2 WHAT IS ARTIFICIAL INTELLIGENCE?

2. What Is Artificial Intelligence?

2.1. Definitions

"By far the greatest danger of Artificial Intelligence," AI theorist Eliezer Yudkowsky has observed, "is that people conclude too early that they understand it."[20] As such, it is important to first properly delineate what AI is – or can be. To do so, we must first grasp what *general* intelligence refers to.

2.1.1. Intelligence

Intelligence is a cornerstone of the human condition – we even named ourselves after it.[21] Intelligence is what allows you to understand (derive meaning from) the sentences you are reading here. It is a mental activity that we, as human beings, are uniquely qualified to exercise. Different elements of and processes within your brain, making use of other parts of your biology (e.g. your eyes), work together to make possible the act of reading and comprehension.

Our understanding of the etymological roots of the word intelligence is somewhat murky, but we do know it is closely related to the activity you are now engaging in – to read and understand. The word consists of two parts: the prefix *inter*, meaning 'between' and the Latin verb *legere*, meaning initially 'to choose, to pick out' and from there evolving into the meaning 'to read'.[22] Intelligence thus implies to gather, to collect, to assemble or to choose, and to form an impression, thus leading one to finally understand, perceive, or know.

Already in the 15th century, the word intelligence was understood as "superior understanding, sagacity, quality of being intelligent".[23] The sense of "information received

20 Eliezer Yudkowsky, "Artificial Intelligence as a Positive and Negative Factor in Global Risk," *Global Catastrophic Risks* 1 (2008): 303., pg 1. The article also appears in the excellent compilation Nick Bostrom and Milan M. Cirkovic, *Global Catastrophic Risks*, 1 edition (Oxford; New York: Oxford University Press, 2011)

21 The term 'homo sapiens' means 'wise man'. For a wonderfully witty tongue-in-cheek comment by Google's Director of Engineering, Open Source and Making Science on humans' narcissistic self-perception, see Chris DiBona, "The Limits of Biological Intelligence," in *What to Think about Machines That Think: Today's Leading Thinkers on the Age of Machine Intelligence*, ed. John Brockman (HarperCollins, 2015)..

22 Douglas Harper, "Online Etymology Dictionary: Arm," accessed September 20, 2016, www.etymonline.com/index.php?term=arm.. The Latin word in turn evolved from a proto-Indo-European root *leg- (1), meaning "to pick together, gather, collect". Reading, therefore, has been etymologically conjectured to imply 'to pick out words'. This PIE root *leg- (1) also makes it closely related to the Latin word *logos* "word, speech, thought", which we know from the various scientific disciplines like sociology, anthropology, microbiology, etc.

23 Douglas Harper, "Online Etymology Dictionary: Intelligence," accessed September 28, 2016, http://www.etymonline.com/index.php?allowed_in_frame=0&search=intelligence.

or imparted, news" was first recorded in the mid-15th century, from which time also stems its other frequent meaning – also in diplomatic and military circles – of "secret information from spies" (1580s).[24] Today, intelligence means different things to different people, both in daily parlance and across various academic disciplines.[25]

Usage ranges from fairly quotidian comparisons of inter-human skill or markers of authority (focusing on the social function of intelligence), to extremely theoretical definitions of intelligence as just one instantiation of a universal class of "optimization processes" which, in the broadest sense, includes even the adaptive mechanisms of natural selection.[26] A comprehensive survey of over 70 definitions of 'intelligence', conducted by Legg & Hutter, argues that it is hard to settle on a single 'correct' term, however they do note that many of the most concise and precise definitions share a number of features: (1) intelligence is a property of some agent that interacts with an environment; (2) intelligence is generally indicative of that agent's ability to succeed at a particular task or stated goal; (3) there is an emphasis on learning, adaptation, and flexibility within a wide range of environments and scenarios.[27]

Combining these features, they distill a fundamental definition of intelligence, as denoting:

> *"an agent's ability to achieve goals in a wide range of environments."*[28]

For the purpose of this paper, we will operationalize this into the more specific and functional definition coined by Stanford University's Formal Reasoning Group, which usefully covers both 'natural' (human and animal) and 'artificial' forms of intelligence:

> *"Intelligence is the computational part of the ability to achieve goals in the world."*[29]

From this definition of intelligence, which refers to both internal processes ("computation…") that act in the service of bringing about *external* results ("…the ability to achieve goals…") across complex, dynamic environments ("…in the world"), we can proceed to home in on a definition of *artificial* intelligence.

2.1.2. Artificial Intelligence

In its broadest sense, AI has been described as "the study of the computations that make it possible to perceive, reason, and act"[30] or "the automation of intelligent behavior",[31]

24 Ibid.
25 For an overview of various definitions, and an attempt to formalize a concept of (machine) intelligence in the broadest reasonable sense, see Shane Legg and Marcus Hutter, "Universal Intelligence: A Definition of Machine Intelligence," *Minds and Machines* 17, no. 4 (2007): 391–444.
26 Cf. Yudkowsky, "Artificial Intelligence as a Positive and Negative Factor in Global Risk.", p. 311
27 Shane Legg and Marcus Hutter, "A Collection of Definitions of Intelligence," *arXiv:0706.3639 [Cs]*, June 25, 2007, http://arxiv.org/abs/0706.3639.
28 Ibid.
29 John McCarthy and Stanford University Formal Reasoning Group, "What Is Artificial Intelligence | Basic Questions," *Formal Reasoning Group*, 2007, http://www-formal.stanford.edu/jmc/whatisai/node1.html.
30 Patrick Henry Winston, *Artificial Intelligence*, 3rd ed., 1992.
31 George F. Luger and William A. Stubblefield, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, 6th ed., 2008. ; cf. also A. Barr and Feigenbaum, eds., *The Handbook of Artificial Intelligence*, vol. 2 (Stanford,

which is driven by a general "study of intelligent agents" both biological and artificial.[32] There are furthermore dozens of definitions and typologies of what constitutes artificial intelligence.[33] However, in concrete terms, and in most applications, AI is defined as non-human intelligence that is measured by its ability to replicate human mental skills, such as pattern recognition, understanding natural language (NLP), adaptive learning from experience, strategizing, or reasoning about others.[34]

Likewise, militaries have also considered AI in a functional context – by the degree to which a machine, in the words of one (older) study of the US Army Sciences Board, "... can incorporate abstraction and interpretation into information processing and make decisions at a level of sophistication that would be considered intelligent in humans",[35] a definition that is also upheld in the more recent Summer Study on Autonomy by the US Defense Science Board, which describes AI as "the capability of computer systems to perform tasks that normally require human intelligence (e.g., perception, conversation, decision-making)."[36] These characterizations, and the explicit link to human performance, derive from the very inception of the field of AI. The original proposal for the seminal 1956 Dartmouth College Summer Project argued 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. An attempt will be made to find how to make* ***machines that use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.****"*[37]

However, while human intelligence has been the common choice as a yardstick for benchmarking or assessing progress in AI development, there are also approaches that do not seek to recreate human intelligence or performance, but instead focus more on systems that approach an ideal-typical 'rational' performance. As a result, most concrete definitions of AI fall into one of four categories (cf. **Figure 3**); these categories represent approaches which take distinct positions on two conceptual dimensions:

1. Whether it emphasises the attainment of specific ('intelligent' or 'sentient') **thought processes (T)** and reasoning, or whether it emphasises ('goal-oriented'; 'effective') **behavior (B)**;
2. Whether it measures success on (1) against **human performance (H)**, or against an ideal concept of intelligence – usually defined as **'rationality' (R)**;[38]

California & Los Altos, California: HeurisTech Press and William Kaufmann, 1982).
32 Cf. Poole, et al., 1998, p.1; Russell, & Norvig, 2009
33 For a comprehensive overview, see also Daniel Faggella, "What Is Artificial Intelligence? An Informed Definition -," *TechEmergence.com*, October 10, 2016, http://techemergence.com/what-is-artificial-intelligence/
34 Cf. Russell, & Norvig, 2009, one of the standard textbooks in the field.
35 Barry J. Brownstein and et al., "Technological Assessment of Future Battlefield Robotic Applications," Proceedings of the Army Conference on Application of AI to Battlefield Information Management (White Oak: US Navy Surface Weapons Center, 1983)., p. 169
36 Defense Science Board, "Report of the Defense Science Board Summer Study on Autonomy" (Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, June 2016), https://www.hsdl.org/?view&did=794641, p. 5
37 J. McCarthy et al., "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," August 31, 1955, http://robotics.cs.tamu.edu/dshell/cs625/2014-09-02.pdf.. Emphasis added.
38 Russell and Norvig, *Artificial Intelligence: A Modern Approach*

	Human Benchmark (H)	Rationality benchmark (R)
Intelligence as Thought Processes (T)	**(T-H) Systems that think like humans** (e.g. cognitive science) *"The exciting new effort to make computers think ... machines with minds, in the full and literal sense"* Haugeland, 1985 *"The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..."* Bellman, 1978	**(T-R) Systems that think rationally** (logic/ laws of thought) *"The study of mental faculties through the use of computational models"* Charniak and McDermott, 1985 *"The study of the computations that make it possible to perceive, reason, and act"* Winston, 1992
Intelligence as goal-oriented behavior (B)	**(B-H) Systems that act like humans** (Cf. Turing test; Winograd Schema Challenge[39]) *"The art of creating machines that perform functions that require intelligence when performed by people"* Kurzweil, 1990 *"The study of how to make computers do things at which, at the moment, people are better"* Rich and Knight, 1991	**(B-R) Systems that act rationally** (rational agents) *"A field of study that seeks to explain and emulate intelligent behavior interms of computational processes"* Schalkoff, 1990 *"The branch of computer science that is concerned with the automation of intelligent behavior"* Luger & Stubblefield, 1993

FIGURE 3: A TYPOLOGY OF DIFFERENT DEFINITIONS OF AI, AND THEIR UNDERLYING APPROACHES, WITH SOME EXAMPLES[39]

Ultimately, while approaches that emphasize internal (e.g. 'T') models are of intellectual and philosophical interest, it is important that the definition of AI which we use does not put arbitrary limits on the inner workings of AI, or the approaches used to creating it.[40] Moreover, from a practical (that is, political or military) perspective, the focus tends to be on the way artificially intelligent systems manage to *act* in the world, that is behavior-focused (B) approaches. Secondly, both in current 'deep learning' applications, as well as in most prospective military applications, AI systems will provide the most added value (the competitive edge), not only by equating human intelligence, but precisely by surpassing it – even if only within a narrow domain, such as information analysis or reaction time.

39 Adapted from Russell and Norvig, *Artificial Intelligence: A Modern Approach.*, p. 5. The sources mentioned are R.E. Bellman, *An Introduction to Artificial Intelligence: Can Computers Think?* (San Francisco: Boyd & Fraser Publishing Company, 1978).; E. Charniak and D. McDermott, *Introduction to Artificial Intelligence* (Massachusetts: Addison-Wesley Reading, 1985).; J. Haugeland, ed., *Artificial Intelligence: The Very Idea* (Cambridge, Mass: MIT Press, 1985).; R. Kurzweil, *The Age of Intelligent Machines* (Cambridge, Mass: MIT Press, 1990).; George F. Luger and William A. Stubblefield, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, 2nd ed. (California: Benjamin/Cummings, 1993).; E. Rich and K. Knight, *Artificial Intelligence*, 2nd ed. (New York: McGraw-Hill, 1991).; R.I. Schalkoff, *Artificial Intelligence: An Engineering Approach* (New York: McGraw-Hill, 1990). and Winston, *Artificial Intelligence*..

40 Cf. Shane Legg and Marcus Hutter, "A Formal Definition of Intelligence for Artificial Systems," *Proc. 50th Anniversary Summit of Artificial Intelligence*, 2006, 197–198.; cf. Faggella, "What Is Artificial Intelligence?"

Accordingly, one strategically (and therefore analytically) relevant definition for our purposes is that of AI as a rational optimization agent that can (enable humans to) act competently in the world (that is, subtype B-R). Therefore, while it is important to use human performance as one measuring stick to understand the threshold at which AI systems (or so-called 'centaur' human-AI teams) can start to outperform and outcompete human adversaries, for most strategic purposes we can focus on the thought processes (T), and specifically the behavioral performance (B) of AI systems, at an equal- or better-than human level of accuracy, speed, or decision quality. This corresponds to the definition provided by Nils J. 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."[41]

2.1.3. Artificial Narrow-, General-, and Superintelligence

There is one additional distinction that is of particular use when thinking about the sequence and speed with which progress on AI will unfold. The literature generally agrees that there are three tiers of AI[42] – which can also be seen as three consecutive generations of AI (see also chapter 3.6 page 55):

- **Artificial Narrow Intelligence** (ANI or "narrow AI")[43]: machine intelligence that *equals or exceeds* human intelligence *for specific tasks*. Existing examples of such systems would be IBM's Deep Blue (Chess) & Watson ('Jeopardy!'), Google's AlphaGo (go), High-Frequency Trading Algorithms, or indeed any specialized automatic systems performing beyond human reach (e.g. Google Translate; spam filters; the guidance systems of point-defense anti-missile cannons etc.).
- **Artificial General Intelligence** (AGI or "strong AI"): machine intelligence meeting the *full range* of human performance across any task; and:
- **Artificial Superintelligence** (ASI): machine intelligence that *exceeds human intelligence* across any task.[44]

41 Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge ; New York: Cambridge University Press, 2010).

42 Another frequently used classification differentiates between weak and strong AI. There seem to be quite a few different ways to characterize the difference (for a divergent one, see Stephen Lucci and Danny Kopec, *Artificial Intelligence in the 21st Century: A Living Introduction*, 2016. but the most frequently used one differentiates between "weak AI"—the variety devoted to providing aids to human thought—and "strong AI"—the variety that attempts to mechanize human-level intelligence (a machine with consciousness, sentience and mind). See Nilsson, *The Quest for Artificial Intelligence*

43 Ray Kurzweil, "Long Live AI," *Forbes.com*, August 15, 2015, http://www.forbes.com/home/free_forbes/2005/0815/030.html

44 Note that even such hypothetical superintelligences might come in different forms, such as 'speed superintelligence' - 'a system that can do all that a human intellect can do, but much faster"; 'collective superintelligence' - 'A system composed of a large number of smaller intellects such that the system's overall performance across many very general domains vastly outstrips that of any current cognitive system'; or 'quality superintelligence' - "A system that is at least as fast as a human mind and vastly qualitatively smarter." Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (OUP Oxford, 2014)., pg. 53, 54, 56

2.2. A Brief History of AI

Contrary to widespread perception AI is not an entirely new discipline. Many of its founding concepts draw on over 2000 years of insights accumulated in philosophy, logic, mathematics, theories of reasoning, cognitive psychology and linguistics.[45] As a practical, applied field, however, AI truly came into its own in the direct wake of the Second World War. The inception of AI as a discipline occurred in wartime research in areas such as cryptography and the calculation of ballistic firing tables for artillery,[46] and it was spurred on by the foundational work of Alan Turing,[47] as well as the work on simple neural networks by Warren McCulloch and Walter Pitts.[48] AI first became a distinct field of study during the small Dartmouth Summer Project held in 1956 – an event that was to herald the first 'spring' for AI research.

2.2.1. Early Enthusiasm: The First AI Spring ('56-75')

Contrary to the ambitious aspirations of the Dartmouth project – its ten attendees had anticipated making "a significant advance" within the space of a single summer[49] – the workshop did not lead to any direct breakthroughs. However, it did give rise to the name 'artificial intelligence', and by introducing all major thinkers to each other, it established the shape of that field for decades to come.[50] The intellectual cross-pollination at Dartmouth, and a subsequent string of high-profile achievements by AI researchers in developing tools capable of human-like performance in narrow fields such as geometrical proofs, algebra, and simple games (e.g. checkers[51]), led to an early golden age of AI research (1956-1974). This was a period in which prototype systems – operating safely within well-defined limits of an artificially simple 'microworld' – again and again refuted established beliefs that *'no machine could ever do X!'*.[52]

In addition to systems based on formal symbolic reasoning, there was a line of 'connectionist' research into what became known as *perceptrons*, an early form of neural networks.[53] During this period, considerable amounts of scientific – and, with the increasing pressures of the Cold War, also military – funds were directed into the field.[54] From the early 1960s to the 1970s, and following a philosophy that it should 'fund people,

45 Russell and Norvig, *Artificial Intelligence: A Modern Approach*., p. 8-15
46 In the shape of ENIAC, the first electronic general-purpose computer, which was developed from 1943-1945 (though not delivered until the end of the war) by John Mauchly and J. Presper Eckert. Cf. Harry L. Reed Jr., "Firing Table Computations on the Eniac," in *Proceedings of the 1952 ACM National Meeting (Pittsburgh)*, ACM '52 (New York, NY, USA: ACM, 1952), 103–106, doi:10.1145/609784.609796
47 e.g. A.M. Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (October 1950): 433–60
48 Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5 (1943): 115–33
49 McCarthy et al., "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence."
50 Russell and Norvig, *Artificial Intelligence: A Modern Approach*., pg. 17
51 Cf. Arthur L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal* 3, no. 3 (July 1959), https://www.cs.virginia.edu/~evans/greatworks/samuel1959.pdf.
52 Cf. Bostrom, *Superintelligence*., pg. 5
53 Russell and Norvig, *Artificial Intelligence: A Modern Approach*., pg. 19-20
54 Daniel Crevier, *AI: The Tumultuous Search for Artificial Intelligence* (New York: BasicBooks, 1993).; Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence*, 4. print (Cambridge: Harvard Univ. Press, 1995),pg. 9

not projects',[55] the Advanced Research Projects Agency (ARPA) was to provide major funding, with few strings attached, to open-ended AI research projects of renowned researchers at MIT, Carnegie Mellon University, and Stanford, amongst others.

2.2.2. 'Grandiose Objectives': The First AI Winter ('74-'80)

This first wave of enthusiasm proved unsustainable, however, and by the early 1970s the field of AI was slipping into the first AI Winter. It became clear that rapid early progress had slowed down, and that AI systems remained far more limited in their capabilities than had initially been expected. There were several reasons for this disillusion, chief amongst them the discovery of the combinatorial explosion of possibilities', which demonstrated that exhaustive search, a key trial-and-error approach underpinning many AI algorithms at the time, would require exponential amounts of computer processing power once one moved away from 'toy problems' to more complex real-world problems – an amount of processing power which was unimaginable at the time. This challenge was aggravated by the technical limitations at the time – chiefly hardware limits on memory and processor speed.[56]

There were also significant internal disagreements; in a seminal 1969 book,[57] Marvin Minsky and Seymour Papert demonstrated that there were limits to the performance of the early perceptrons of the connectionist school. As a result of this work, research funding on these early neural nets was mostly abandoned for a decade, and work focused more on programming formal rules into symbolic systems, which would feed into the later expert system' of the 1980s.

All these diverse problems led to significant funding cutbacks on both sides of the Atlantic; in 1966, a scathing report by the Automatic Language Processing Advisory Committee of the US government on the slow progress in natural language processing and automatic translation led to the termination of all research funding by the US National Research Council.[58] In 1973, the publication of the Lighthill Report[59], which was highly critical of the failure of AI research to live up to its 'grandiose objectives', led to a complete cessation of British research for over a decade, sending a bow wave of funding cuts in AI research across Europe. At the same time, ARPA was deeply disappointed in the returns on its initial investment in AI, specifically in speech recognition, but also more broadly. After the passage of the 1969 Mansfield Amendment, which required US military research funding to focus more on "mission-oriented direct research, rather than basic undirected research"[60], it did not take long for nearly all military funding to dry up.

---

55 Daniel Crevier, *AI: The Tumultuous Search for Artificial Intelligence* (New York: BasicBooks, 1993). p. 65
56 Bostrom, *Superintelligence*., pg. 6-7
57 Marvin Minsky and Seymour Papert, *Perceptrons* (Oxford, England: M.I.T. Press, 1969).
58 John Hutchins, "The History of Machine Translation in a Nutshell," 2005, http://www.hutchinsweb.me.uk/Nutshell-2005.pdf.
59 James Lighthill, "Artificial Intelligence: A General Survey," Artificial Intelligence: A Paper Symposium (Science Research Council, 1973), http://www.math.snu.ac.kr/~hichoi/infomath/Articles/Lighthill%20Report.pdf.
60 The Mansfield Amendment was passed as part of the Defense Authorization Act of 1970. See US Congress, "Defense Authorization Act," Pub. L. No. 91–121 (1969).. See also Committee on Innovations in Computing and Communications, National Research Council, "Chapter 9: Developments in Artificial Intelligence | Funding a

2.2.3. Expert Systems: The Second AI Spring ('80-'87)

A second wave came in the 1980s, with the advent of so-called expert systems – rule-based programmes that answered questions or solved problems within a small domain of specific knowledge.

Such programmes emulated decision-making processes of a human expert, which had been hard-coded into a formal logical language. Expert systems could be used as support tools for decision-makers or executives. At the same time, after almost a decade of relative standstill, research on neural networks, the aforementioned connectionist school, saw a renaissance spurred on by progress in physics and computer sciences, as well as the development of the backpropagation algorithm,[61] which made it possible to train multi-layered neural networks capable of learning a wider range of useful functions.

Meanwhile, the promise of ANI-like expert systems – applications of AI which were recognized as economically useful by the broader industry[62] – led to interest and funding for research, starting with Japan's launch (in 1981) of the 10-year *Fifth-Generation Computer Systems Project*, a public-private partnership that sought to develop a massively parallel computing architecture capable of rapidly drawing on vast stores of rules, producing AI capable of natural conversation, translation, and image interpretation.[63] Spurred on by fears of falling behind, many Western governments followed suit and provided funding to AI research – such as the Microelectronics and Computer Technology Corporation (MCC) in the US, and the Alvey Project in the UK. Simultaneously, DARPA founded its Strategic Computing Initiative in the process tripling its funding for research.[64] As a result, various software and hardware companies converged around expert systems and sales went from a few millions in 1980 to $2 billion by 1988.[65]

2.2.4. 'Not the Next Wave'; The Second AI Winter ('87-'93)

Once again, however, the boom turned into a bust, as research programs in Japan, the US and Europe failed to meet their objectives. Many specialized AI hardware companies collapsed after 1987, as the desktop computers from Apple and IBM rapidly outcompeted more expensive AI-specific ones. With only some exceptions, expert systems proved of limited practical utility. Small expert systems generally were ineffective or added little

---

Revolution: Government Support for Computing Research," in *Funding a Revolution: Government Support for Computing Research*, 2008, https://web.archive.org/web/20080112001018/ http://www.nap.edu/readingroom/books/far/ch9.html

61 The first algorithm underlying backpropagation had already been discovered in 1969, by Arthur Bryson and Yu-Chi Ho, Arthur Earl Bryson and Yu-Chi Ho, *Applied Optimal Control: Optimization, Estimation, and Control.* (Blaisdell Publishing Company or Xerox College Publishing, 1969).. Its application to neural nets was developed by Paul Werbos, Paul Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences" (Ph.D., Harvard University, 1974).. However the algorithm was only popularized in the community in the mid-1980s by David Rummelhart, in David E. Rumelhart, James L. McClelland, and CORPORATE PDP Research Group, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations* (Cambridge, MA, USA: MIT Press, 1986)

62 Russell and Norvig, *Artificial Intelligence: A Modern Approach*., p23-24

63 Ibid., p24; Bostrom, *Superintelligence*., p7

64 Pamela McCorduck, *Machines Who Think*, 2nd ed. (Natick, MA: A.K. Peters, Ltd., 2004).

65 Russell and Norvig, *Artificial Intelligence: A Modern Approach*.

value, large ones could be prohibitively expensive to create, test and keep updated and they were also fairly 'brittle', as they tended to crash in the face of unusual inputs.[66] At the end of the 1980s, the leadership at DARPA's Information Processing Technology Office took the view that AI was not "the next wave" and that researchers had once again over-promized and under-delivered.

This led to severe funding cuts across the Strategic Computing Initiative, leaving only programs of direct military relevance – including an autonomous battle tank program, never delivered, and the DART logistics & battle management system, which would later prove invaluable during the First Gulf War.[67]

2.2.5. A Third, Sustained AI Spring ('93-'11)

After the relative disappointment of the previous decades, many AI researchers abandoned long-term dreams of developing 'human-level' AI, and instead moved into fragmented sub-fields focused on rigorously solving specific problems or applications.[68] Moreover, the traditional logic-based paradigm, which had reached its summit with the imposing but brittle expert systems of the 1980s, found itself progressively challenged by techniques both old and new. Perhaps the most important of these came in the form of a reinvigorated interest in neural networks and genetic algorithms – approaches which distinguished themselves from the earlier expert systems by exhibiting 'graceful degradation' – referring to small errors in the assumptions resulting in only small reductions in performance, rather than a full crash.

As a result of this pragmatic turn, further fuelled by Moore's Law and the attendant advances in hardware capabilities,[69] the field of AI quietly began to bloom.[70] It began not only to achieve some of its decades-old goals, but was also able to increasingly, if invisibly, permeate both the technology industry, and, frequently, aspects of daily life.[71] In the form of direct applications or downstream spinoff technologies, research was finding utility in a wide range of fields, from games (notably the famous 1997 chess victory, of IBM's Deep Blue over world champion Garry Kasparov); to logistics, spacecraft and satellite monitoring; robotics; traffic management; medical diagnostics;[72]speech recognition,[73] autonomous vehicles and Google's search engines, to name but a few.[74]

---

66 Douglas Lenat and R.V. Guha, *Building Large Knowledge-Based Systems* (Addison-Wesley, 1989).. See also the account in Stuart Russell, "Defining Intelligence," *EDGE*, February 7, 2017, https://www.edge.org/conversation/stuart_russell-defining-intelligence

67 McCorduck, *Machines Who Think*., pg. 430-43

68 Ibid

69 Cf. Ray Kurzweil, *The Singularity Is Near* (Viking Press, 2005)., p274.

70 Cf. Jiqiang Niu et al., "Global Research on Artificial Intelligence from 1990–2014: Spatially-Explicit Bibliometric Analysis," *ISPRS International Journal of Geo-Information* 5, no. 5 (May 16, 2016): 66, doi:10.3390/ijgi5050066

71 Russell and Norvig, *Artificial Intelligence: A Modern Approach*.

72 Ibid., pp. 26-27;

73 Cf. The Economist, "Are You Talking to Me? -," *The Economist*, June 7, 2007, http://www.economist.com/node/9249338.

74 Ajit Nazre and Rahul Garg, "A Deep Dive in the Venture Landscape of Artificial Intelligence and Machine Learning," September 2015, http://slideplayer.com/slide/7002258/.

Military applications incorporating ANI also grew in prominence such as for instance unmanned aerial vehicles (UAVs). Military UAVs had previously seen limited use, by the US in the Vietnam War, and by the Israeli Defense Force during the Yom Kippur War ('73) and (with great success) against the Syrian air force during the First Lebanon War ('82).[75] However, from the '90s, and with the first deployment of the Predator UAV during the Balkan Wars, these systems started to become a widely used component of military operations – and UAVs have made appearances in almost every conflict since.

Finally, the performance of the previously mentioned DART tool for automated logistics planning and scheduling, development of which had survived even the drawdown in DARPA's Strategic Computing Initiative, during Operation Desert Storm in 1991, led DARPA to claim that this single application had more than paid back their thirty-year investment in AI.[76]

In addition to private funding, public funding picked up as well: since the mid-2000s, both the European Union (under its Seventh Framework Programme) and the US (through DARPA's Grand Challenge Program and the Cognitive Technology Threat Warning System) have begun to provide renewed broad funding for AI research. As such, even though it took until the mid-2000s for AI researchers to shrug off any negative legacies associated with the term AI – many instead choosing to describe their work as 'knowledge-based systems'; 'informatics', or 'computational intelligence', to name a few[77] – Research gradually but surely has entered its third, and so far most sustained 'spring' a steady blooming.

2.2.6. Big Data, Deep Learning, and an Artificial Intelligence Revolution ('11-present)

In recent years, the '3rd reincarnation'[78] of AI appears to have hit a tipping point. This success is driven by specific factors which have produced a dramatic improvement in the predictive accuracy of algorithms. Some of these factors are conceptual, including advances in neuroscience as well as computer science – most notably the work of Geoffrey Hinton and Salakhutdinov[79], which pioneered powerful new techniques for enabling neural network pattern-recognition. Other factors of note are increases in affordable computing power and faster networks; cloud infrastructures; the growth in the Internet of Things and Big Data – and specifically the open-source availability of very large datasets (sometimes

75 Mary Dobbing and Chris Cole, "Israel and the Drone Wars: Examining Israel's Production, Use and Proliferation of UAVs" (Drone Wars UK, January 2014), https://dronewarsuk.files.wordpress.com/2014/01/israel-and-the-drone-wars.pdf.

76 Sara Reese Hedberg, "DART: Revolutionizing Logistics Planning," *IEEE Intelligent Systems* 17, no. 3 (May 2002): 81–83, doi:10.1109/MIS.2002.1005635.; Stephen E. Cross and Edward Walker, "DART: Applying Knowledge Based Planning and Scheduling to CRISIS Action Planning.," in *Intelligent Scheduling*, ed. Monte Zweben and Mark Fox (San Francisco, CA: Morgan Kaufmann, 1994), 711–29.

77 John Markoff, "Behind Artificial Intelligence, a Squadron of Bright Real People," *The New York Times*, October 14, 2005, http://www.nytimes.com/2005/10/14/technology/behind-artificial-intelligence-a-squadron-of-bright-real-people.html.

78 Nazre and Garg, "A Deep Dive in the Venture Landscape of Artificial Intelligence and Machine Learning.", slide 2

79 G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science* 313, no. 5786 (July 28, 2006): 504–7, doi:10.1126/science.1127647.

generated from social-media networks) for use in the training and testing of large-scale machine learning networks.[80]

Furthermore, more than ever before, AI research has benefited tremendously from a huge boom in the level, diversity and sources of funding and talent, including from major private sector players such as Apple, Amazon, Baidu, Google, Facebook, IBM, Microsoft.[81] For many of these, AI systems increasingly are not just one more asset by which to marginally increase existing profits: instead, they are at the heart of the enterprise's business model. AI research has also benefited from investments by traditional top-spenders on research, amongst others from the automotive industry – with $1 billion investment programs by Toyota[82] and (in February 2017) Ford Motors,[83] alongside major programs by Mercedes-Benz,[84] BMW,[85] – as well as the pharmaceutical sector.[86]

Finally, all this has also been facilitated by a new mode of cumulative knowledge-building, which marks a shift away from closed epistemic communities with limited interconnections that are tightly policed by gatekeepers, towards an open-source model of research, debate, scholarship and content creation (as in the case of Google Translate, an application which, calibrated with reference to existing document translations, and continually improved by massed user feedback, has rapidly bypassed the accuracy of any of the traditional translation programs, which relied on painstaking manual coding by linguistic experts).

Whatever other effects (good and bad) this shift towards collaborative ecosystem-like crowdsourcing has had, it has certainly lowered barriers to access, created greater (reputational and financial) rewards for success and innovation, and led to the proliferation of ever more comprehensive and widely available data, which in turn has allowed for the rapid implementation, deployment, testing and iteration over new AI applications.[87]

80 Nazre and Garg, "A Deep Dive in the Venture Landscape of Artificial Intelligence and Machine Learning."
81 Ibid., slide 7
82 Rachel Metz, "Toyota Investing $50M for Autonomous-Car Research at Stanford and MIT," *MIT Technology Review*, 2015, https://www.technologyreview.com/s/541046/toyota-investing-50m-with-stanford-mit-for-autonomous-car-research/.; Hans Greimel, "What's behind Toyota's Big Bet on Artificial Intelligence?," *Automotive News*, 2015, http://www.autonews.com/article/20151109/OEM06/311099937/whats-behind-toyotas-big-bet-on-artificial-intelligence%3F.
83 Mike Isaac and Neal E. Boudette, "Ford to Invest $1 Billion in Artificial Intelligence Start-Up," *The New York Times*, February 10, 2017, https://www.nytimes.com/2017/02/10/technology/ford-invests-billion-artificial-intelligence.html
84 Mark Bergen, "Mercedes-Benz Wants to Beat Google, Uber to Our Driverless Future," *Recode*, November 26, 2015, http://www.recode.net/2015/11/26/11620962/mercedes-benz-wants-to-beat-google-uber-to-our-driverless-future
85 Horatiu Boeriu, "BMW Sees New Opportunities in Artificial Intelligence," *BMW BLOG*, June 9, 2016, http://www.bmwblog.com/2016/06/09/bmw-sees-new-opportunities-artificial-intelligence/.
86 Cf. CB Insights, "From Virtual Nurses To Drug Discovery: 106 Artificial Intelligence Startups In Healthcare," *CB Insights - Blog*, February 3, 2017, https://www.cbinsights.com/blog/artificial-intelligence-startups-healthcare/., for an description of startups applying AI to medical applications, see João Medeiros, "The Startup Fighting Cancer with AI," *WIRED UK*, accessed February 25, 2017, http://www.wired.co.uk/article/ai-cancer-drugs-berg-pharma-startup.. For one overall overview of AI investments by sector, see Bank of America Merrill Lynch, "Robot Revolution - Global Robot & AI Primer," Primer, Thematic Investing (Bank of America Merrill Lynch, December 2015), http://www.bofaml.com/content/dam/boamlimages/documents/PDFs/robotics_and_ai_condensed_primer.pdf.
87 Stanford University, "Artificial Intelligence and Life in 2030 One Hundred Year Study on Artificial Intelligence | Report of the 2015 Study Pane.", p. 14-17

This potent combination of powerful hardware, extensive funding, (relatively) open development, and big (labelled) test datasets, has resulted in remarkable achievements in numerous fields. This includes the long-intractable field of natural language processing: starting with IBM Watson's famous 2011 Jeopardy! victory, this track record has continued with the public debut of voice-responsive Virtual Personal Assistants (like Apple's SIRI, or Microsoft's Cortana), culminating in the 2015 DeepMind project which trained a deep neural net on over 300.000 CNN and Daily Mail articles, after which it was able to use information from these articles to accurately answer 60% of the queries put to it.[88] Other innovations marked the first challenges to the hallowed Turing test, including the 2013 announcement, by startup Vicarious announcing in 2013 that its AI had achieved a 90% success rate at solving CAPTCHA tests;[89] and the development of 'Eugene Goostman', a chatbot which in 2014 managed to fool 33% of human judges at the Royal Society that it was a 13-year old boy.[90]

Other advances have come in the field of face recognition, with various algorithms developed by Google (GoogLeNet), Facebook (DeepFace – 2013) and Yahoo (DeepDense – 2015)[91] having achieved better-than-human identification rates. Finally, Google DeepMind has achieved remarkable successes in developing self-teaching AI agents capable of superhuman performance at common video games (ATARI Deep-Q – 2013)[92] and the game of go (AlphaGo – 2016).[93]

There has also been growing interest in the (cyber)security applications of these technologies; in the Summer of 2016, the 'Mayhem' AI (Carnegie Mellon University) won the DARPA Cyber Grand Challenge in automatic system vulnerability detection & patching.[94]

---

88 Emerging Technology from the MIT Tech Review, "Google DeepMind Teaches Artificial Intelligence Machines to Read," *MIT Technology Review*, 2015, https://www.technologyreview.com/s/538616/google-deepmind-teaches-artificial-intelligence-machines-to-read/.. See also the original paper at Karl Moritz Hermann et al., "Teaching Machines to Read and Comprehend," *arXiv:1506.03340 [Cs]*, June 10, 2015, http://arxiv.org/abs/1506.03340.

89 Rachel Metz, "AI Startup Vicarious Claims Victory Over Captchas," *MIT Technology Review*, 2013, https://www.technologyreview.com/s/520581/ai-startup-says-it-has-defeated-captchas/.

90 The Guardian, "Computer Simulating 13-Year-Old Boy Becomes First to Pass Turing Test," *The Guardian*, June 9, 2014, sec. Technology, https://www.theguardian.com/technology/2014/jun/08/super-computer-simulates-13-year-old-boy-passes-turing-test.; others however questioned the significance of the contest, cf .Gary Marcus, "What Comes After the Turing Test?," *The New Yorker*, June 9, 2014, http://www.newyorker.com/tech/elements/what-comes-after-the-turing-test

91 Cf. Emerging Technology from the MIT Tech Review, "The Face Detection Algorithm Set to Revolutionize Image Search," *MIT Technology Review*, 2015, https://www.technologyreview.com/s/535201/the-face-detection-algorithm-set-to-revolutionize-image-search/.; see also the original paper at Sachin Sudhakar Farfade, Mohammad Saberian, and Li-Jia Li, "Multi-View Face Detection Using Deep Convolutional Neural Networks," *arXiv:1502.02766 [Cs]*, February 9, 2015, http://arxiv.org/abs/1502.02766..

92 Volodymyr Mnih et al., "Playing Atari with Deep Reinforcement Learning," 2013.; Cade Metz, "Teaching Computers to Play Atari Is A Big Step Toward Bringing Robots Into the Real World," *WIRED*, 2015, https://www.wired.com/2015/12/teaching-ai-to-play-atari-will-help-robots-make-sense-of-our-world/

93 In February 2017, another AI company, Libratus, also announced a breakthrough in AI systems playing poker against the best human players. cf. Cade Metz, "Inside the Poker AI That Out-Bluffed the Best Humans," *WIRED*, 2017, https://www.wired.com/2017/02/libratus/.

94 Nonetheless, Mayhem was still soundly defeated by all-human hacking teams at the DEFCON summit soon after. See Devin Coldewey, "Carnegie Mellon's Mayhem AI Takes Home $2 Million from DARPA's Cyber Grand Challenge," *TechCrunch*, 2016, http://social.techcrunch.com/2016/08/05/carnegie-mellons-mayhem-ai-takes-home-2-million-from-darpas-cyber-grand-challenge/

Finally, there has also been a growing awareness of the societal and operational concerns revolving around the accountability of ANI systems which are based on opaque machine learning capabilities. For instance, in the fall of 2016, DARPA issued a funding call for projects focused on 'Explainable Artificial Intelligence' (XAI) – "new or modified machine learning techniques that produce explainable models that, when combined with effective explanation techniques, enable end users to understand, appropriately trust, and effectively manage the emerging generation of Artificial Intelligence (AI) systems."[95]

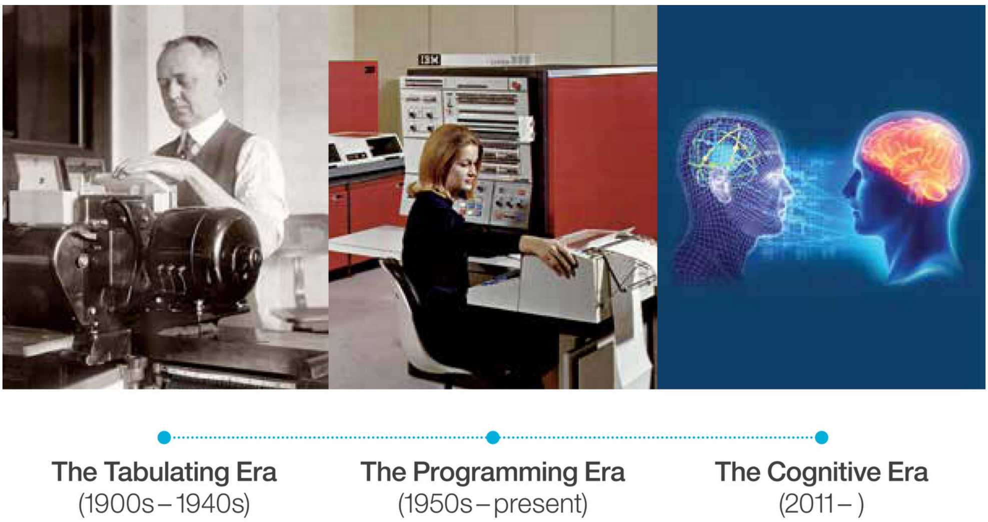

FIGURE 4: IBM'S VIEW OF THE HISTORY OF COMPUTING[96]

Recent developments strongly suggest that AI research is now moving into the third of three main 'eras' in the evolution of AI, identified by IBM's John Kelly.[97] In the first era (the 'Tabulating Era' – 1900-1940s), humans created single-purpose mechanical systems that essentially were able to perform a single function: to count. They used punched cards to input and store data, and to eventually instruct the machine what to do – in a very primitive way. In the second era (the 'Programming Era' – 1950s-present), the advent of electronic systems allowed for the evolution of digital "computers" that were able to perform ever more sophisticated if/then logical operations and loops, with instructions coded in software that was programmed by humans. With the rise of powerful deep learning systems, and their application throughout research and industry, we are at last entering the Cognitive Era, in which these software systems no longer have to be fully programmed, but can instead start 'learning' autonomously through training, use,

95 DARPA, "Grant Opportunity - DARPA-BAA-16-53 - Explainable Artificial Intelligence (Xai)" (Department of Defense - DARPA - Information Innovation Office, n.d.), http://www.grants.gov/web/grants/view-opportunity.html?oppId=287284.. See also David Gunning, "Explainable Artificial Intelligence (XAI)," *DARPA*, accessed September 23, 2016, http://www.darpa.mil/program/explainable-artificial-intelligence.
96 J. E. Kelly, "Computing, Cognition and the Future of Knowing," *Whitepaper, IBM Reseach*, 2015, http://www.academia.edu/download/44915350/Computing_Cognition_WhitePaper.pdf
97 Ibid

and user feedback, in ways which may resemble human learning, but which often can no longer be fully comprehended by it.[98] To understand the opportunities – and limits – offered by such learning machines, we will now briefly review the AI discipline, and its distinct approaches at developing machines that can accurately represent the world, learn, and take appropriate, and increasingly intelligent decisions.

2.3. AI: A Cookbook of Components, Approaches and Design Architectures

For those unfamiliar with the field, or too familiar with popular culture, it may sometimes appear as if the development of Artificial Intelligence, insofar as it seeks to reproduce human intelligence (singular), is about a single technology – a unified search to develop that kernel of sentience in machines, which humans already possess. Yet nothing could be further from the truth: in practice, far from involving just one single technology or unified discipline, AI is a collection of distinct fields working on different technologies that, working together in the appropriate environments, allow for intelligent behavior to emerge.

2.3.1. Design Specifications: The Components of Intelligence

In this context, it helps to recall that the human variety of intelligence is similarly complex and multifaceted. When we think of human intelligence, we tend think of a singular abstract cognitive ability that we see as a trait of an individual, making her 'smarter' (e.g. more knowledgeable; more capable) than others. When we do so, however, we tend to forget how many different – (neuro)biological – components have to coalesce in order for any form of intelligence to materialize. Let us return to the example of intelligence we started with: reading. As you read these sentences, the center of the retinas in your eyes has to use motor movement to scan the page carefully. This scan is then transferred as chemical signals to the visual cortex in the brain which recognizes patterns in a row of black marks against a white background. Your visual system then progressively extracts graphemes, syllables, prefixes, suffixes, and word roots as it parses our natural language. Through various routes – which again involve various parts of our brain – it converts those few black marks into an entire semantic universe of meanings by cross-referencing them against various other pieces of information stored in your brain[99]. And all of this is just the 'legere' part of 'intelligence' – the overall process whereby reading (or hearing, or talking, or experiencing, etc.) translates into intelligence that is even more complex[100].

98 For a thought-provoking analysis of the 'inscrutability' of deep-learning based AI networks, see Aaron M. Bornstein, "Is Artificial Intelligence Permanently Inscrutable?," *Nautilus*, 2016, http://nautil.us/issue/40/learning/is-artificial-intelligence-permanently-inscrutable.

99 For a fascinating overview of this process, see Stanislas Dehaene, *Reading in the Brain: The Science and Evolution of a Human Invention* (New York: Viking, 2009).. For a more recent overview of the different parts of the brain that are known to be involved in this process, see Stephani Sutherland, "What Happens in the Brain When We Read?," *Scientific American Mind* 26, no. 4 (August 7, 2015): 14–14

100 For a very readable overview of what cognitive neuroscientists have been discovering worldwide about the

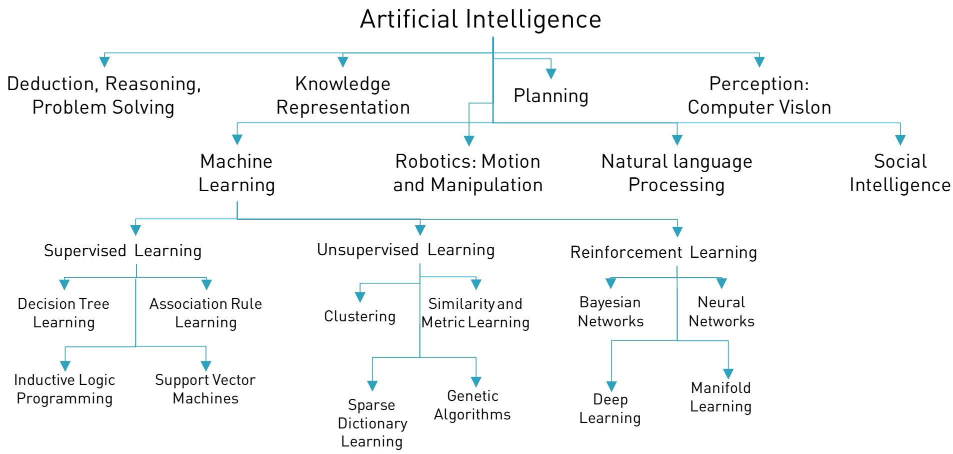

FIGURE 5: AN OVERVIEW OF NOTABLE APPROACHES AND DISCIPLINES IN AI AND MACHINE LEARNING[101]

By analogy to these components constituting human cognition, what, then, is the library of intelligent sensory-, information-processing-, pattern-matching- and categorization-functions underlying artificial intelligence? As represented in Figure 5, at its highest level, AI problems cluster around several classes. The first of these concerns the problem of correctly **parsing inputs**, which deals with issues of perception; computer vision; natural language processing; and taking appropriate cues from social intelligence; the second of these involves correctly **planning & executing outputs** ('behavior'), which involves appropriate processes for knowledge representation, prioritization, planning and, in embodied AI systems, robotics (system motion; collision avoidance; manipulators).

2.3.2. Machine Learning: 5 Schools of Thought

Within the distinct subfields of AI, however, there is one field which has been the most responsible for the recent advances in AI system implementations which have driven this third, sustained AI spring: this field revolves around approaches to achieve **machine learning** – helping an AI system learn to identify deep, hidden patterns in existing datasets, or learn to match specific features in data with specific responses or outputs.[102] There are, broadly speaking, **five main 'schools' of artificial intelligence machine learning**, taking their main inspiration from different fields of science.[103]

various brain dynamics that ultimately lead to a conscious state, see Stanislas Dehaene, *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (New York, New York: Viking, 2014).

101 Based on Nazre and Garg, "A Deep Dive in the Venture Landscape of Artificial Intelligence and Machine Learning.", slide 3

102 Note that these problem classes frequently overlap: computer vision frequently utilizes machine learning to achieve the pattern-matching of imagery with certain tags. However the key point is that such machine learning approaches are the more broadly applicable learning functions. Computer vision, as an 'input' problem, is therefore a subset of the 'learning problem' of machine learning.

103 This typology is derived from the work of machine learning expert Pedro Domingos. Cf. Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, 1 edition (New York:

The first, the aforementioned *connectionists,* take their cue from neuroscience, seeking to study how neurons work and encode knowledge by strengthening synapses between them. As a result, they have developed the neural-net algorithm called backpropagation, which learns from user-provided samples ('correct' input/output pairs). The neural net feeds inputs (such as pictures) through layers of artificial neurons until they emit a final output, then checks whether this output is in accordance with the pre-specified ('correct') output in the training data, a process which is continued until the algorithm becomes able to correctly predict the correct output even for new inputs (in effect demonstrating inductive logic).

Having revived the once-defunct research agenda which originated in the work on perceptrons, connectionists are currently arguably the most famous and successful school, having pioneered the backpropagation learning algorithms which, under the name of 'deep learning', have driven much of the success and applications pioneered by AI giants such as Google, Facebook and Baidu – and which are used for image- and voice recognition, as well as choosing search results and ads. There is an irony in the fact that some of the greatest successes in the quest for non-human intelligence have therefore been driven by attempts which rely on a biomimicry which seeks to approximate, or at least draws inspiration from, the human neural architecture. Nonetheless, in spite of the connectionist success, there are still considerable limits to deep learning systems, as argued by other camps in machine learning. As such, *'Evolutionists'* take their cue from evolutionary biology, and believe they can build AI by emulating (and speeding up) natural selection in digital environments, through genetic algorithms.

Evolutionary programming has had a long and storied history in adaptive design, assembling patents from the generation of non-intuitive designs for devices such as radio receivers, amplifiers and, most recently, 3D printers.[104]

Conversely, other machine-learning researchers argue that instead of biology (whether neuroscience or evolution), we should instead work from 'first principles' in computer science and logic. As such, *Bayesians* are a school of thought which is inspired by statistics; Bayesian networks (or 'belief networks') try to encode probability estimates for a large number of different competing hypotheses, with respective belief probabilities updated as new information becomes available.[105] Though capable of being deployed separately,

Basic Books, 2015).; and Pedro Domingos, "The Race for the Master Algorithm Has Begun," *WIRED UK*, 2016, http://www.wired.co.uk/article/master-algorithm-pedro-domingos.. For an illuminating interview, see Pedro Domingos, The Knowledge Project: Pedro Domingos on Artificial Intelligence, interview by Shane Parrish, Podcast, August 30, 2016, http://theknowledgeproject.libsyn.com/pedro-domingos-on-artificial-intelligence.. See also Cade Metz, "AI's Factions Get Feisty. But Really, They're All on the Same Team," *WIRED*, 2017, https://www.wired.com/2017/02/ais-factions-get-feisty-really-theyre-team/..

104 Cf. Domingos, "The Race for the Master Algorithm Has Begun.". For an overview of this field, see John R. Koza, "Survey of Genetic Algorithms and Genetic Programming," in *WESCON/'95. Conference record.'Microelectronics Communications Technology Producing Quality Products Mobile and Portable Power Emerging Technologies'* (IEEE, 1995), 589, http://ieeexplore.ieee.org/abstract/document/485447/.; for an account of its history, see Stephanie Forrest and Melanie Mitchell, "Adaptive Computation: The Multidisciplinary Legacy of John H. Holland," *Communications of the ACM* 59, no. 8 (August 2016), http://web.cecs.pdx.edu/~mm/jhh-cacm.pdf

105 cf. Judea Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988)

Bayesian networks can also enhance the decision making ability (and increase the transparency) of connectionist neural networks, and as such will likely be incorporated in the designs of self-driving cars and other AI-guided systems – systems for which it is important that the cause of an accident or error can be clearly ascertained.[106]

The fourth school, *symbolists*, is closest to the classic strand of knowledge-based AI, still pursuing a general-purpose learning algorithm which can freely combine rules and fill in the gaps in its knowledge. Ultimately, symbolic learning aims at mimicking the thought process of scientists themselves; looking at data to formulate hypotheses, testing these hypotheses against the data to refine them, and deducing new knowledge.

For instance, in 2014 'Eve', an artificially-intelligent 'robot scientist' based on symbolic learning and equipped with basic knowledge of molecular biology, made an autonomous discovery, showing that a compound with established anti-cancer properties might also help in the fight against malaria.[107]

Finally, the last school in machine learning is *analogisers*. These systems, inspired by human psychology, seek to operate on the basis of analogy – to match new cases with the most similar such situation which has been encountered in the past.

Ultimately, all of these schools have their own strengths, and encounter their own challenges, in their mutual pursuit of a single 'master algorithm': the foundation of machines truly able to learn in chaotic, new environments. It is certainly clear that we are still some time away from true or 'general' artificial intelligence, and there may yet be hitches or barriers for many of these approaches. And yet, the sheer pace of development over the past few years, the qualitative improvements in the perception-, learning- and decision-capabilities of current systems, and the manner in which the different approaches to AI are increasingly able to catalyze and strengthen each other's research, all strongly suggest that something has shifted in the field of AI. One way or another, we are now on a highway towards increasingly more capable forms of machine intelligence – with all the danger and opportunity that entails.

106 Cade Metz, "AI Is About to Learn More Like Humans—with a Little Uncertainty," *WIRED*, 2017, https://www.wired.com/2017/02/ai-learn-like-humans-little-uncertainty/.

107 Kevin Williams et al., "Cheaper Faster Drug Development Validated by the Repositioning of Drugs against Neglected Tropical Diseases," *Journal of The Royal Society Interface* 12, no. 104 (March 6, 2015): 20141289, doi:10.1098/rsif.2014.1289.. See also University of Cambridge, "Artificially-Intelligent Robot Scientist 'Eve' Could Boost Search for New Drugs," *University of Cambridge*, February 4, 2015, http://www.cam.ac.uk/research/news/artificially-intelligent-robot-scientist-eve-could-boost-search-for-new-drugs.; Andy Extance, "Robot Scientist Discovers Potential Malaria Drug," *Scientific American*, February 2015, https://www.scientificamerican.com/article/robot-scientist-discovers-potential-malaria-drug/..