Chapter 1

**Google Brain**

*Fundamentals about google brain & Introduction to A.I.*

Links:

Google Brain Team : <https://ai.google/research/teams/brain/>

Google A.I. : https://ai.google/

* 1. **About Google brain**

Google Brain is a deep learning artificial intelligence research team at Google. Formed in the early 2010s. Google Brain combines *open-ended machine learning research* with systems engineering and Google-scale computing resources.

**1.1.1 History**

The so-called "Google Brain" project began in 2011 as a part-time research collaboration between Google Fellow Jeff Dean, Google Researcher Greg Corrado, and Stanford University professor Andrew Ng. Ng had been interested in using deep learning techniques to crack the problem of artificial intelligence since 2006, and in 2011 began collaborating with Dean and Corrado to build a large-scale deep learning software system, DistBelief, on top of Google's cloud computing infrastructure. Google Brain started as a Google X project and became so successful that it was graduated back to Google: Astro Teller has said that Google Brain paid for the entire cost of Google X.

In June 2012, the New York Times reported that a cluster of 16,000 computers dedicated to mimicking some aspects of human brain activity had successfully trained itself to recognize a cat based on 10 million digital images taken from YouTube videos. The story was also covered by National Public Radio and SmartPlanet.

In March 2013, Google hired Geoffrey Hinton, a leading researcher in the deep learning field, and acquired the company DNNResearch Inc. headed by Hinton. Hinton said that he would be dividing his future time between his university research and his work at Google.

**1.2 Projects**

**1.2.1 Artificial-intelligence-devised encryption system**

In October 2016, the Google Brain ran an experiment concerning the encrypting of communications. In it, two sets of A.I.'s devised their own cryptographic algorithms to protect their communications from another A.I., which at the same time aimed at evolving its own system to crack the A.I.-generated encryption. The study proved to be successful, with the two initial A.I.s being able to learn and further develop their communications from scratch.

In this experiment, three A.I.s were created: Alice, Bob and Eve. The goal of the experiment was for Alice to send a message to Bob, which would decrypt it, while in the meantime Eve would try to intercept the message. In it, the A.I.s were not given specific instructions on how to encrypt their messages, they were solely given a loss function. The consequence was that during the experiment, if communications between Alice and Bob were not successful, with Bob misinterpreting Alice's message or Eve intercepting the communications, the following rounds would show an evolution in the cryptography so that Alice and Bob could communicate safely. Indeed, this study allowed for concluding that it is possible for A.I.s to devise their own encryption system without having any cryptographic algorithms prescribed beforehand, which would reveal a breakthrough for message encryption in the future.

**1.2.2 Image enhancement**

In February 2017, Google Brain announced an image enhancement system using neural networks to fill in details in very low resolution pictures. The examples provided would transform pictures with an 8x8 resolution into 32x32 ones.

The software utilizes two different neural networks to generate the images. The first, called a "***conditioning network***," maps the pixels of the low-resolution picture to a similar high-resolution one, lowering the resolution of the latter to 8x8 and trying to make a match. The second is a "***prior network***", which analyzes the pixelated image and tries to add details based on a large number of high resolution pictures. Then, upon upscaling of the original 8x8 picture, the system adds pixels based on its knowledge of what the picture should be. Lastly, the outputs from the two networks are combined to create the final image.

This represents a breakthrough in the enhancement of low resolution pictures. Despite the fact that the added details are not part of the real image, but only best guesses, the technology has shown impressive results when facing real-world testing. Upon being shown the enhanced picture and the real one, humans were fooled 10% of the time in case of celebrity faces, and 28% in case of bedroom pictures. This compares to previous disappointing results from normal ***bicubic*** scaling, which did not fool any human.

***1.2.3 Google Translate***

The Google Brain project contributed to Google Translate. In September 2016, ***Google Neural Machine Translation (GNMT)*** was launched, an ***end-to-end learning framework***, able to learn from a large number of examples. While its introduction has increased the quality of Google-Translate's translations for the pilot languages, it was very difficult to create such improvements for all of its 103 languages. Addressing this problem, the Google Brain Team was able to develop a ***Multilingual GNMT*** system, which extended the previous one by enabling translations between multiple languages. Furthermore, it allows for ***Zero-Shot Translations***, which are translations between two languages that the system has never explicitly seen before- Google announced that Google Translate can now also translate without transcribing, using neural networks. This means that it is possible to translate speech in one language directly into text in another language, without first transcribing it to text. According to the Researchers at Google Brain, this intermediate step can be avoided using neural networks. In order for the system to learn this, they exposed it to many hours of Spanish audio together with the corresponding English text. The different layers of neural networks, replicating the human brain, were able to link the corresponding parts and subsequently manipulate the audio waveform until it was transformed to English text.

**1.2.4 Robotics**

Different from the traditional robotics, robotics researched by the Google Brain Team could automatically learn to acquire new skills by machine learning. In 2016, the Google Brain Team collaborated with researchers at Google-X to demonstrate how robotics could use their experiences to teach themselves more efficiently. Robots made about 800,000 grasping attempts during research. Later in 2017, the team explored three approaches for learning new skills:

* through reinforcement learning,
* through their own interaction with objects, and
* through human demonstration.

To build on the goal of the Google Brain Team, they would continue making robots that are able to learn new tasks through learning and practice, as well as deal with complex tasks.

**1.2.5 In Google products**

The project's technology is currently used in the

* Android Operating System's speech recognition system,
* photo search for Google+ and
* video recommendations in YouTube.

**1.3 Team and location**

Google Brain was initially established by Google Fellow Jeff Dean and visiting Stanford professor Andrew Ng. In 2014, the team included Jeff Dean, Quoc Le, Ilya Sutskever, Alex Krizhevsky, Samy Bengio and Vincent Vanhoucke. In 2017, team members include Anelia Angelova, Samy Bengio, Greg Corrado, George Dahl, Michael Isard, Anjuli Kannan, Hugo Larochelle, Quoc Le: Chris Olah, Vincent Vanhoucke, Vijay Vasudevan and Fernanda ViegasJ,Chris Lattner, who created Apple's new programming language Swift and then ran Tesla's autonomy team for six months joined Google Brain's team in August 2017.

Google Brain is based in Mountain View, California and has satellite groups in Accra, Amsterdam, Beijing, Berlin, Cambridge (Massachusetts), London, Montreal, New York City, Paris, Pittsburgh, Princeton, San Francisco, Tokyo, Toronto, and Zurich.

**1.4 Reception**

Google Brain has received coverage in Wired Magazine, the New York Times, Technology Review, National Public Radio, and Big Think.

**1.5 See also**

• Artificial intelligence

* Glossary of artificial intelligence
* Quantum Artificial Intelligence Lab - run by Google in collaboration with NASA and Universities Space Research Association
* TensorFlow.

**References**

1. *"Brain Team mission - Google Al"*
2. *Machine Learning Algorithms and Techniques Research at Google. Retrieved May 18: 2017*
3. *"Research at Google", research.google.com. Retrieved 2018-02-16.*
4. *"Google's Large Scale Deep Neural Networks Project". Retrieved 25 October 2015.*
5. *Jeff Dean and Andrew Ng (26 June 2012). "Using large-scale brain simulations for machine learning and A.I." its Official Google Blog. Retrieved 26 January 2015.*
6. *Markoff, John (June 25, 2012). "How Many Computers to Identify a Cat? 16,000". The New York Times. New York Times. Retrieved February 11, 2014.*
7. *Jeffrey Dean; et al. (December 2012). "Large Scale Distributed Deep Networks" (PDF). Retrieved 25 October 2015.*
8. *Conor Dougherty (16 February 2015). "Astro Teller, Google's 'Captain of Moonshots,' on Making Profits at Google-X". Retrieved 25 October 2015.*
9. *A Massive Google Network Learns To Identify — Cats"i£. National Public Radio. June 26, 2012. Retrieved February 11, 2014.*
10. *Shin, Laura (June 26, 2012). "Google brain simulator teaches itself to recognize cats". SmartPlanet. Retrieved February 11, 2014.*
11. *A "U of T neural networks start-up acquired by Google"^ (Press release). Toronto, ON. 12 March 2013. Retrieved 13 March 2013.*
12. *Abadi, Martin; Andersen, David G. (2016). "Learning to Protect Communications with Adversarial Neural Cryptography".*

*arXiv: 1610.069183. Bibcode:2016arXiv161006918Ae9.*

1. *Dahl, Ryan; Norouzi, Mohammad; Shlens, Jonathon (2017). "Pixel Recursive Super Resolution". arXiv: 1702.007833. Bibcode:2017arXiv170200783D&.*
2. *"Google Brain super-resolution image tech makes "zoom, enhance!" real"^. arstechnica.co.uk. 2017-02-07. Retrieved 2017-05-15.*
3. *"Google just made 'zoom and enhance' a reality ~ kinda". cnet.com. Retrieved 2017-05-15.*
4. *"Google uses A.I. to sharpen low-res images". engadget.com. Retrieved 2017-05-15.*
5. *Schuster, Mike; Johnson, Melvin; Thorat, Nikhil. "Zero-Shot Translation with Google's Multilingual Neural Machine Translation System"^. Google Research Blog. Retrieved 15 May 2017.*
6. *Reynolds, Matt. "Google uses neural networks to translate without transcribing". New Scientist. Retrieved 15 May 2017.*
7. *'The Google Brain team — Looking Back on 2016". Research Blog. Retrieved 2017-12-18.*
8. *"Speech Recognition and Deep Learning". Google Research Blog. August 6, 2012. Retrieved February 11, 2014.*
9. *"Improving Photo Search: A Step Across the Semantic Gap". Google Research Blog. June 12, 2013.*
10. *'This Is Google's Plan to Save YouTube". Time. May 18, 2015.*
11. *Google Brain team website. Accessed 13.05.2017.* [*https://research.google.com/teams/brain/*](https://research.google.com/teams/brain/)
12. *Etherington, Darrell (Aug 14, 2017). "Swift creator Chris Lattner joins Google Brain after Tesla Autopilot stint" TechCrunch. Retrieved 11 October 2017.*
13. *"Research at Google" , research.google.com. Retrieved 2017-08-01.*
14. *Levy, Steven (April 25, 2013). "How Ray Kurzweil Will Help Google Make the Ultimate A.I. Brain". Wired. Retrieved February 11,2014.*
15. *Wohlsen, Marcus (January 27, 2014). "Google's Grand Plan to Make Your Brain Irrelevant". Wired. Retrieved February 11, 2014.*
16. *Hernandez, Daniela (May 7, 2013). "The Man Behind the Google Brain: Andrew Ng and the Quest for the New ". Wired. Retrieved February 11, 2014.*
17. *Hof, Robert (April 23, 2013). "Deep Learning: With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart". Technology Review. Retrieved February 11, 2014.*
18. *Regalado, Antonio (January 29, 2014). "Is Google Cornering the Market on Deep Learning? A cutting-edge comer of science is being wooed by Silicon Valley, to the dismay of some academics". Technology Review. Retrieved February 11, 2014.*
19. *"Ray Kurzweil and the Brains Behind the Google Brain"^. Big Think. December 8, 2013. Retrieved February 11, 2014.*

**1.6 Do we need a Newton in the field of artificial intelligence?**

Yes and no. The problem is rolled up with an old philosophical problem called “Free Will”. Humans claim to have this. Any human level robot would also have to have Free Will or at least believe it does as much as humans believe they have it

I could build an AI in 10 minutes that claims to have Free Will but I could not build one that looks to itself introspectively and finds that it must have a kind if free will because it sees not fundamental constraints to what it can do. No one can build an AI like that. We have no clue.

To continue along the current pathway where we make more and more useful machines like better self driving cars and Siri who makes fewer mistakes. We might even build a doctor AI that is better the any human physician. We are making progress on those lines. Our machines will get better but they will remain mindless machines that just run programs. The AI might cure cancer but it will not really care or know about cancer. It will just be an app running on a computer.

But no one has clue how to make a self aware general purpose AI. To build that will requires a scientific breakthrough. Engineering technical advances are not heading us in the direction, direction of machine who looks inward and determines feels as if it must have free will.

My dog is certainly is a self aware conscious being that can learn. Making a machine like a dog is not even on our horizon. My dog is not that smart. My iPhone understands more english. Being a self aware being is not the same as being intelligent . you can have one without much of the other.

Is a Newton required? Yes if you want a self aware AI who can act on it’s own interests and has what we call “Free Will”. No if all you want is a very useful machine that might be a simulation of an ultra-smart human.

Here is the $1,000,000 question: Can you write a program that does something that it was not programmed to do? You can’t simply roll dice and choose a behavior because the dice rolling and behaviors are programs. Maybe writing a few dozen kinds of steps then randomly combine them. No, you had to program that. THAT is the problem. Anything you write is just automation.

Can you write a program that TRUELY has free will? I want *proof* not just arguments. Maybe it will take another Issac Newton to not only prove it but invent a whole new field of mathematics needed for the proof and then after we all fully understand the human mind/body problem we build a human equivalent robot.

**1.7 The Turing test**

The Turing test, developed by Alan Turing in 1950, is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human. Turing proposed that a human evaluator would judge natural language conversationsbetween a human and a machine designed to generate human-like responses. The evaluator would be aware that one of the two partners in conversation is a machine, and all participants would be separated from one another. The conversation would be limited to a text-only channel such as a computer keyboardand screen so the result would not depend on the machine's ability to render words as speech. If the evaluator cannot reliably tell the machine from the human, the machine is said to have passed the test.

**1.8 The Top-Down Approach**

There are at least two major problems scientists have been facing for decades that have impeded their efforts to create robots: pattern recog­nition and common sense. Robots can see much better than we can, but they don’t understand what they see. Robots can also hear much better than we can, but they don’t understand what they hear.

To attack these twin problems, researchers have tried to use the “top-down approach” to artificial intelligence (sometimes called the “formalist” school or GOFAI, for “good old-fashioned AI”). Their goal, roughly speaking, has been to program all the rules of pattern recog­nition and common sense on a single CD. By inserting this CD into a computer, they believe, the computer would suddenly become sell- aware and attain humanlike intelligence. In the 1950s and 1960s great progress was made in this direction, with the creation of robots that could play checkers and chess, do algebra, pick up blocks, and so forth. Progress was so spectacular that predictions were made that in a few years robots would surpass humans in intelligence.

The top-down approach to artificial intelligence resulted in huge, clumsy robots that took hours to navigate across a special room that contained only objects with straight lines, that is, squares and triangles. If you placed irregularly shaped furniture in the room the robot would be powerless to recognize it.

For example, when we enter a room, we immediately recognize the floor, chairs, furniture, tables, and so forth. But when a robot scans a room it sees nothing but a vast collection of straight and curved lines, which it converts to pixels. It takes an enormous amount of computer time to make sense out of this jumble of lines. It might take us a frac­tion of a second to recognize a table, but a computer sees only a collec­tion of circles, ovals, spirals, straight lines, curly lines, corners, and so forth. After an enormous amount of computing time, a robot might fi­nally recognize the object as a table. But if you rotate the image, the computer has to start all over again. In other words, robots can see, and in fact they can see much better than humans, but they don’t un­derstand what they are seeing. Upon entering a room, a robot would see only a jumble of lines and curves, not chairs, tables, and lamps.

Our brain unconsciously recognizes objects by performing tril­lions upon trillions of calculations when we walk into a room-an activity' that we are blissfully unaware of. The reason that we are un­aware of all our brain is doing is evolution. If we were alone in the for­est with a charging saber-toothed tiger, we would be paralyzed if we were aware of all the computations necessary to recognize the danger and escape. For the sake of survival, all we need to know is how to run. When we lived in the jungle, it simply was not necessary for us to be aware of all of the ins and outs of our brain’s recognizing the ground, the sky, the trees, the rocks, and so forth.

In addition to pattern recognition, the second problem with the de­velopment of robots is even more fundamental, and that is their lack of “common sense.” Humans know, for example,

* Water is wet.
* Mothers are older than their daughters.
* Animals do not like pain.
* You don’t come back after you die.
* Strings can pull, but not push.
* Sticks can push, but cannot pull.
* Time does not run backward.

But there is no line of calculus or mathematics that can express these truths. We know all of this because we have seen animals, water, and strings, and we have figured out the truth by ourselves. Children learn common sense by bumping into reality. The intuitive laws of bi­ology and physics are learned the hard way, by interacting with the real world. But robots haven’t experienced this. They know only what has been programmed into them beforehand.

In the past, mathematicians have tried to mount a crash program that could amass all the laws of common sense once and for all. The most ambitious attempt is CYC (short for encyclopedia), the brainchild of Douglas Lenat, the head of Cycorp.

But since the firm’s founding in 1984, its credibility has suffered from a common problem in AI: making predictions that generate headlines but are wildly unrealistic. Lenat predicted that in ten years, by 1994, CYC would contain 30 to 50 percent of “consensus reality.” To­day CYC is not even close. As the scientists of Cycorp have found out, millions and millions of lines of code need to be programmed in order for a computer to approximate the common sense of a four-year-old child. So far the latest version of the CYC program contains only a pal­try 47,000 concepts and 306,000 facts. Despite Cycorp’s regularly opti­mistic press releases, one of Lenat’s coworkers, R. V. Guha, who left the team in 1994, was quoted as saying, “CYC is generally viewed as a failed project.... We were killing ourselves trying to create a pale shadow of what had been promised.”

In other words, attempts to program all the laws of common sense into a single computer have floundered, simply because there are so many laws of common sense. Humans learn these laws effortlessly be­cause we tediously continue to bump into the environment throughout our lives, quietly assimilating the laws of physics and biology, but ro­bots do not.

**1.9 The Bottom-Up Approach**

Because of the limitations of the top-down approach to artificial intel­ligence, attempts have been made to use a “bottom-up” approach in­stead, that is, to mimic evolution and the way a baby learns. Insects, for example, do not navigate by scanning their environment and reducing the image to trillions upon trillions of pixels that they process with su­percomputers. Instead insect brains are composed of “neural net­works,” learning machines that slowly learn how to navigate in a hostile world by bumping into it. At MIT, walking robots were notori­ously difficult to create via the top-down approach. But simple buglike mechanical creatures that bump into the environment and learn from scratch can successfully scurry around the floor at MIT within a mat­ter of minutes.

Rodney Brooks, director of MIT’s famed Artificial Intelligence Lab­oratory, famous for its huge, lumbering “top-down” walking robots, became a heretic when he explored the idea of tiny “insectoid” robots that learned to walk the old-fashioned way, by stumbling and bump­ing into things. Instead of using elaborate computer programs to math­ematically compute the precise position of their feet as they walked, his insectoids used trial and error to coordinate their leg motions us­ing little computer power. Today many of the descendants of Brooks’s insectoid robots are on Mars gathering data for NASA, scurrying across the bleak Martian landscape with a mind of their own. Brooks believes that his insectoids are ideally suited to explore the solar system.

One of Brooks’s projects has been COG, an attempt to create a me­chanical robot with the intelligence of a six-month-old child. On the outside COG looks like a jumble of wires, circuits, and gears, except that it has a head, eyes, and arms. No laws of intelligence have been programmed into it. Instead it is designed to focus its eyes on a human trainer, who tries to teach it simple skills. (One researcher who be­came pregnant made a bet as to which would learn faster, COG or her child by the age of two. The child far surpassed COG.)

For all the successes in mimicking the behavior of insects, robots using neural networks have performed miserably when their pro­grammers have tried to duplicate in them the behavior of higher or­ganisms like mammals. The most advanced robot using neural networks can walk across the room or swim in water, but it cannot jump and hunt like a dog in the forest, or scurry' around the room like a rat. Many large neural network robots may consist of tens to perhaps hundreds of “neurons”; the human brain, however, has over 100 bil­lion neurons. C. elegans, a very simple worm whose nervous system has been completely mapped by biologists, has just over 300 neurons in its nervous system, making its nervous system perhaps one of the simplest found in nature. But there are over 7,000 synapses between these neurons. As simple as C. elegans is, its nervous system is so com­plex that no one has yet been able to construct a computer model of this brain. (In 1988 one computer expert predicted that by now1 we should have robots with about 100 million artificial neurons. Actually, a neural network with 100 neurons is considered exceptional.)

The supreme irony is that machines can effortlessly perform tasks that humans consider “hard,” such as multiplying large numbers or playing chess, but machines stumble badly when asked to perform tasks that are supremely “easy” for human beings, such as walking across a room, recognizing faces, or gossiping with a friend. The rea­son is that our most advanced computers are basically just adding ma­chines. Our brain, however, is exquisitely designed by evolution to solve the mundane problems of surv ival, which require a whole com­plex architecture of thought, such as common sense and pattern recog­nition. Survival in the forest did not depend on calculus or chess, but on evading predators, finding mates, and adjusting to changing envi­ronments.

MIT’s Marvin Minsky, one of the original founders of AI, summarizes tiie problems of AI in this way: “*The history of AI is sort of funny be­cause the first real accomplishments were beautiful things, like a ma­chine that could do proofs in logic or do well in a calculus course. But*

*then we started to try to make machines that could answer questions about the simple kinds of stories that are in a first-grade reader book. There’s no machine today that can do that*.”

Some believe that eventually there will be a grand synthesis be­tween the two approaches, the top-down and bottom-up, which may provide the key to artificial intelligence and humanlike robots. After all, when a child learns, although he first relies mainly on the bottom- up approach, bumping into his surroundings, eventually he receives instruction from parents, books, and schoolteachers, and learns from the top-down approach. As an adult, we constantly blend these two ap­proaches. A cook, for example, reads from a recipe but also constantly samples the dish as it is cooking.

Hans Moravec says, *“Fully intelligent machines will result when the mechanical golden spike is driven uniting the two efforts,”* proba­bly within the next forty years.

**1.10 Machine learning (ML)**

**Machine learning** (**ML**) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysisthrough unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

**1.11 Deep learning**

**Deep learning** (also known as **deep structured learning**or **hierarchical learning**) is part of a broader family of machine learning methods based on artificial neural networks. Learning can be supervised, semi-supervised or unsupervised.

Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases superior to human experts.

Neural networks were inspired by information processing and distributed communication nodes in biological systems. ANNs have various differences from biological brains. Specifically, neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analog.

**1.12 Data mining**

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases” process, or KDD. Aside from the raw analysis step, it also involves database and data management aspects, data pre­processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. The difference between data analysis and data mining is that data analysis is used to test models and hypotheses on the dataset, e.g., analyzing the effectiveness of a marketing campaign, regardless of the amount of data; in contrast, data mining uses machine-­learning and statistical models to uncover clandestine or hidden patterns in a large volume of data.

The term "data mining" is in fact a misnomer, because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (*mining*) of data itself. It also is a buzzword and is frequently applied to any form of large-scale data or information processing(collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence (e.g., machine learning) and business intelligence. The book *Data mining: Practical machine learning tools and techniques with Java* (which covers mostly machine learning material) was originally to be named just *Practical machine learning*, and the term *data mining* was only added for marketing reasons. Often the more general terms (*large scale*) *data analysis* and *analytics* – or, when referring to actual methods, *artificial intelligence* and *machine learning* – are more appropriate.

The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps.

The related terms *data dredging*, *data fishing*, and *data snooping* refer to the use of data mining methods to sample parts of a larger population data set that are (or may be) too small for reliable statistical inferences to be made about the validity of any patterns discovered. These methods can, however, be used in creating new hypotheses to test against the larger data populations.

