa/17dmrdpmach_pa.pdf
[40] See https://breakingdefense.com/2017/11/our-artificial-intelligence-sputnik-moment-is-now-eric-schmidt-bob-work/?_ga=2.65416942.1702442390.1509614577-220094446.1509614577
[41] For a short video discussion of the topic, see www.youtube.com/watch?v=Lmy_TAMDXdA
[42] As an example, see www.theregister.co.uk/2018/02/08/second_ai_winter/
[43] See description at www.nextgov.com/analytics-data/2018/02/us-air-forces-next-ai-project-about-kick-high-gear/145929/
[44] See www.ft.com/content/e33a6994-447e-11e8-93cf-67ac3a6482fd
[45] See www.wired.com/story/for-superpowers-artificial-intelligence-fuels-new-global-arms-race/
[46] See http://science.sciencemag.org/content/360/6388/474.1.full
[47] See www.cnn.com/2017/11/29/politics/us-military-artificial-intelligence-russia-china/index.html
[48] See https://defensesystems.com/articles/2016/05/04/darpa-chief-limits-of-artificial-intelligence.aspx
[49] See a related discussion at https://motherboard.vice.com/en_us/article/jpg4w7/elite-scientists-have-told-the-pentagon-that-ai-wont-threaten-humanity

REFERENCES

Allen, Gregory. (2017). Project Maven brings AI to the fight against ISIS. *Bulletin of the Atomic Scientists*.

ALPAC. (1966). *Languages and machines: computers in translation and linguistics.* A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington: National Academy of Sciences, National Research Council.

Barr, A. & Feigenbaum, E. (1981). *The Handbook of Artificial Intelligence, Volume 1*. William Kaufmann, Inc.

Brooks, R. (2017). The Seven Deadly Sins of AI Predictions. *MIT Technology Review.* Nov/Dec 2017.

Cohen, Zachery. (2017). US risks losing artificial intelligence arms race to China and Russia. CNN. www.cnn.com/2017/11/29/politics/us-military-artificial-intelligence-russia-china/index.html).

Corrigan, Jack. (2017). Three-Star General Wants Artificial Intelligence in Every New Weapon System. NextGov.com. www.nextgov.com/cio-briefing/2017/11/three-star-general-wants-artificial-intelligence-every-new-weapon-system/142225/

Churchill, O. (2018). China's AI dreams. *Nature: International journal of science, 553*, S10–S12.

Dijkstra, E. (1972). Turing Award Lecture. *Communications of the ACM,* 15, 859–66.

Dreyfus, H. and Dreyfus, S. (1986). *Mind Over Machine*, New York: The Free Press.

Economist. (2002). AI by another name. *The Economist.com.* www.economist.com/node/1020789

Feigenbaum, E.A. (1977). *The Art of Artificial Intelligence: Themes and Case Studies*. Conference Proceedings of the International Joint Conference on Artificial Intelligence, 1014–29.

Gass & Chapman (eds). (1985). *Theory and Application of Expert Systems in Emergency Management Operations*. Proceedings of a Symposium at the Department of Commerce, Washington, DC, 24–25 April 1985.

Goodfellow, I., *et al*. (2015). *Explaining and Harnessing Adversarial Examples*, Cornell University Library. Published online, https://arxiv.org/abs/1412.6572.

Griggs, T. & Wakabayashi. (2018). How a Self-Driving Uber Killed a Pedestrian in Arizona. *The New York Times*, 21 March 2018.

Hu, S.D. (1987). *Expert Systems for Software Engineers and Managers*, New York: Springer Science and Business Media.

Knapp, Brandon. (2018). Here's where the Pentagon wants to invest in artificial intelligence in 2019. *Defense News*.

Knight, W. (2017). The Dark Secret at the Heart of AI. *MIT Technology Review, May/June 2017.* Published online, www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.

Liebowitz, Jay. (1997). Worldwide Perspectives and Trends in Expert Systems: An Analysis based on the Three World Congresses on Expert Systems. *AI Magazine, 18*, 115–19.

Minsky, M. & Papert, S. (1972). *Perceptrons: An Introduction to Computational Geometry*, Cambridge: The MIT Press.

Moor, James. (2006). *The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years*. American Association for Artificial Intelligence, Winter 2006, 87–91.

Newell, A. & Simon, H.A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM*, 19, 113–26.

Olazaran, M. (1996). A Sociological Study of the Official History of the Perceptrons Controversy. *Social Studies of Science*, 26, 611–59.

Potember, R. (2017). Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD. JASON, The Mitre Corporation. JSR-16-Task-003, Jan 2017.

Rich, E. & Knight, K. (1991). *Artificial Intelligence*, second ed., McGraw-Hill.

Richbourg, R. (2018). It's Either a Panda or a Gibbon: AI Winters and the Limits of Deep Learning. *War on the Rocks.com*. Published online, https://warontherocks.com/2018/05/its-either-a-panda-or-a-gibbon-ai-winters-and-the-limits-of-deep-learning/.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*, 386–408.

Rummelhart, D., Hinton, E. & Williams, R. (1986). Learning representations by back propagating errors. *Nature, 323*, 533–36.

Samuel, A. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *44*, 206–26.

Serbu, J. (2018). Defense Innovation Board to tackle DoD's software acquisition problems, using software. Federal news Radio, 30 Jan 2018.

Serbu, Jason. (2018). DoD strategy for AI has implications ranging from intel to business reform. Federal news Radio.

Shapiro, E. (1983). The fifth generation project—a trip report. *Communications of the ACM*, 26, 637–41.

Silver, D., *et al.* (2016). Mastering the game of Go with deep neural networks and tree search. *Nature 529*, 484–89.

Somers, J. (2017). Is AI Riding a One-Trick Pony? *MIT Technology Review, Nov/Dec 2017*.
Soloway, E., *et al.* (1987). Assessing the Maintainability of XCQN-in-RIME: Coping with the Problems of a VERY Large Rule-Base. *Proceedings of AAAI-87*, 824–29
Stefferud, E. (1963). The Logic Theory Machine: A Model Heuristic Program. Rand Memorandum, RM-3731-CC.
Stewart, Phil. (2017). China racing for artificial intelligence military edge over US. *Financial Review*.
Tucker, P. (2018). The Pentagon is Building an AI Product Factory. *Defense One, 19 April*.
van Wesel, P. (2017). Challenges in the Verification of Reinforcement Learning Algorithms. NASA/TM-2017-219628.
Weizenbaum, J. (1966). ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*, 36–45.
Whitehead, A.N. & Russell, B. (1963). *Principia Mathematica*, Cambridge: Cambridge University Press.
Winston, P. & Pendergrast, K. (1986). *The AI Business: The Commercial Use of Artificial Intelligence*, Cambridge: MIT Press.