Research Paths Towards Thinking A.I.

Shiv Contractor (Ed.)
Michael Moosmeier (Ed.)
Sebastian Schuon (Ed.)

Contents

Foreword

Klaus Diepold

These days the academic community spends significant amount of energy discussing innovations in teaching and learning. In particular, the advent of digital tools causes a frenzy about the introduction of digital innovations in teaching. Such digital innovations include video based teaching materials aka MOOCs (massive open online courses), various concepts for "inverted classrooms", the use of tablets, clickers, online voting and feedback and other computer-based tools as well as using all sorts of digital media and much more.

I am all for experimenting with new formats of teaching, and a reasonable use of digital tools can be helpful and inspiring. Any new way of teaching, be it digitally supported or based on less technically demanding methods and tools shall be evaluated if they actually help students to be better educated. This last question taken by itself is already a challenge, because we lack reliable methods to measure the success of education. Looking at examination results is certainly not enough, just as much as student evaluations are seldom more than an assessment of the well-being. These methods are prone to all sorts of secondary influences rendering the results close to useless. One measure of success that I regard as rather reliable and instructive is to what amount I succeed in activiating the students in a class. By activation I mean that students are actually going out on their own acquiring facts and knowledge, digesting and actively discussing the material they found among themselves. This way they create new knowledge and experience. Even though this knowledge may not be new to the world or the scientific community, it is new to the students and it is active in the minds by virtue of the process the students went through. Besides the new knowledge they pick up, they also gain experience in the process of collecting and digesting knowledge, being critical and constructive as well as experiencing the power of communication and intellectual exchange.

The book you hold in hands is one of these didactical experiments. The students were set up to collect, acquire, digest and produce new knowledge for themselves. This year we ventured to explore future research directions for Artificial Intelligence. This also serves for the students as a preparation for choosing a topic for their final project before concluding their Master's degree at TUM.

One aspect that I find instructive to measure the success of this course format is the amount of effort and time students invest in the course voluntarily, without me, the instructor, urging or requiring them to work harder. They just do it, because they feel inspired and because they are curious. Funny enough, this extra engagement on the students side earned me an exhorting message from my Dean of Study, who felt that our students were overly burdened by the course. This exhortation was the result of the students' course evaluation, where students indicated that they've worked many more hours than accounted for the by the assigned credits. However, the students also acknowledged that they loved the course in spite of the long hours. To me, this is a strong indicator that we did something right. I am exuberantly happy about the outcome of the course, which is exactly this book and I am proud of the students who proved very convincingly that they are maturing academically and that they can create original research-related output way beyond reproducing scripts and lecture notes. In spite of this personally felt success, I still had to promise to the Dean of Study that next year we will return to a format with reduced work load. I am not sure if the students can keep up to this promise if I succeed to fire them up to a similar amount, possible with one of my next didactical experiments.

Introduction

Klaus Diepold
Michael Moosmeier
Sebastian Schuon

Dear Reader,

Welcome to the proceedings of the seminar “Brain, Mind and Cognition”. This section provides background information about the genesis of this book, why it exists, how it was conceived and how it was finally produced. This account shall serve to also communicate the didactical concept underlying the process that eventually produced the book. From this account it is also clear that the content of the book represents a snapshot of current thinking about the future research on intelligent systems as seen by students just about to enter the world of science. The book is the result of a one-semester seminar dedicated to the topic of Brain, Mind and Cognition. The seminar is an elective course in the curriculum of the Master of Science in Electrical and Computer Engineering, which is offered by the Faculty of Electrical and Computer Engineering of the Technical University of Munich. The seminar consists mainly of weekly meetings for 2 hours to discuss and work on the subject. The work is organized as group discussions and team work. Students have to read, write, review and present their findings on a weekly basis. To this end new digital e-learning tools and methods were employed along with discussion styles such as world-cafe or speed-dating discussions. Presentations were given as Pecha Kuchas[1], which facilitate for highly focused and condensed presentation sessions, while also creating a spontaneous and fun atmosphere. Throughout the seminar participating students should learn fundamental aspects of scientific research along with honing the skills in oral and written communication in a scientific or technical field. The intention for the students in the course is to develop ideas and paths for future research in the field leading towards insights and methods necessary to design and implement intelligent systems in the broader sense of the word. One of the fundamental aspects is to better understand what “thought” actually is and how the technical implementation of a thinking machine may move forward as many scientific and technical sub-disciplines evolve into the future.

The seminar in total was structured in three major stages:
1st Stage: Individual Reading, Writing, Reviewing, Discussing
2nd Stage: Team - Researching, Presenting, Discussing
3rd Stage: Team - Projecting - Writing - Reviewing

During the first stage the seminar started out with all students jointly reading one book by Eric Baum entitled “What is thought?”. Between subsequent meetings students agreed on a set of chapters to read until the next meeting. Students were also asked to reflect on their reading by writing short essays along some high-level guiding questions. Each essay had the size of 5000 characters (incl. spaces). The students uploaded their essays before the next meeting using an e-learning platform (Moodle). Furthermore, students were randomly assigned 5 essays of their fellow students to read and review. During this stage, the students discussed the content of the book during the weekly meetings, using various forms of discussion, such as world cafe, speed dating discussions, fishbowl discussions. This form of reading, writing, reviewing and discussing generated a shared domain of knowledge to facilitate the later stages in the process. It also conveyed fundamental information about the field of study on intelligent systems or Artificial Intelligence. This first stage took about 4 weeks. The second stage started with a workshop where the students tried to identify major fields of science and technology, which were considered essential to push the topic of intelligent systems into the future. By the end of the workshop, the student agreed on a list containing the dominating fields and domains. Subsequently, students could assign

[1] `https://www.pechakucha.org/faq` (accessed April 26, 2019)

themselves to one of these items on the list to further study the field in more detail. During the next five weeks, the teams of students researched their chosen field compiling information about the state of the art and the major trends. During the weekly contact hours one student per team delivered a Pecha Kucha presentation highlighting the group's findings during the past week for all others to understand and participate. The presentations were followed by discussions on open points. A Pecha Kucha presentation consists of 20 slides, where each slide is shown for 20 seconds. A complete Pecha Kucha presentation hence takes 6:40 minutes and forces students to be concise and to focus on essentials. The one purpose of the presentations is to disseminate the collected information to the fellow students. Another objective is to sharpen the sense of the presentes to think about their target audience and to taylor the amount and the level of detail of the presentations to match the expectation of their target audience.

The book is written by students mainly for students. It does not claim to contain and communicate ultimate truths, but rather tries to project current facts and trends into future directions of research based on a intense investigation of trends and possibilities. The book may prove to be a helpful tool to orient students interested in the study and the development of intelligent systems, AI, machine learning and so on as a basis to narrow down on a topic for their Master thesis projects or even beyond. Not least of it, the book may also display interesting ideas and anticipations, which may be helpful even for more seasoned researchers to communicate with young people and transmit the excitement for science and research on intelligent systems using a language and level that students can digest and appreciate. We hope, you the reader, will find inspiration in the chapters and material to further lead the discussion about Thought and Consciousness of machines. If you have any remarks on this book, our process or the course itself, we would love to hear from you!

Chapter 1

Data Nurturing AI

Michael Hausler
Jannick Lippert
Thomas Lober
Martin Satzke
Ke Shuai

1.1 Abstract

In this chapter, we will mainly discuss how an increasing amount of data affects artificial intelligence (AI). First, the historical development and past milestones of data storage and data analysis are presented; starting off as cave paintings to databases and nowadays, technologies such as automatic clustering. The amount of gathered data has significantly increased in all avenues over the last few decades and this progress will accelerate in the future. It is driven by trends like the increasing use of mobile devices, social networks, and the internet of things (IoT). Apart from the amount of the data, the diversity and complexity of the data has also been increasing. Therefore, we have described different approaches to analyze structured as well as unstructured data. In this context, the term 'Big Data' is discussed with its dimensions and technical approaches like data mining and cloud computing. We have also mentioned the impact of Deep Learning (DL) approaches for automatically extracting features from data as well as for pattern recognition. This results in the exposition of the current challenges in dealing with big amounts of complex data. Features in high-dimensional spaces have to be extracted and analyzed to make the data usable for applications in the field of machine learning (ML). In addition to new techniques for learning highly-dimensional and imbalanced data, privacy and security issues are also evaluated as research trends for the future. Future opportunities for ML with Big Data include the topic how the rising amount and complexity of data influences AI approaches.

1.2 Past Milestones

At the onset of human life, there was information transfer only in spoken form. The first way of making this spoken data tangible was painting on cave walls. Techniques on how and where to store data were developed, from animal skin and walls to papyrus and eventually, paper. With the invention of punch cards in the 18$^{\text{th}}$ century, information storage improved. In 1890, the Hollerith tabulating machine was able to read punch cards automatically and to speed up the process of the American census. Technology further improved with the invention of magnetic drums and magnetic tapes by the middle of the 20$^{\text{th}}$ century, representing the early form of computer memory [41].

Since ancient times, improvements in data storage technology have been necessary to satisfy the human urge to store and manage data. For development in terms of capability, the technology of databases had to be improved. The hardware-specialized database machines in the 1970s were replaced by software-based database systems that could be run on general-purpose computers. Digital technology caused a massive magnification of data volume in the late 1980s from several gigabytes to even a terabyte. For a single computer system, the maximum capabilities for storing and processing data were reached, which cleared the way for data parallelization. The idea of apportioning data and tasks to a cluster of separate hardware laid the foundation for several architec-

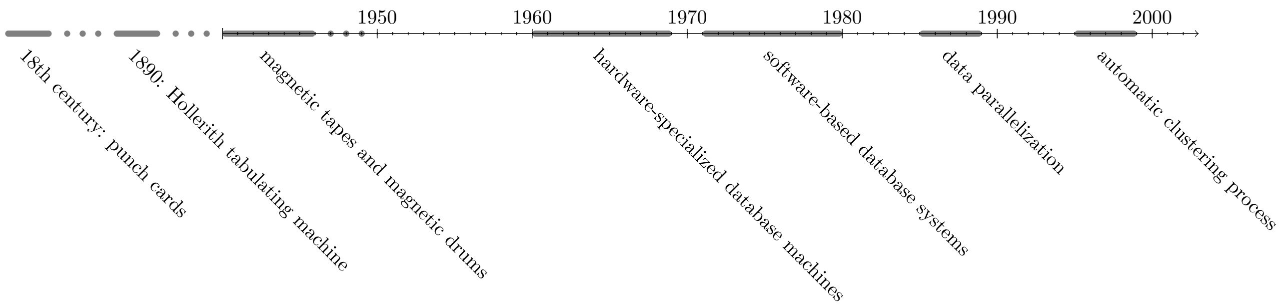


Figure 1.1: Timeline with main milestones of data analysis; *source: own image*

tures of parallel databases.

In the late 1990s, with the rise of the internet and the associated massive magnitude of semi-structured and unstructured data up to petabytes, parallel databases in the common way came to their limitation. Developments like Google File System and MapReduce programming model enabled an automatic clustering process to parallelize and compute processes on a variety of servers [22]. The presented milestones in the field of data analysis are shown in figure 1.1.

1.2.1 Analysis of Databases

The approach of extracting information out of collected data is called analysis of databases. The analysis breaks the whole into separable components, which the user applies in decision-making. This data set can be deterministic like survey points of a physicist who is searching for a fitting curve. Furthermore, the data is applied in non-deterministic situations, e.g. extracting data from a customer walking through a store. The basic steps that data analysis runs through are information extracting, cleaning, exploration, modelling, and implementation. First, some data has to be chosen and must be converted in a computer usable format. In the cleaning step, duplicated or bad data has to be removed. Afterwards, the data can be explored. It is important to understand the messages contained by the data. Extracting and cleaning may be applied iteratively for a data set that fits better to the corresponding use case. Once the data is accurate, mathematical formulas or models (e.g. regression) may be applied to extract correlation or causation. The last step is to implement the procedure. A program fed by some input has to deliver information that benefits the user [4].

1.2.2 Introduction to Big Data

Big Data is a blanket term for the non-traditional strategies and technologies needed to gather, organize, and process insights from large data sets. Big Data refers to data sets which have a size beyond the capabilities of the current database technology. While the problem of working with data that exceeds the computing power or storage of a single computer is not new, the pervasiveness, scale, and value of this type of computing has greatly expanded in recent years. Like Steve Lucas, the former Global Executive Vice President and General Manager of SAP Database & Technology said in [29], companies have always stored copious amounts of information. However, currently, the focus has shifted to utilizing this data to create value. Today, we have technologies like Hadoop, which make it possible to access a tremendous amount of data in a functional and practical way with an aim of generating value from this data. The availability of inexpensive hardware makes this easier and more feasible to quickly retrieve and process information at lower costs than ever before. As can already be observed, it will have a big impact in the next decades on several fronts such as social sciences, business, health care, entertainment, and sports.

Despite having used the term ‘Big Data’ so often, we have no general definition. Defining this term is difficult due to the fact that projects, vendors, practitioners, and business professionals use it differently. Keeping that in mind, generally speaking, Big Data is the category of computing strategies and technologies that are used to handle large data sets. The common scale of Big Data sets is constantly shifting and may vary significantly depending on the organization [29].

Another definition of Big Data is given by IBM. According to them, Big Data is characterized by either or all of the three ‘V’s listed as follows [33]:

- Volume: A large amount of data is generated from various sources. New data sources like IoT (Internet of Things) and social networks are factors that make the volume rise significantly.

- Variety: Multiple types and formats of data, consisting of structured and unstructured data, have to be analyzed.

- Velocity: Increased velocity in data generation also requires an increase in the velocity of data processing. Big Data is more dynamic than conventional data and influences the frequency of decision making. However, these decisions also react on the processing chain of the generated data. This means that there is a new dimension in the velocity of data.

After an appropriate definition of the term has been made, a closer look into the world of Big Data can be taken. Because of the technological development in the last years, humans are now at a level where they are socially connected and tracked to a great extent. All actions performed are recorded indelibly. We currently own devices that can observe and analyze every movement we make and this is just the beginning. This immense amount of available data is the basis of Big Data. All the milestones mentioned above have led to this level of data processing. Despite the fact that 'Big Data' as a buzzword is approaching its peak in the life cycle, there are reasons to believe that this trend will only accelerate in the coming years [13].

With the development of the Internet of Things, we are collecting even more data. The billions of sensors are collecting new data every second in all our devices. With this cornucopia of data, produced image, audio, and video material can be administered and analyzed with new technologies and approaches. All this offers a whole new way of understanding the world. The current data deluge is revolutionizing the way research is carried out and resulting in the emergence of a new fourth paradigm of science based on data-intensive computing. The actual science is strongly influenced by data, leading to a new data-centric way of conceptualizing, organizing, and carrying out research activities. This could pave the way to discovering new approaches to solve problems that were previously considered extremely hard or, in some cases, impossible. Also, it would not be presumptuous to believe that this new form of research may eventually lead to serendipitous discoveries [37]. In the following section, there will be an overview of the current research concerning techniques and technology for data analysis and storage.

1.3 Current Research

After introducing the history of data analysis and the buzzword Big Data, we want to present different topics in current research. Fields like Data Mining, MapReduce, and Cloud Services connect data with artificial intelligence in some way or the other.

1.3.1 Data Mining

Data Mining is a subset of 'Knowledge Discovery in Databases'. It describes the algorithms, methods, and systems to automatically extract patterns, correlations, and predictions from data [44]. The processes of data mining provide the ability to extract complex patterns from large and diverse data sets that would be impractical or not even possible to obtain manually [8].

In general it is not limited to Big Data, but with the its rise in the the aforementioned three dimensions, there are several challenges in terms of software and hardware that have to be dealt with [44]. The process of data mining can be divided into the following three tiers, as summarized in [44]:

Tier I: Data Mining Platform Big Data mining requires the computation of less structured data that goes beyond the scope of common databases. Several different data sources with diverse data formats like text, image, and video data have to be mined for extracting useful information. Especially, for hidden relationships in unstructured data, high computing performance is needed. This task is deployed by cluster computers, where one single data mining task is divided into several small tasks, which are split up between the computing nodes of the cluster. Parallel programming tools like MapReduce and Enterprise Control Language make this possible [44].

Tier II: Big Data Semantics and Application Knowledge The process of designing algorithms and systems for data mining can be improved, if application and domain knowledge is used in the right places. By improving the structure and performance of these systems through expert knowledge, patterns and predictions become more accurate. In the field of data mining, the aspect of regulations has to be considered. It is essential for achieving benefit of Big Data, to share data and results between multiple parties, but on the other hand, there are privacy concerns, when sensitive and personal information is shared. The field of privacy preserving data mining tries to achieve middleground for these two concerns. By restricting the access to sensitive data to a defined group of people through certification and access control in addition to anonymizing data fields to avoid association of data to an individual, a compromise between data sharing and protecting privacy is made [44].

Tier III: Big Data Mining Algorithms The challenge of Big Data Mining systems is to deal with unstructured data by combining and extracting the knowledge of several, heterogeneous data sources and formats. Uncertain and incomplete data impede this process. Through fusion processes of several data sources, the relevance and trustfulness of information must be correlated to achieve a global optimization goal for the needed application. This current challenge of generating structured data will be discussed in the next section [44]. One driver to manage challenges in terms of data mining are competitions. A famous example is the '$1 million Netflix Prize', which was granted in 2006. The challenge was to improve the company's recommendation system for video rental and streaming services by 10 percent. This recommendation system is a data mining algorithm based on predictions from previous streams and rentals. Competitions are a good way to foster new ideas and solutions for existing problems in data mining. Although the prize was awarded only to one team in 2009, the competition incited many researchers across the globe to tackle this topic and innovate in the field of data mining [8].

1.3.2 Generating Structured Data

Like explained earlier, see 1.2.2 on page 12, the amount of data sources and collected data is already huge and will further grow in the future. Moreover, the definition of Big Data introduced by IBM states variety as a key indicator of Big Data. The data formats are getting more diverse as the data comes from a growing number of sources. The motivation behind interpreting the data is to get a structure into the gathered data. Having achieved this, the data will be able to be interpreted to achieve significant results. In the past few years, research has worked on different approaches to structure data. In the following, two of them are presented. While the first concentrates on a general attempt to structure different types of data in one step, the second only uses specific data formats to extract information.

MapReduce and Hadoop A tool with universal data input to generate structured data and additionally analyze it is MapReduce. It is an approach for data processing with one single master that assigns tasks to all connected slaves in the cluster. The efficient parallelism is the biggest feature of this tool, allowing the control of 'Big Data' in a reasonable time. This is done by breaking the large data set into small units. These units are processed by several parallel nodes [10]. The programmer defines a map function with a key and value pairs as well as a reduce function to merge all values labeled with one key. The Map and Reduce functions are usually simple and equivalent to usual database commands. In this manner, MapReduce can process and combine data from different data storage systems as it groups data with the same key together. As open source implementation of MapReduce, Apache Hadoop is often used in both academic and industrial applications [11].

Structuring Social Media Data MapReduce and Hadoop are general approaches to structure and analyze a lot of unstructured, gathered data. By defining functions, these tools are adapted for a special problem. A different approach is to design tools for a specific data format and analysis. Therefore, only selected data is used from one specific data format. Especially, social media is a growing data supplier with a lot of potential, where the generated data is often analyzed with specific tools. Social media gives the possibility to get insights into human behaviour with relatively little effort as compared to traditional sociological methods, e.g. it is possible to observe mood oscillations of millions of people in 84 countries [40].

While the range of applications is wide, Big Data research focuses on the platform Twitter. There are no direct messages and very few private profiles, which lead to a huge available data set. The data has a simple and clean structure with very few basic functions and a maximum of 140 characters per tweet. This makes it easy to process and analyse compared to other platforms like Facebook [40]. As done in [6], the 'Google-Profile of Mood States'-tool was used to analyze public mood over a certain time based on twitter data in six dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). As positive mood encourages to run an economical risk, the researches checked whether there is a connection between the mood in tweets and the rate of the Dow Jones Industrial Average. They showed that they can significantly improve models to predict the stock price by the dimensions 'Calm' and 'Happiness' [6].

1.3.3 Cloud with Big Data

By cloud, we mean an infrastructure that provides on-demand resources or services over the Internet, usually, at the scale and reliability of a data center. A storage cloud provides storage services (block or file-based services); a data cloud provides data management services (record-based, column-based or object-based services); and a compute cloud provides computational services. Often, these are stacked together to serve as a computing platform for developing cloud-based applications [20].

Cloud computing refers to both the applications delivered as services over the Internet as well as the hardware and systems software in the data centers that provide those services. The services themselves have for long been referred to as Software as a Service (SaaS) [3]. Cloud computing platforms have been designed with two important restrictions: First, clouds have assumed that all the nodes in the cloud are co-located, i.e. within one data centre or that there is a relatively small bandwidth available between the geographically distributed clusters containing the data. Second, these clouds have assumed that individual inputs and outputs to the cloud are relatively small, although the aggregate data managed and processed is extremely large [20].

Our assumption is that there are high-speed networks (10 Gb sK1 or higher) connecting various geographically distributed clusters and the cloud must support both the ingestion and the return of relatively large data sets [20]. Cloud computing involves the delivery of computing services to end-users with the promise of faster provisioning, dynamic allocation, improved manageability, and reduced maintenance. Typically, cloud computing is delivered in a service model, such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), etc. [3].

An emerging trend is to upload data to the cloud. Data as a Service (DaaS) is based on the concept that data can be provided on demand to the user regardless of geographic or organizational separation of the data provider and consumer. Moving information systems such as building management systems to the cloud may have several benefits in terms of cost and service provision [3].

However, there are several obstacles for Cloud Computing. One being 'Business Continuity and Service Availability' - organizations worry about whether utility computing services will have adequate availability. Another one is 'Data Confidentiality' - cloud users face security threats both from outside and inside the cloud. Also, 'Data Transfer Bottlenecks' could be a problem. Applications continue to become more data-intensive. As a result of wanting to avoid slow online transfer, data transfer costs are an important issue. Lastly, 'Bugs in Large-Scale Distributed Systems' - errors in these very large-scale distributed systems can be quite troublesome. A common occurrence is that these bugs cannot be reproduced in smaller configurations [3].

Many people may find cloud and data center quite similar but there are actually some obvious differences between them. In data center, data will be accessible as long as the local network remains stable. Having all or most of the data stored in one location makes it accessible to third parties and unwanted external agents, both virtually and physically. However, using cloud with online access, one will never lose track of his data as long as there is a stable internet connection. If the organization's location is compromised via fire, break-in, flooding, etc., the data will remain untouched and unharmed at its remote location. Anything online is more susceptible to virtual attack. For example, is it more likely that a hacker isolates a cloud storage system than a data center. Usually, most essential and critical data are placed in a data center and less confidential information on the cloud to be more easily accessible[1].

1.3.4 Applications of Deep Learning in Big Data Analytics

Standard artificial neural networks (ANN) consist of different connected neurons, which are little processors with an internal transfer function from input to output. By connecting those neurons in different layers, different stages are built. By activating the first stage with sensors perceiving the environment, a chain of neurons is activated dependent on the weights, which are implemented in the neurons to change the transfer function. The learning is done by finding weights that result in the desired behaviour of the Input-Output function. According

[1] www.networkspecialists.com/the-cloud-vs-the-data-center-whats-the-difference/ (accessed April 26, 2019)

to the problem, this requires more or less chains of computation as every step has one transfer function to change the current value [35]. DL was inspired by the automatic feature extraction from data of primary sensorial areas in the human brain [30]. While shallow neuronal networks with a few models have been used for decades, the aim of DL is to assign weights to neurons in many consecutive layers [35]. ANN became effective when efficient gradient descent methods for supervised learning methods were used to optimize the weights of a network.

The usage of backpropagation methods for networks with many layers was not that promising, see [43]. DL became helpful, especially in pattern recognition applications after using new unsupervised learning methods and new ML techniques like kernel machines [36]. The proceedings also made it interesting for more general reinforcement applications with unsupervised learning processes [35]. This shows, that DL is tightly linked with unsupervised learning, which promises to extract patterns from data without labelling it. This allows the process to extract a more complex representation from more complex data, which has been previously discussed in this chapter. The DL structure makes it possible to find global connections in data points and makes the learning process independent from human classification [30].

1.4 Current Challenges

Since Big Data technologies are in their infancy, there are still some issues that have to be solved. The following chapter introduces the current challenges, which Big Data research, especially analysis, is facing right now.

1.4.1 Significant Fields in Big Data Research

Access to a large amount of data and the previously introduced technologies in data analysis show great promise. Therefore, data science enjoys an immense popularity in various fields of academia, industry, and government. Now, let us take a look at the significance of Big Data research and which challenges the research is confronting at the moment.

Big Data has a strong significance on the development of nations. Certainly, Big Data is a sector of industry, which obviously increases the Gross Domestic Product (GDP). Moreover, Data is becoming a resource of modern society, which brings profit in use of a value-added chain, consisting of data acquisition, analysis, curation, storage, and usage. However, there is more potential in Big Data. The capacity of accumulating, processing, and utilizing vast amounts of data will become a new landmark of a country's strength [24]. The information can help the government to make political decisions. It is possible to collect data for identifying the health, education, living standard, and social environment of citizens in a certain area to guide the national development. Generally, it helps people to better understand the present and interpret the times in which they live. Every intelligence seeks to perceive the future, so does Big Data. For instance, it helps in finding diseases and cures or mitigation strategies for them [1].

Even if the exponentially growing volume of available data makes the Big Data approach possible, it is one of the toughest challenges. This large amount of information cannot be simply stored. Since the data has to be accessed from a database, the input/output speed matters too, especially for real time applications. On the one hand, both issues can be overwhelmed by development of better hardware. On the other hand, software technologies like MapReduce are decreasing data size. In the field of hardware, systems have to be developed, which better suit the requirements of Big Data. For software, the key is to develop technologies for analyzing data to obtain better knowledge, which reduce the data by selecting the data and effective features. These algorithms have to handle data with high complexity and be insusceptible to uncertainty and inconsistencies [1]. Ideally, these conclusions have the ability of parallelization. Therefore, they are usable in real time applications [24].

Another challenge is the preservation of privacy. Since the aim of Big Data is the generalization, individual-related data is insignificant. Regardless, there is a need for multi-level security which is protects privacy of individuals and security of enterprises as well as states [28]. A closer look into this challenge and its future trends are presented in section 1.5.1.

1.4.2 Challenges of Big Data Analysis

As mentioned above, Big Data promises new levels of scientific discovery and economic values. Previously, we have also described the difference between Big Data and the traditional small-, medium- or large-scale data. In the current part, we want to have a focus on the challenges of analyzing Big Data. Due to the drastic price drop to get a very Big Data collection, we have a massive amount of high-dimensional and unstructured data. The current trend of data being produced and stored more massively and cheaply is likely to remain or even accelerate in the future [17]. With this large amount of data, it has become important to develop a valid statistical analysis for this data. In general, there are two main goals of high-dimensional data analysis. The first one is to develop effective methods that can accurately predict the future observations. The second one is to gain insight into the relationship between the features and response for scientific purposes in parallel to the first goal[15]. Additionally, to achieve these goals, it is very important to understand the heterogeneity as well as the commonality across different sub populations.

Now, let us focus on the exact challenges that arise when we try to analyze the data. The characteristics of Big Data are high dimensionality and large sample size. With these features, there are three special challenges. Firstly, the high dimensionality brings noise accumulation, spurious correlations, and incidental homogeneity. Secondly, the high dimensionality combined with a large sample size creates issues such as heavy computational cost and algorithmic instability. Also, the massive samples in Big Data are typically aggregated from multiple sources at different time points using different technologies. This creates issues of heterogeneity, experimental variations, and statistical biases, requiring the development of more adaptive and robust procedures [14].

To get along with this, new statistical thinking and computational methods are needed. Most of the methods that we use today to analyze data are struggling with either the massive size of the data or the high-dimensionality of the data. To design effective statistical procedures for exploring and predicting Big Data, we need to address Big Data problems such as heterogeneity, noise accumulation, spurious correlations, and incidental homogeneity, in addition to balancing the statistical accuracy and computational efficiency [14]. This is one of the future trends that is given a high research effort nowadays (see chapter 1.5.2 on the following page).

1.5 Future Trends

The previous part showed the past milestones of data analysis, current research topics, and challenges that researchers are facing right now. The following part will cover certain topics that may have an impact in the future development of AI and are therefore, important to be investigated more closely.

1.5.1 Security and Privacy

Currently, the amount of stored personal data is increasing along with better analysis, extending the insights to various connections. Due to more valuable and sensitive information, there is a call for data security and privacy, which existing schemes cannot satisfy. Evidently, there is quite a contradiction between Big Data and security. Since it is a hard task to arrange these two problems, it will have a heavy impact on the development of Big Data. The danger of Big Data is not only the data, which is to be kept private, but also the fact that data mining makes it possible to generate new insights into a person's behaviour. Therefore, the information about a single human has to be secret in every layer of Big Data (collection, mining, and decision making). The process starts with the data provider, where there is a trade-off between security and the benefits of participation in data mining. Since the aim of companies is to add value through users' data, they have to make the data provider feel safe so that they offer more data. Due to different opinions about these trade-offs and subjective notions of privacy, applications should implement an option, enabling the common user to define the information that can be uploaded. Once the data is collected, it contains sensitive data that enables re-identification of the provider [23]. Narayanan et al. [31] show that it is possible to re-identify people, especially if one has additional information. So, anonymization operations have to guarantee anonymity with no significant loss in data value. Fung et al. [18] describe that the available projections for this purpose are generalization, suppression, anatomization, permutation, and perturbation.

One approach is decentralized learning, which keeps the client's data in his possession. Federated learning is an ML setting that is training a high-quality model under the leadership of a central server from a federation of participating devices. In the first step, the clients are downloading the current model from the central server. Each client adapts and improves its model, which is sent to the central server. These updates are merged with the origin model to enhance its behaviour [25]. Federated learning benefits from its structure, which is sharing generalized but not explicit data. Privacy can be achieved by generating new safety mechanisms and by exploiting predestined topology. For institutions, it should be pretty clear that building trust is a slow process, but one can lose it very fast. Therefore, only investing in privacy will advance the development in Big Data.

1.5.2 Learning in High-Dimensional Data

For ML algorithms, the rise in dimensionality of data has had a big impact over the last decades and will play an influential role in the next years. Out of data that is created in a huge variety of formats, the extracted features are represented in a high-dimensional space. With the exponentially growing complexity, diversity, and quantity in the last and next few years, common pattern recognition techniques will reach their boundaries. Hard requirements like real time applications make it due to improve the infrastructure and methods to learn from high-dimensional data in order to hold these algorithms computationally tractable. Promising techniques to reduce the dimensionalities of data are presented in this section [19][27].

By providing lower-dimensional data, common ML techniques are computationally able to process the extracted features in high dimensions. Gao et al. cluster the process dimension reduction into three different ways [19]:

Feature Transformation The set of input features that represent the data is reduced by exploiting redundancies. Thereby, a smaller set of features is created as a combination of former input variables. A common technique is Principal Components Analysis, which reduces the dimensions by orthogonally projecting the data on a linear subspace to achieve relationship information. More recently, success has been achieved by DL networks, which include auto-encoders. Common auto-encoders are extended to generalized auto-encoders, which reduce the dimensions by iteratively investigating the relations between the features through manifold learning [19].

Feature Encoding The dimension of the data is reduced by encoding the features into compact code. In this field of research, the method of hashing has been granted success for the future. High efficiency in computation and storage is achieved by encoding the feature vectors in the high-dimensional space that represent the data into a bit-string of binary code. Distances in the Hamming space using efficient bit operations correlates the proximity of different features. These hashing methods are extended to Multiple Feature Hashing, Cross-Media Hashing or Complementary Hashing [19].

Feature Selection Only the most relevant features of the input space are kept with filter and wrapper methods. The computationally light filter methods, which are not built upon any ML algorithm, have recently been devoted a lot of research effort. Features are ranked and filtered with a filter method and afterwards, irrelevant features are further eliminated by a wrapper method. Sequential and heuristic wrapper approaches evaluate the feature subset with a sort of black box predictor by maximizing the predictors objective function [19]. Apart from feature selection without ML algorithms, a lot of research has been granted methods of online ML, in particular, batch learning when dealing with large-scale and high-dimensional data in the real world. Although batch learning reaches high prediction accuracy, there are drawbacks in terms of efficiency and scalability. Therefore, Wu et al. present promising approaches regarding online feature selection algorithms, which reach highly competitive accuracy in comparison to batch learning algorithms, but are computationally lighter [45].

Especially for real world applications, to learn in Big Data, efficiency, scalability, and computational intensity is crucial currently and certainly also in the future.

1.5.3 Learning from Imbalanced Data

Ordinary learning algorithms assume the distribution of different classes in the data set as roughly equal. Unfortunately, in real world applications, this is often not the case and the number of samples is highly unequal between classes both in binary and in multiclass decision problems[26]. Examples for such applications are software defects, natural disasters, cancer gene expressions, or fraud credit card expressions. These events are infrequent and have heavy cost, if undetected [21]. The research has been dealing with this problem for more than two decades already but there is still work to do as learning from imbalanced data is important to be able to use additional Big Data sets to improve approaches in artificial intelligence [38].

Approaches for Learning from Imbalanced Data Currently, three different approaches for learning from imbalanced data are in use and focus of research [26]:

- The first approach is the data-level method, which changes the data set by oversampling or undersampling. In practice, this means to generate more samples of the minority class or delete samples of the majority class. This can help but does not solve the general problem as generalization ability must be created also from the minority class and no features from the majority class must be lost.

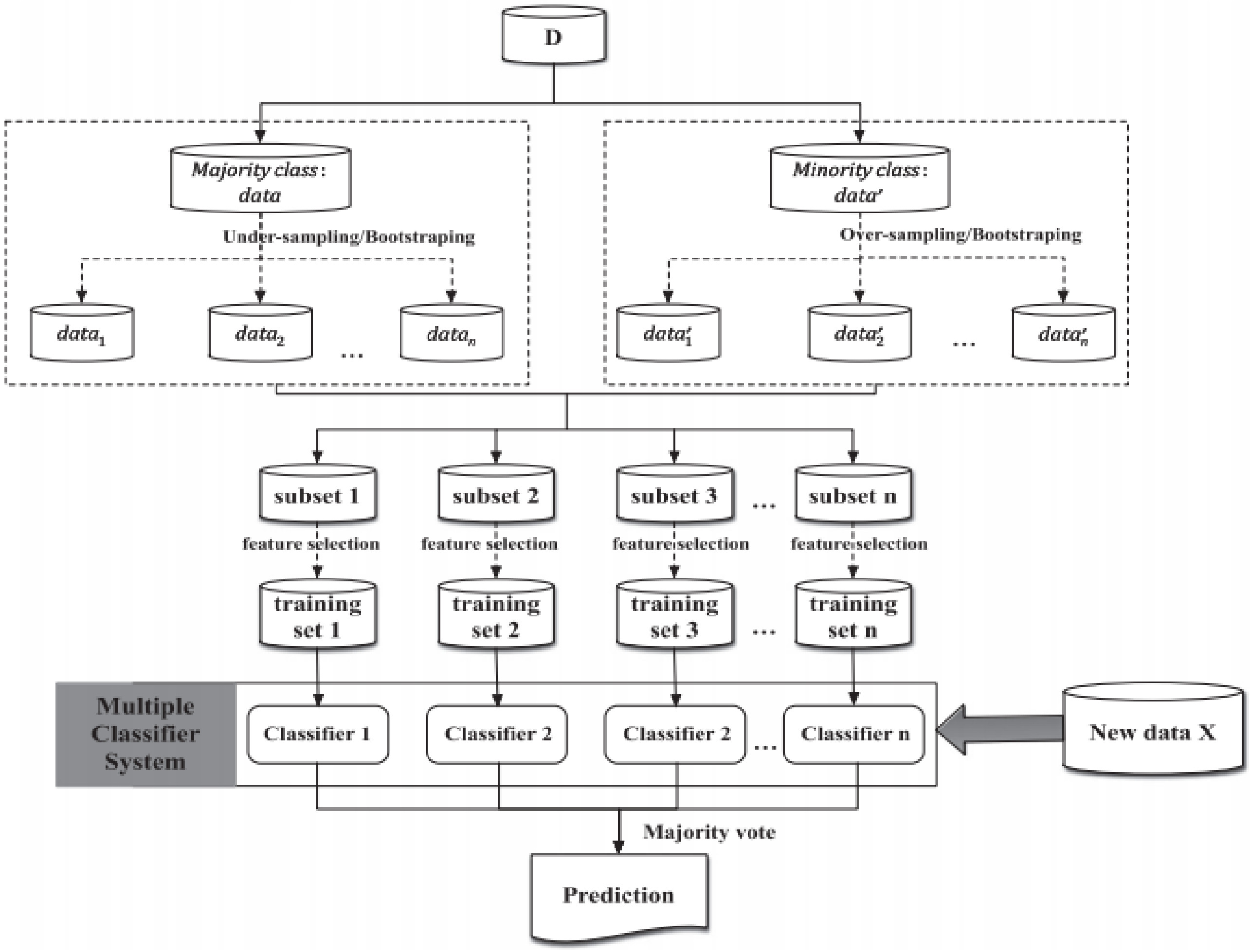


Figure 1.2: Parallel ensemble framework; *source: [21]*

- Algorithm-level methods are another approach. They move the bias towards the minority group. However, this requires a good knowledge of the learning algorithm and tends to fail while handling skewed data sets.
- The third approach is a hybrid version of the first two, trying to combine the strengths and reduce weaknesses.

While a lot of research was already done on binary classification problems with imbalanced data, multiclass data classification is much more difficult and a future topic.

General Challenges of Learning with Imbalanced Data Sets In the real world, the samples are clearly not separated from samples of another class. The signals are underlying noise and there are usually outlier samples. Just removing the outliers increases the problem as the number of samples decreases. How to deal with such important samples is still a topic for research [26]. Moreover, the minority samples often overlap with the majority class. For a better accuracy score, a usual classifier will classify those regions with the majorities label [12]. This also effects the Algorithm-Level and Hybrid Methods introduced earlier. The difficult samples need a more differentiated treatment of samples as the bias is usually adjusted according to the respective class and not to properties of a specific sample [26]. An additional challenge is the ratio between classes. While the current research concentrates on ratios from 1:4 to 1:100, real world applications often have ratios between 1:1000 to 1:5000 [9]. A very popular approach for imbalanced data is Ensemble learning, which is highly competitive and robust. The several base classifiers are combined after performing independently [21]. Nonetheless, there is a lack of insight into performance with skewed classes. It is unclear whether diversity in data sets of majority class and minority class is equally important. Moreover, there is no defined process to construct ensembles, like the relationship between number of classifiers and characteristics of imbalanced data [26]. The scheme of a parallel ensemble learning process with over- and undersampling as prepossessing is shown in figure 1.2 on the previous page.

Challenges in Multiclass Classification The mentioned challenges for binary problems are mostly also present in multi-class problems, but are even more challenging there. Furthermore, there are additional challenges that arise in multi-class problems: One approach to reduce the complexity in multi-class classification is to do a class decomposition and reduce the classification to several binary classification problems [16]. To improve this, future work may consider to work with different classifications depending on the distribution of classes as some pairs may be more equally distributed than others and must be treated differently for good results. An additional challenge for research is to fuse the binary problems to a classification for several classes again. These approaches need more insight on classification of differently skewed distributions and how ensemble solutions can be implemented here for better results [26].

Semi- and Unsupervised Learning from Imbalanced Data A popular method for unsupervised learning are clustering algorithms. They discover groups in data on their own and reduce the problem complexity. In clustering, groups with highly different size reduce the effectiveness of the algorithm [26]. There are centroid-based and density-based methods. A future chance could be to use more density based or hybrid methods as the density-based methods are more robust against imbalance while centroid-based are more popular. Additionally, clustering could be used to structure the samples of the minority class itself, tested in [42] by Wang and Chen.

Imbalanced Distribution in Big Data As in Big Data, introduced in section 1.2.2, the data complexity is higher and the number of data sources increases, the imbalance of the data also increases. Therefore, algorithms must be scalable, efficient, and able to handle heterogeneous and atypical data. Systems like MapReduce (see section 1.3.2) cannot use oversampling methods as the different mappers are differently partitioned and the creation of artificial samples is not satisfactory. A possible solution is a global arbitration unit working as a supervisor or considering the relationship between samples of one class for oversampling [26].

Conclusion on Learning with Imbalanced Data The impact of progress in learning with imbalanced data can affect the possibilities of artificial intelligence for a wider range of applications. In the current research, a study by Haixiang et al. [21] showed that in the last ten years, the most researched applications in the field of imbalance data were chemical and biomedical engineering, financial management, and information technology. Besides emergency management, they see security management as one of the domains, which could most benefit from progress in this area. Therefore, the diversity in ensembles, methods for unsupervised learning, and multi-class classification problems must be further researched. Furthermore, the field of online learning data streams should be considered as an important method to learn Big Data online.

1.5.4 The Opportunities of Machine Learning with Big Data

Owing to the new ML techniques that have been discovered in the recent years, there have been big societal impacts in a wide range of applications. For example, in the field of computer vision, speech processing, health, and Internet of Things, just to name a few. In all of these fields, the rise of the available amount of data and suitable analyses methods give us an opportunity to make considerable progress. Especially, the advent of the Big Data era has spurred interest into ML algorithms and gaining new insights into various business applications and human behaviors [47]. It is clear that Big Data provides unprecedentedly rich information for ML. However, ML is struggling by handling the big amount of data. With an ever-expanding universe of Big Data, ML has to grow and advance in order to transform Big Data into actionable intelligence [47]. These future ML techniques need to enable users to uncover underlying structures and make predictions from large data sets. If the ML algorithms develop in this direction, they will have great potential and will be an essential part of Big Data analytics [39]. In this chapter, we want to focus on the opportunities of ML on Big Data. Therefore, we would like to introduce a framework that is centered on ML and follows the phases of preprocessing, learning, and evaluation. In addition, the framework is also comprised of four other components that influence and are influenced by ML, namely Big Data, user, domain, and system [47].

The framework of ML on Big Data (MLBiD) is shown in figure 1.3 on the following page. MLBiD is centered on the ML component, which interacts with the other components. All the components interacted in both directions. In the following section, we only focus on the Big Data part and skip the rest.

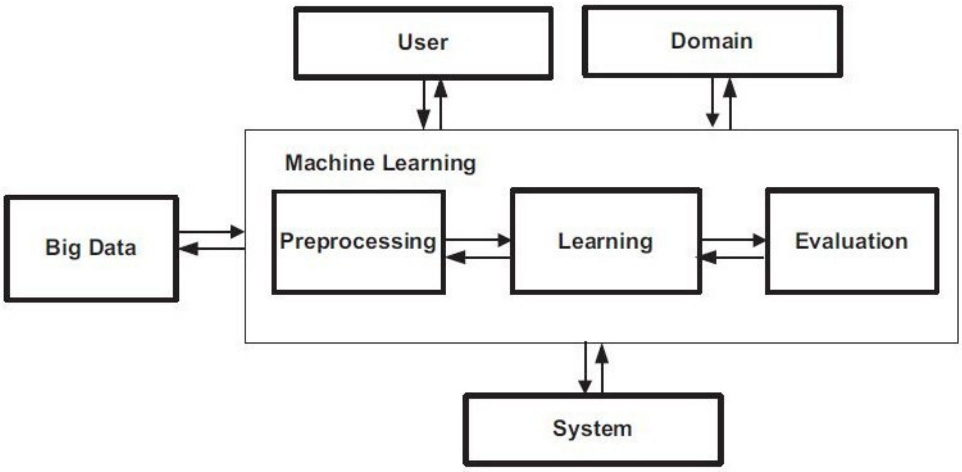


Figure 1.3: A framework of MLBiD; *source: [47]*

One of the biggest problems that Big Data and ML need to deal with is the processing of data. Currently, significant effort in deploying an ML system goes into the design of pre-processing pipelines and data transformations that result in a representation of data that can support effective ML [5]. There still exist several issues in the data pre-processing that will need attention in the future. Data preparation has always been costly because it needs human labor and a large number of options to choose from. The other problem is that some conventional data assumptions do not hold for Big Data. On the other hand, Big Data creates the opportunity for reducing the reliance on human supervision by learning from massive, diverse, and streaming data sources directly [47]. In general, you can split up the future topics into six most important challenges:

Data redundancy The first problem that needs to be addressed is the data redundancy. This means that two or more data samples represent the same entity. For this problem, the traditionally used problems in the past 20 years are no longer feasible for Big Data. To this end, a possible solution could be Dynamic Time Warping as it is much faster than the state-of-the-art Euclidean distance algorithms [34].

Noise Data The second challenge is the noise data. Missing and incorrect values, data sparsity, and outliers can introduce noise to ML. Traditional solutions have problems with handling Big Data here as well. For instance, manual methods are no longer feasible due to their lack of scalability. Nevertheless, it is important to handle the noise because interesting patterns may lie in this noisy data. Accurate predictive analytics of Big Data can be used to estimate missing values such as replacing incorrect readings due to malfunctioned sensors or broken communication channels. To achieve this, researchers are making efforts to scale up outlier detection (e.g., ONION) to enable analysts to effectively explore anomalies in large data sets [47].

Data set heterogeneity Another critical issue is the data set heterogeneity. Due to Big Data, we face multi-view data from diverse types of repositories in disparate formats and from different samples of the population. This data is highly heterogeneous. This multi-view heterogeneous data is mostly unstructured text, audio, or video format and has a varying level of importance for a learning task. Of course, all of them are not treated as equally important. Now, it is important to learn from multiple views in parallel and then ensemble multiple results by learning the importance of feature views to the task. For this, a method that is needed, which is robust to data outliers and can address optimization difficulties and convergence issues [7].

Data discretization Another problem we willl encounter in the future is data discretization. Some ML algorithms such as decision trees and Naïve Bayes can only deal with discrete attributes. Discretization translates quantitative data into qualitative data, procuring a non-overlapping division of a continuous domain. The purpose of attribute discretization is to find concise data representations as categories, which are adequate for the learning task to retain as much information as possible in the original continuous attribute [47]. However, unfortunately, most of the existing discretization approaches are not efficient to cope with Big Data. A future approach could be to firstly sort the data based on the value of a numerical attribute and then split into fragments of the original class attribute. These fragments, which are summarized by the percentage composition of different classes, are viewed as super instances and the target of discretization [46].

Labeling data Additionally, to the already mentioned challenges, there is the issue of labeling data correctly. Traditional data annotation methods are labor-intensive. Several alternative methods have been suggested to address this challenge. For instance, online crowd-generated repositories can serve as a source for free annotated training data, which can capture a large variety in terms of both class number and intra-class diversity [32].

Feature representation and selection The last major topic that affects data pre-processing in Big Data to a great extent is the feature representation and selection. The performance of ML is heavily dependent upon the choice of data representation or features [5]. Feature selection helps enhance the performance of ML by identifying prominent features. It essentially selects different subsets of features and data, aggregating them at different levels of granularity. This contributes to reducing the amount of Big Data. However, feature engineering requires prior domain knowledge, human ingenuity, and is often labor-intensive [5]. To address the weakness of current feature engineering algorithms when dealing with Big Data, various solutions have been proposed such as distributed feature selection or a low-rank matrix approximation (e.g., standard Nyström method) [47].

As we can see in the broad range of topics that are of significance, there will undoubtedly be a large amount of work in the future. Here, the challenging part will be to deal with Big Data in combination with ML. In conclusion, we can say that ML and Big Data must develop hand in hand as they are strongly connected and depend on each other.

1.5.5 Big Data and Machine Learning in Medicine

Here, we will discuss the possibility of Big Data and Machine Learning being used in the field of Clinical Medicine. Having access to a large amount of data, with the help of ML algorithms, we will be able to develop better diagnosis methods in order to achieve a higher accuracy.

Machine Learning, conversely, approaches problems as a doctor progressing through residency might - by learning rules from data. Starting with patient-level observations, algorithms sift through vast numbers of variables, looking for combinations that reliably predict outcomes [48]. ML is similar to regression models to some extent. Although, ML algorithms are better as they can handle enormous numbers of predictors - sometimes, remarkably, more predictors than observations - and combine them in nonlinear and highly interactive ways [48]. In this way, ML makes itself a better way to deal with Big Data, where the volume and complexity of data is high.

Health care, biomedical research, and population health are generating massive, complex, distributed, and often dynamic sets of data. The size and complexity of this data will pose both challenges and opportunities to health organizations [48]. Take the radiograph as an example, in a conventional way, the doctor might use certain features on the graph to deduce important results. But, with the help of ML, we can take every pixel of the radiograph as a variable and predict the outcome with a large amount of variables being taken into consideration. Usually, if we use the right model and enough input data to train the algorithm, we will get an algorithm that has an even higher accuracy than human doctors [48].

Actually, there are a significant benefits to using AI methods to perform medical diagnostics. Machines are not dependent on environmental influences that effect human alertness, performance, and concentration, e.g. distraction through tiredness. Of course, if we need a very high performance from the algorithm, there is still some work to be done. First, we need to make sure whether the quantity and quality of the input data can meet the demand. Highly 'data hungry' ML algorithms often require millions of observations to reach acceptable performance levels. In addition, biases in data collection can substantially affect both performance and generalizability [48]. Another typical problem is overfitting. Since ML algorithms can easily overfit predictions, they produce undesirable outcomes. To overcome this problem, we have to test our models on truly independent validation data sets [48].

In the future, we could reasonably expect some positive changes in clinical medicine. First, ML will dramatically improve prognosis because data could be drawn directly from EHRs (electronic health records) or claim databases, allowing models to use thousands of rich predictor variables, which produce more accurate diagnosis. Second, ML will displace much of the work of radiologists and anatomical pathologists. These physicians focus largely on interpreting digitized images, which can easily be fed directly to algorithms instead. Then, ML will improve diagnostic accuracy. However, this could take a decade to realize since there are still some technical problems to be solved. For example, for each diagnosis, the model needs to be built and validated individually [48].

Besides, Big Data can also be used to personalize medicine. Since developed and developing countries have very different abilities to generate and analyze Big Data, the gap of medical resources between the rich and poor would become even bigger in the future, especially when the ML algorithms are be used in clinical medicine on a large scale. As a result, due to lack of required technology, poor countries may lag behind when ML is involved in health care. However, we believe that the general level of medical care for everyone would still improve by a great extent in the future [2].

1.6 Conclusion

In this article, we have described the effect of data on the development of artificial intelligence, beginning with an introduction to data history and the trend of Big Data. Followed by information about the current research in data mining, the generation of structured data, cloud applications, and DL in Big Data analytics. The last chapter dealt with different future data topics that will influence artificial intelligence. Privacy and security mechanisms were presented as they will be even more challenging in future due to the significant rise in the amount of data. Since development of artificial intelligence depends on the data, risks must be minimized to preserve privacy and control data flows. Otherwise, users will try to completely shield their private data. Furthermore, techniques are shown to reduce the dimension of the space in which the features are presented to learn from high-dimensional data. As problems are exponentially growing with complexity, diversity, and quantity, common pattern recognition techniques have reached their boundaries.

Another topic of research is learning from imbalanced data as ordinary learning algorithms assume a uniform distribution between classes. For imbalanced data sets, approaches to change distribution with over- and under-sampling and change classifiers bias are presented. This topic also deals with the promising future topics of multi-class classification, unsupervised learning, and diversity in ensemble learning. In addition, the future of the connection between ML and data is explained since new ML techniques have been developed in the past years. These techniques will have big societal impacts in a wide range of applications. In addition to this development, the evolution of Big Data has been considerable over the past few years. It is clear that Big Data provides unprecedentedly rich information for ML, which makes it possible to further develop artificial learning approaches.

However, the struggle of handling Big Data to discover the hidden value is also coped with other ML algorithms in terms of extracting features, reducing dimensions, and recognizing patterns. Overall, it is important that ML and Big Data approaches are developed hand in hand in the next years in order to be able to handle these challenges. An exemplary application of the hand in hand development of approaches in ML and Big Data is clinical medicine. When there is enough high-quality data input, the ML algorithms can be trained to produce better outcomes than humans. The required high-quality input data can be extracted from the raw data of observations through ML algorithms.

The current developments in AI rely on a vast amount of data for training, validating, and testing ML algorithms. For future breakthroughs in creating more powerful and general AI methods, new approaches for data processing and analysis must be developed and improved. For these developments, it is not sufficient to just improve the current algorithms in data analysis. Rather, it requires completely new concepts. Some actual trends in research (presented in section 1.5) may give an insight into future developments.

References

[1] Debi Prasanna Acharjya and Kauser Ahmed. "A survey on big data analytics: challenges, open research issues and tools". In: *Int. J. Adv. Comput. Sci. Appl* 7.2 (2016), pp. 1–11.

[2] David Meyre Akram Alyass Michelle Turcotte. "From big data analysis to personalized medicine for all: challenges and opportunities". In: *BMC Medical Genomics* (2015), pp. 1–2.

[3] Michael Armbrust et al. "A view of cloud computing". In: *Communications of the ACM* 53.4 (2010), pp. 50–58.

[4] Julius S Bendat and Allan G Piersol. *Random Data: Analysis and Measurement Procedures*. John Wiley and Sons, 2010.

[5] Y. Bengio, A. Courville, and P. Vincent. "Representation Learning: A Review and New Perspectives". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35.8 (2013), pp. 1798–1828.

[6] Johan Bollen, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market". In: *Journal of computational science* 2.1 (2011), pp. 1–8.

[7] Xiao Cai, Feiping Nie, and Heng Huang. "Multi-view K-means Clustering on Big Data". In: *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. 2013, pp. 2598–2604.

[8] Christopher Clifton. "Encyclopædia britannica: definition of data mining". In: *Retrieved on* 9.12 (2010), p. 2010.

[9] Wojciech Marian Czarnecki and Krzysztof Rataj. "Compounds activity prediction in large imbalanced datasets with substructural relations fingerprint and EEM". In: *Trustcom/BigDataSE/ISPA*. Vol. 2. IEEE. 2015, pp. 192–192.

[10] Muthu Dayalan. "MapReduce: Simplified Data Processing on Large Cluster". In: *International Journal of Research and Engineering* 5.5 (2018), pp. 399–403.

[11] Jeffrey Dean and Sanjay Ghemawat. "MapReduce: a flexible data processing tool". In: *Communications of the ACM* 53.1 (2010), pp. 72–77.

[12] Misha Denil and Thomas Trappenberg. "Overlap versus imbalance". In: *Canadian Conference on Artificial Intelligence*. Springer. 2010, pp. 220–231.

[13] Vasant Dhar, Matthias Jarke, and Jürgen Laartz. "Big Data". In: *Business & Information Systems Engineering* 6.5 (Oct. 2014), pp. 257–259.

[14] Jianqing Fan, Fang Han, and Han Liu. "Challenges of Big Data analysis". In: *National Science Review* 1.2 (2014), pp. 293–314.

[15] Jianqing Fan and Jinchi Lv. "Sure independence screening for ultrahigh dimensional feature space". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.5 (2008), pp. 849–911.

[16] Alberto FernáNdez et al. "Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches". In: *Knowledge-based systems* 42 (2013), pp. 97–110.

[17] Damien François. "High-dimensional Data Analysis: From Optimal Metrics to Feature Selection". PhD thesis. Université catholique de Louvain, 2008.

[18] Benjamin C M Fung et al. "Privacy-preserving data publishing: A survey of recent developments". In: *ACM Computing Surveys* 42.4 (2010).

[19] Lianli Gao et al. "Learning in high-dimensional multimedia data: the state of the art". In: *Multimedia Systems* 23.3 (2017), pp. 303–313.

[20] Yunhong Gu and Robert L Grossman. "Sector and Sphere: the design and implementation of a high-performance data cloud". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 367.1897 (2009), pp. 2429–2445.

[21] Guo Haixiang et al. "Learning from class-imbalanced data: Review of methods and applications". In: *Expert Systems with Applications* 73 (2017), pp. 220–239.

[22] Han Hu et al. "Toward scalable systems for big data analytics: A technology tutorial". In: *IEEE access* 2 (2014), pp. 652–687.

[23] M. Jensen. "Challenges of Privacy Protection in Big Data Analytics". In: *IEEE International Congress on Big Data*. 2013, pp. 235–238.

[24] Xiaolong Jin et al. "Significance and challenges of big data research". In: *Big Data Research* 2.2 (2015), pp. 59–64.

[25] Jakub Konečný et al. "Federated Learning: Strategies for Improving Communication Efficiency". In: *NIPS Workshop on Private Multi-Party Machine Learning.*

[26] Bartosz Krawczyk. "Learning from imbalanced data: open challenges and future directions". In: *Progress in Artificial Intelligence* 5.4 (2016), pp. 221–232.

[27] Zhihan Lv et al. "Next-generation big data analytics: State of the art, challenges, and future research topics". In: *IEEE Transactions on Industrial Informatics* 13.4 (2017), pp. 1891–1899.

[28] Natalia Miloslavskaya and Aida Makhmudova. "Survey of Big Data Information Security". In: *IEEE International Conference on Future Internet of Things and Cloud Workshops.* IEEE. 2016, pp. 133–138.

[29] Michael Minelli, Michele Chambers, and Ambiga Dhiraj. *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses.* John Wiley & Sons, 2012.

[30] Maryam M Najafabadi et al. "Deep learning applications and challenges in big data analytics". In: *Journal of Big Data* 2.1 (2015), pp. 1–21.

[31] A. Narayanan and V. Shmatikov. "Robust De-anonymization of Large Sparse Datasets". In: *IEEE Symposium on Security and Privacy (sp 2008).* 2008, pp. 111–125.

[32] Long-Van Nguyen-Dinh et al. "Combining Crowd-generated Media and Personal Data: Semi-supervised Learning for Context Recognition". In: *Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia.* 2013, pp. 35–38.

[33] D. E. O'Leary. "Artificial Intelligence and Big Data". In: *Intelligent Systems, IEEE* 28.2 (2013), pp. 96–99.

[34] Thanawin Rakthanmanon et al. "Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping". In: *ACM Transactions on Knowledge Discovery from Data* 7.3 (2013), p. 10.

[35] Jürgen Schmidhuber. "Deep learning in neural networks: An overview". In: *Neural Networks* 61 (2015), pp. 85–117.

[36] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT press, 2002.

[37] C. Thanos, S. Manegold, and M. Kersten. "Big Data - Introduction to the special theme". In: *ERCIM News* 89 (2012), pp. 10–11.

[38] Isaac Triguero et al. "ROSEFW-RF: the winner algorithm for the ECBDL'14 big data competition: an extremely imbalanced big data bioinformatics problem". In: *Knowledge-Based Systems* 87 (2015), pp. 69–79.

[39] Chun-Wei Tsai et al. "Big data analytics: a survey". In: *Journal of Big Data* 2.1 (2015).

[40] Zeynep Tufekci. "Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls." In: *ICWSM* 14 (2014), pp. 505–514.

[41] Amit Verma, Iqbaldeep Kaur, and Inderjeet Singh. "Comparative analysis of data mining tools and techniques for information retrieval". In: *Indian Journal of Science and Technology* 9.11 (2016).

[42] Yangtao Wang and Lihui Chen. "Multi-exemplar based clustering for imbalanced data". In: *13th International Conference on Control Automation Robotics & Vision.* IEEE. 2014, pp. 1068–1073.

[43] Paul J Werbos. "Backpropagation through time: what it does and how to do it". In: *Proceedings of the IEEE* 78.10 (1990), pp. 1550–1560.

[44] Xindong Wu et al. "Data mining with big data". In: *IEEE transactions on knowledge and data engineering* 26.1 (2014), pp. 97–107.

[45] Yue Wu et al. "Large-scale online feature selection for ultra-high dimensional sparse data". In: *ACM Transactions on Knowledge Discovery from Data* 11.4 (2017).

[46] Y. Zhang and Y. Cheung. "Discretizing Numerical Attributes in Decision Tree for Big Data Analysis". In: *IEEE International Conference on Data Mining Workshop.* 2014, pp. 1150–1157.

[47] Lina Zhou et al. "Machine learning on big data: Opportunities and challenges". In: *Neurocomputing* 237 (2017), pp. 350–361.

[48] M.Phil. Ziad Obermeyer M.D. and Ph.D. Ezekiel J. Emanuel M.D. "Predicting the Future — Big Data, Machine Learning, and Clinical Medicine". In: *HHS Author Manuscripts* (2016), pp. 1–3.

Chapter 2

Avante-garde Computational Power

Junwen Li
Tobias Mayr
Ziqi Zhang
Johannes von Stetten
Fabian Uhl

2.1 Abstract

Declared in 1965, Moore's Law describes a doubling in the number of components per integrated circuit and has been proven accurate over several decades. However, in recent years, the pace of progress has slowed down, with some applications already getting pushed to their physical limits [12]. Therefore, new ways to optimize computational processes are needed to satisfy the increasing demand for faster and more efficient electronic devices. Recent research promises a wide variety of new technologies like neuromorphic, quantum, and biological computing as well as advances in the field of heterogeneous and cloud computing. This chapter aims to give an overview of today's standings of these CPU studies, its achievements and current challenges with a special regard to artificial intelligence. Lastly, an outlook to future research directions is provided.

2.2 Introduction

"Computation is a bottleneck right now for machine learning" said Reza Zadeh, an adjunct professor at Stanford University as well as founder and CEO of Matroid[1]. Feeding big data into deep learning software to train it is exponentially more resource intensive than before. The computing industry has relied on Mooores's law and Dennard scaling for 50 years to achieve more powerful computation. According to Moore's law, the number of the transistors in a dense integrated circuit doubles about every two years [33]. Meanwhile, Dennard scaling describes how the amount of power that transistors use scales down as they shrink [13]. Whereas, currently, these two phenomena are difficult to continue because the transistor cannot get smaller. Therefore, we need new solutions to satisfy our increasing demand for computational power.

Current limitations of computing power are also due to shortcomings of the von Neumann architecture [35]. Many of today's computers at least partially implement the von Neumann architecture. As shown in figure 2.1 on the next page, the von Neumann architecture consists of a central processing unit which communicates with a memory unit over a single sequential bus. In a von Neumann architecture, data and program instructions are both stored in the same memory unit and share the single bus. The limited data transfer rate of the shared bus causes the 'Von Neumann Bottleneck', which limits the processing speed [3]. This problem has become more prevalent as CPU speed and memory size have increased, whereas the data transfer rate could not increase proportionally.

[1] `https://www.technologyreview.com/s/607917/how-ai-can-keep-accelerating-after-moores-law/` (accessed April 26, 2019)

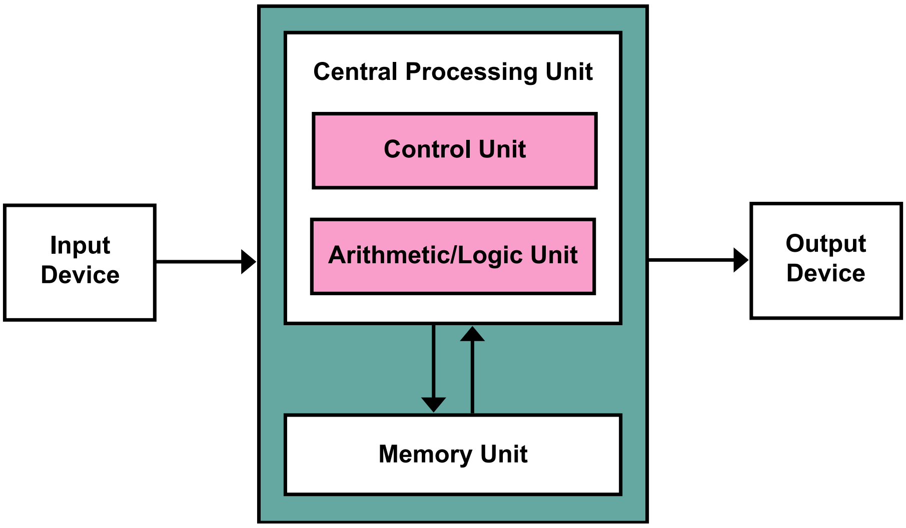


Figure 2.1: Von Neumann architecture diagram; *source:* [2]

2.3 Neuromorphic Computing

The term 'Neuromorphic' was introduced by Carver Mead [30]. In his work, he used Very Large-Scale Integration (VLSI) systems to create analog circuits which modeled neuronal structures. Nowadays, neuromorphic computing describes more generally, computing systems which use biologically-inspired non-von Neumann architectures. Since the introduction by Carver Mead, the field underwent a massive increase in publications. The review article [42] gives a comprehensive overview of the development in the field.

A particular example for a neuromorphic computing system is TrueNorth by IBM [31]. The TrueNorth digital chip uses an architecture which is strongly inspired by neurobiological concepts. The inspirations are used on different levels: TrueNorth consists of a multichip network which resembles the long range connections between cortical regions. Each chip is made from a network of neurosynaptic cores inspired by the cortex's two-dimensional sheet. The overall network is designed to connect a neuron on any core to an axon on any core just like in the brain. Each of the neurosynaptic cores is loosely inspired by the idea of a canonical cortical microcircuit. Therefore, each core has a structure in which axons exist as inputs, neurons as outputs, and synapses as directed connections from axons to neurons. In addition, the architecture models aspects of neuronal behaviour like spiking and axonal delay. The whole architecture is implemented using 28-nm CMOS technology.

Furthermore, neuromorphic hardware supported the development of spiking neural networks (SNN), which are often called 'the third generation of neural networks' [28]. These neural networks work on time series of 'spikes', which are short impulses in the signal. Therefore, SNNs can additionally use time to encode information by spiking frequency. Typically, a neuron in an SNN accumulates the incoming spikes and emits a spike itself, if the neuron's internal potential exceeds a certain threshold.

2.3.1 Design of Neuromorphic Chips

As neuromorphic hardware breaks up with the widely used von Neumann architecture for computers, different levels of design have to be taken into account. While on von Neumann computing systems, only the software has to be adapted, whereas on neuromorphic systems, there is a co-development between hardware and software. Currently, the community is focusing on developing appropriate hardware despite lacking a helpful software for easy programming [42].

Several steps have to be taken for the design of a neuromorphic system as described in [42]. First of all, the underlying neurobiological processes have to be modeled. This typically involves modelling the interaction between the building blocks like neurons and synapses, and modelling the inner workings of those building blocks. Depending on the application, one can choose between models which are more or less biologically plausible. Normally, more biologically plausible models are used for simulating biological brains for neuroscience.

[2] `https://commons.wikimedia.org/wiki/File:Von_Neumann_Architecture.svg` (accessed April 26, 2019)

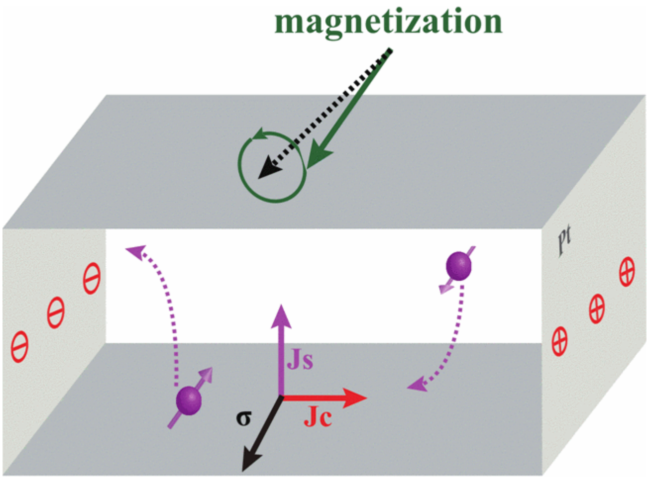


Figure 2.2: Spin affecting electrical and magnetic fields; *source: Zeng et al. [52]*

Meanwhile, more abstract models can be used to implement existing approaches such as convolutional neural networks (CNN) more effectively for technical applications.

If the model is chosen, one has to implement algorithms that can run on the specified models. A particular challenge for the implementation of algorithms is the learning part. In this context, one can choose between on-line or off-line learning. While on-line learning is typically seen as more resource-intensive, neuromorphic systems are said to simplify it, fundamentally. Looking at the current developments in the field, a major problem seems to be the lack of proper algorithms for neuromorphic hardware [42]. This is mainly based on the fact that all existing algorithms have been developed for von Neumann architectures and worsened because there is no standard neuromorphic hardware architecture until now. Therefore, for the development of a neuromorphic algorithm, several factors have to be taken into account: the model, device type, type of learning and the desired performance. The chosen models and algorithms then have to be implemented in hardware. Unlike the beginning of neuromorphic hardware, many current implementations are not analog anymore but digital or mixed digital-analog systems [42]. In addition to the high level system design, often, novel or less common devices like memristors or spin-based devices are incorporated because of their very special characteristics. Furthermore, there have been recent innovations in the choice of the materials for neuromorphic hardware like the application of graphene [22].

A new, popular field of research used for improving the performance of neuromorphic chips is Spintronics. The idea of this technology is to use electron spin rather than charge to process information, promising energy-efficiency, non-volatility, and near-zero leakage (see figure 2.2) [38].

The physical foundation for this field of research is the ability to induce a spin in the electrons in semi-conducting and organic materials as well as metals. The spinning electrons produce polarized currents inside the materials that affect both electrical and magnetic fields. Aligning the spin in magnetic multi-layered stacks causes the 'short-circuit-effect', meaning that half of the electrons with the same spin direction create a short-circuit in all magnetic layers. Another point is that the internal electrical resistance of the material is drastically reduced by the magnetic field. This effect is called giant magneto-resistance (GMR) [16]. Research papers of Chen, Sharad et al. proved that devices based on spintronic technology are able to mimic the computational behaviour of synapses and neurons while working more energy- and space-saving than conventional CMOS implementations [9][43].

2.3.2 Application of Neuromorphic Chips

There are two main applications of neuromorphic chips: neuroscience and technical applications [42]. In neuroscience, neuromorphic chips can be used to speed up simulations of processes in the brain and therefore, help obtain new insights. Till date, many technical applications for neuromorphic hardware have been tested. In most of the applications, neuromorphic chips are used to implement some kind of neural network. On the other hand, there have been also approaches where general graph problems have been solved [7]. Typical fields for technical applications are imitating biological sensing [11], interfacing to biological systems [19], robotics & con-

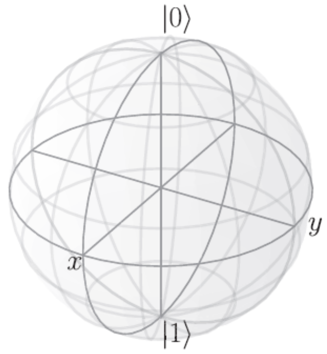


Figure 2.3: Bloch sphere for qubit in quantum computing; *source: [51]*

trol [27], image & video processing [46], language processing, smart sensors, and general data classification [48].

Very commonly, neuromorphic implementations are used for image-based tasks like edge detection, image compression and image classification. The already mentioned platform TrueNorth was also used for multi-object detection and classification in videos. It had to recognize and classify people, cyclists, cars etc. in a surveillance camera footage. While operating on a resolution of 400x240 pixels and a framerate of 30 frames per second, it only used 63mW [31]. In particular, many implementations work with standard data sets like the MNIST database [24], which is a collection of handwritten digits to recognize. This is done to make neuromorphic approaches comparable to older approaches as these have been exhaustively tested on the standard data sets. Another way to compare neuromorphic chips to other implementations is by applying them on basic benchmarking tests like realizing simple logic gates e.g. AND or XOR [50].

Another big area of application is imitating biological sensing systems like vision or hearing. For this application, non-linear components can be used to directly recreate a biological sensing system. The resulting neuromorphic chips can be used to apply the sensing systems for technical purposes or to better understand the real systems [11].

2.4 Quantum Computing

Quantum computing is computing using quantum-mechanical phenomena, such as superposition and entanglement. A quantum system makes a surprisingly efficient computer; quantum algorithms may provide quadratic or exponential speedup over the more well-known classical counterparts.

A qubit is the fundamental building block of quantum computing and is a two-level quantum state. The general pure state of a qubit on this basis is with the constraint $|\alpha|^2 + |\beta|^2 = 1$. With the constraints on α and β, these two numbers define a point on the surface of the unit sphere in three dimensions and this sphere is called the Bloch sphere (see figure 2.3). Its purpose is to give a geometric explanation to single-qubit operations [51].

2.4.1 Quantum Machine Learning

In quantum machine learning, algorithms are developed to solve typical problems of machine learning using efficiency of quantum computing. The expectation is that in the near future, such machines will be commonly available for applications and will be able to help process the growing amount of global information. Making use of the efficiency of quantum computing makes it possible to speed up the computation of AI, exponentially.

The most important application of quantum machine learning methods to solve actual problems is probably the task of pattern classification. For example, scientists intend to use feature vectors that contain pre-processed information on patients and their correctly diagnosed disease to create algorithms that can efficiently classify the disease of the patient. Closely related to pattern classification are other tasks such as pattern completion, pattern recognition, and associative memory. Associative memory is retrieving a memory vectors from a number of stored memory vectors upon an input.

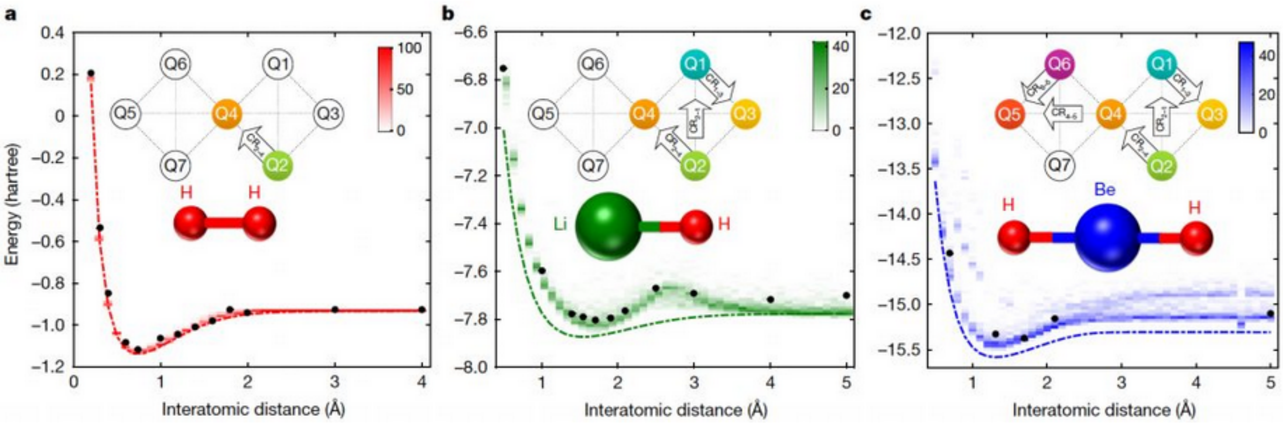


Figure 2.4: Application of quantum chemistry; *source:* [3]

A support vector machine (SVM) is a machine learning algorithm used for linear discrimination, which is a subcategory of pattern classification. To find the hyperplane that is the best discrimination between two class regions and serves as a decision boundary, we need to calculate the mathematical formulation that contains a critical inner product - kernel K. K is a Matrix which creates new feature vectors from original feature vectors and calculating it can get really expensive in terms of computational resources. Thus, it is crucial for SVM to evaluate this Matrix efficiently. Quantum computing is a possible solution and calculation can be processed parallely [29].

Neural network models are also a widely used technology for artificial intelligence. A challenge for pattern classification with a neural network is the computational cost for the backpropagation algorithm. Backpropagation computes the gradients that are needed in calculation of weights. For this problem, quantum searches are proven to be faster than comparable classical searches [29].

2.4.2 Application of Quantum Computing

An application of quantum computing is to measure molecular energy from IBM. In 2017, IBM's scientists developed a new approach to simulate molecules on a quantum computer that may revolutionize chemistry and material science[3]. The scientists use six qubits on a quantum computer to address molecular structure problems - the largest molecule simulated on a quantum computer to date. The results demonstrate a path of exploration for near-term quantum systems to enhance our understanding of complex chemical reactions that could lead to practical applications. Figure 2.4 describes an application in quantum chemistry: there are experimental results (black filled circles), exact energy surfaces (dotted lines), and density plots (shading) of outcomes from numerical simulations, for several inter-atomic distances for H2 (left), LiH (middle), and BeH2 (right)[3].

IBM Q is an industry-first initiative to build commercially available universal quantum computers for business and science. The IBM scientists put a 20-qubit machine in the cloud and made it available for the world to use, explore, and learn from. They are also working on a prototype with 50 qubits, which will be for the future of IBM Q. However, IBM has not published details of its system in any journal and without looking into details it is hard to comment[4].

D-Wave Systems Inc. is the world's first company to sell quantum computers. It declared that it has a 2,000 qubit quantum computer. But, this does not mean that D-Wave Systems has access to a quantum computer 400 times bigger than IBM. The underlying reason behind is probably, the different approaches to quantum mechanics in the two companies[5].

[3] `https://www.ibm.com/blogs/research/2017/09/quantum-molecule/` (accessed April 26, 2019)
[4] `https://www-03.ibm.com/press/us/en/pressrelease/51740.wss` (accessed April 26, 2019)
[5] `https://www.theregister.co.uk/2017/03/06/ibm_has_cloud_access_to_quantum_computer_400_times_smaller_than_dwave_system/` (accessed April 26, 2019)

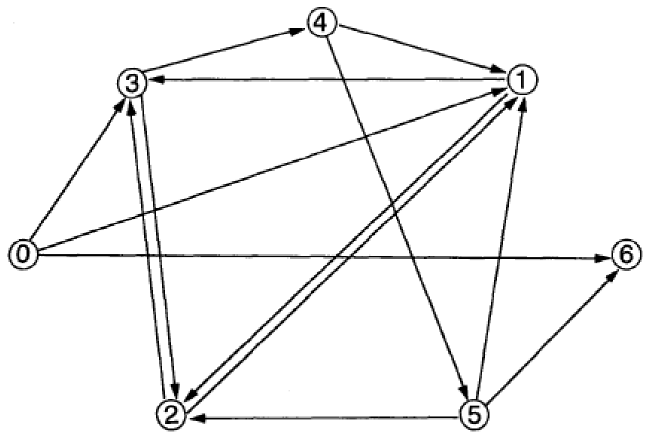


Figure 2.5: Hamiltonian Path Problem; *source: [25]*

2.5 Biological Computing

Currently, biological computing is a promising research area. Unlike silicon-based devices, computers made from biological materials are bio-compatible and have several characteristics such as self-organization, massive parallelization, and functional complexity [34]. In biological computing, chemical reactions in the biological material are used to finish certain computing tasks. Biological computers provide a platform for performing Biological Computing. Scientists take advantage of biological material like DNA, RNA, and protein as the main raw materials in biological computers. Information is propagated in biological computers in the form of waves. When waves go through the protein molecules, the structure of the protein molecules will be changed. In this way, information can be either stored or propagated.

Biological computers have small volume but very high power. More than hundred million circuits can be integrated on a small area of biological computers. Furthermore, because biological materials can process chemical reactions with only little energy, the thermal loss is low compared to the conventional electronic computers. A fascinating advantage of biological computers is that biological material can repair itself, if it does not work well. This makes biological computers more reliable and robust. In addition, biological computers are highly capable of storing data. One gram of DNA can store as much data as one billion CDs [8]. In the aspect of synchronized processing, biological computers also perform very well.

2.5.1 Past Milestones

The original idea to design a biological computer came from Professor Feynman in 1959. He mentioned in one of his lectures that people should design and construct a computer which can manipulate DNA and use it to perform calculations.

After the 1970s, people found out that some important components in a digital circuit, for example, diodes and logical gates like AND-Gate and OR-Gate can be implemented by using biological material. Furthermore, DNA can carry different information under different circumstances. Thus, biological chips can be produced after careful cultivation and are able to deal with problems at an extremely high speed.

One important milestone in the area of Biological Computing is the achievement of Professor Adleman in 1994 [1]. He used biological computing to solve the famous Hamiltonian path problem. In the Hamiltonian path problem, there are several nodes and the aim is to find a path which connects all of the nodes and all nodes should be entered only once (see figure 2.5).

To solve this problem, Professor Adleman randomly generalized many DNA sequences and each sequence represented one node. According to the sequences of nodes, the sequences of all paths can be derived. A path-sequence consists of the second half sequence of the origin and the first half sequence of the terminal. Then a complement of each sequence of the node can be derived according to the principle of complementary base pairing, namely A and T is a pair, and C and G is a pair. Then, these complements can play a role as a splint which connects two paths. For example, the complement of node 3 can connect path 2-3 and 3-4. In this way, a longer

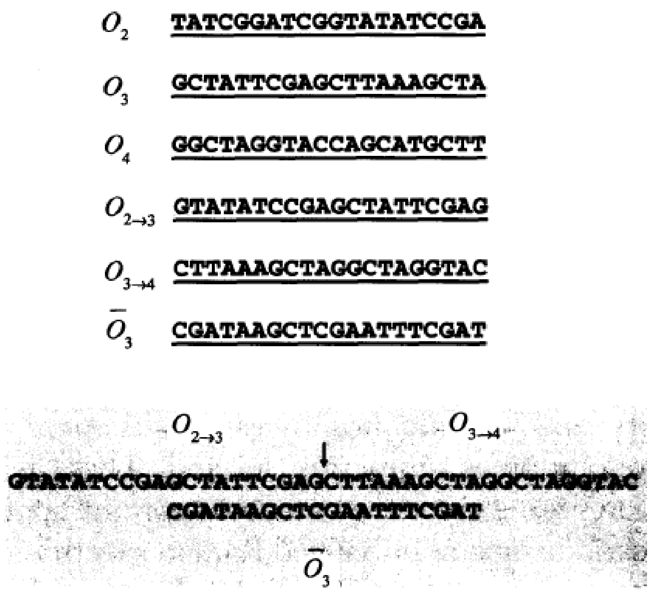


Figure 2.6: Encoding; *source: [25]*

DNA sequence can be obtained and extended. Figure 2.6 gives an example to the process of solving this problem.

Although Professor Adleman only solved a Hamiltonian path problem containing only seven nodes, it is the first successful example to use biological computing to solve practical problems. The influence of this success is self-evident. It proves that the DNA can be controlled and its internal biochemistry can be thought of as computing processes. The success of Professor Adleman led to a wide research in the area of biological programming and some recent papers and news indicate that biological programming is a research area with high potential on accelerating computing processes and it is possible to execute computations by means of DNA or cellular components.

2.5.2 Current Research

In 2016, the MIT developed a programming language for bacteria, enabling users to write their own algorithms and functions such as detecting the temperature of the environment[6]. After compiling the functions, a particular DNA sequence is generated. That is, the algorithms are carried by the DNA sequence. Then, the DNA sequence is put into a cell and the functions run inside the cell. The DNA sequence with important information will chemically react with other DNA sequences in the cell. In this way, the algorithms can be 'biologically' implemented.

In 2010, another paper introduced minimal models of molecular communication [34]. These models use cells to implement the function of diodes, AND-Operation, OR-Operation and NOT-Operation. Figure 2.7 on the next page shows how an AND-Operation can be achieved by using such a cell model. The H means that the cell is highly excitable and if one of its neighbour cells is excited, it will be also excited after a time delay t. The L means that this cell cannot be excited until two arbitrary cells in its neighbourhood are excited. So, if A or B is excited, C will not be excited. But, if both A and B are excited, then C will be excited after t and the output Y will also be excited after 2t. Other models will realize different operations in digital circuits.

A review paper in 2010 [40] indicates that the hardware accelerators can make biological computing faster. With the help of hardware, the algorithms in biological computing can be strongly accelerated. Central processing unit (CPU), graphics processing unit (GPU), field-programmable gate array (FPGA), and cell broadband engine (CBE) can all play a role as an accelerator in biological computing. Some data in this paper shows that FPGAs can execute pairwise sequence alignment, a very significant algorithm in biological computing, a hundred times faster than without the help of hardware. The Multi-core Network-on-Chip improves the computing speed even more surprisingly. The algorithm runs 22.000 times faster than before. Other algorithms in biological

[6] `http://news.mit.edu/2016/programming-language-living-cells-bacteria-0331` (accessed April 26, 2019)

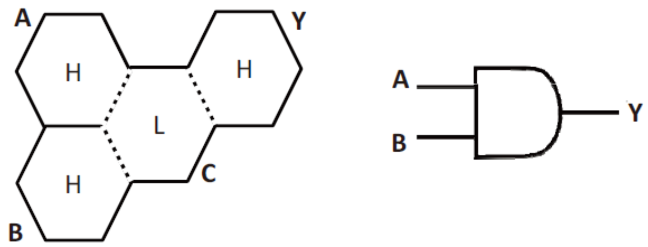


Figure 2.7: Implementing AND-Operation by cellular model; *source: [34]*

computing such as the phylogenetic tree reconstruction strategies can be accelerated with the help of hardware as well.

2.5.3 Current Challenges

The biggest problem that exists in biological computing is that it is difficult and sometimes intractable to extract the useful information. For example, in the experiment of Professor Adleman, the DNA finished the calculation in an hour, but it took Professor Adleman seven days to select the right answer for the problem because the right answer was mixed with many false sequences. This big problem of how to extract the useful and required information effectively from a mixture of useful and useless information, still remains.

In addition, to execute the computing biologically, a large number of DNA sequences are necessary. This increases the cost of solving problems through biological computing.

2.5.4 Conclusion

In conclusion, biological computing is a promising research area and it has already made sizable achievements. Some successful examples of biological computing show that it is feasible to compute biologically. Since chemical reactions are executed very quickly in biological material, it possesses a high calculation speed. Furthermore, biological materials possesses a fascinating potential in storing data and repairing themselves when damaged. Yet, the problems of applying biological computing cannot be ignored. The biggest obstacle of applying biological computing on a large scale is the difficulty separating useful computing results from useless ones. A further barrier is the cost of executing biological computing because a lot of DNA sequences must be generated before performing the computations.

2.6 Heterogeneous Computing

Due to the upcoming physical limit in the integrated circuits technology, new approaches for a better computing power must be found. The multi-core technology allows the use of various processors in parallel, which permits utilizing previous technology to improve the performance in a significant way. Homogeneous computing sets the focus on CPUs, consisting of identical processor architectures, whereas a heterogeneous computing CPU is build of different special processors with various characteristics [32].

2.6.1 Special Processors

In recent years, machine learning and other new tasks have changed the requirements that must be handled by a computer and especially, by its CPU. So, the learning task in neural networks already moved from the CPU to the GPU for a better performance [41]. The special processors are produced for a different range of applications. For instance, the graphics processing unit (GPU) is developed for the several but small graphical computations, which are often vector or matrix computations. To increase computational performance, the processor uses efficiently pipeline and is able to parallely compute identical calculations. Another special processor is the digital signal processor (DSP). It is designed to handle continuous signals like an audio or video data stream and can be found in plasma TVs. Furthermore, the low power consumption of the DSPs predestines them for mobile devices like smart-phones.

2.6.2 Software

The multi-core technology enhances the method of programming a software. While a one-core CPU processes a list of tasks sequentially, a multi-core CPU can process tasks in parallel. However, the tasks that can be handled in parallel and the ones that depend on the solution of another task must be preset by the programmer. Furthermore, the distribution of the tasks gets more complex with the different instruction set of individual special processors. To handle these difficulties, a fundamental restructuring of the software architecture can be offered. In contrast to the symmetric multiprocessing (SMP) system, which relies on a one kernel model with one OS and a shared memory, the asymmetric multiprocessing (AMP) system combines more than one kernel with more than one OS and allows parallel to shared memory and a separate memory model. Thus, the modularization takes place on the lowest level of the software stack[7].

2.6.3 Benefits of Heterogeneous Computing

Three of the most important characteristics of a CPU are the performance, power consumption, and security. To improve the performance of the CPU, tasks can be split into parts with different requirements and assigned to the most efficient processor. The resulting computational power is a combination of the ultimate (best performance) computational power of the individual special processors. This performance optimization depends on the constellation of special processors used and the distribution algorithms. To lower the power consumption in low computational power phases, the heterogeneous computing architecture is be able to switch off power intensive processors and work only with a few or one economical processor. At least to improve the security of a CPU, tasks with a high safety-relevance can be separated and computed on a special processor. With the introduced AMP software system, a CPU architecture with a separate memory can be realized. This makes it possible to encapsulate security tasks[7].

2.7 Cloud Computing

Information technology can be found in almost every part of the human life, be it transportation, healthcare, telecommunications, or the Internet of Things. Cloud computing is a new, yet rapidly emerging technology based on internet computing in central computation centres in which shared resources are provided on internet to other users on demand [47]. The idea behind this concept is to add computing capacity and to increase flexibility between users without the need to invest in new infrastructure or having additional maintenance costs (see figure 2.8 and 2.9 on the following page) [10]. In past few years have observed an increase in the significance of this technology and currently, a lot of research is ongoing to improve its implementation.

Cloud infrastructure is very robust and is always available. Computing services have to be highly reliable, scalable, and capable of being flexible to its users. In contrast to grid computing, cloud computing aims at serving multiple users at the same time over a universal connection. Cloud services can be accessed by using a standardized interface over a network [39].

Usually, cloud systems can be divided into three different types of service and depicted in a pyramid model. Infrastructure as a Service (IaaS) is a model that provides basic IT-resources like computational power, storage space and capacity. The user has full control over the system in this model.

Platform as a Service (PaaS) is a model that provides programming and development tools for cloud-based applications. Usually, memory, message queuing, and MySQL services as well as security and monitoring tools are provided when the app is deployed.

Software as a Service (SaaS) implies that the cloud system provider only provides self-made applications to the user. This is in contrast to the other two models which are using third party programs and software.

Popular and successful examples for big cloud computing services are the AWS Amazon cloud service and the Microsoft Azure cloud servers.

[7] `https://blog.nxp.com/tech-in-depth/3-reasons-why-embedded-heterogeneous-systems-are-more-efficient` (accessed April 26, 2019)

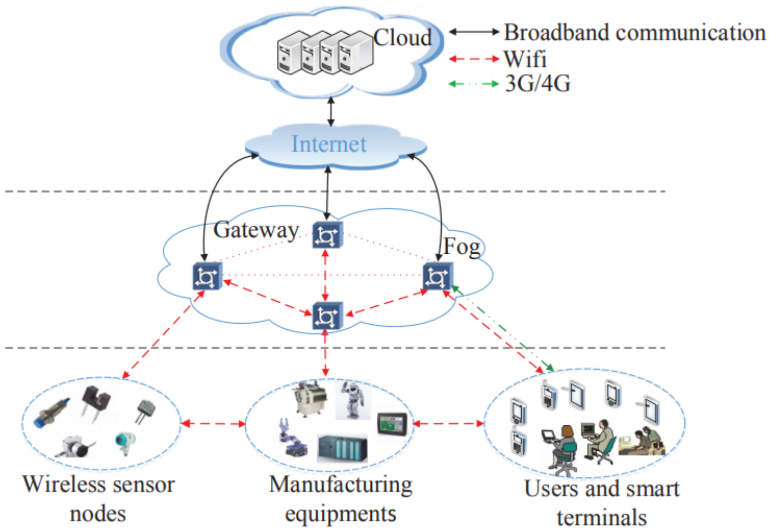


Figure 2.8: Example of IoT cloud computing; *source: [45]*

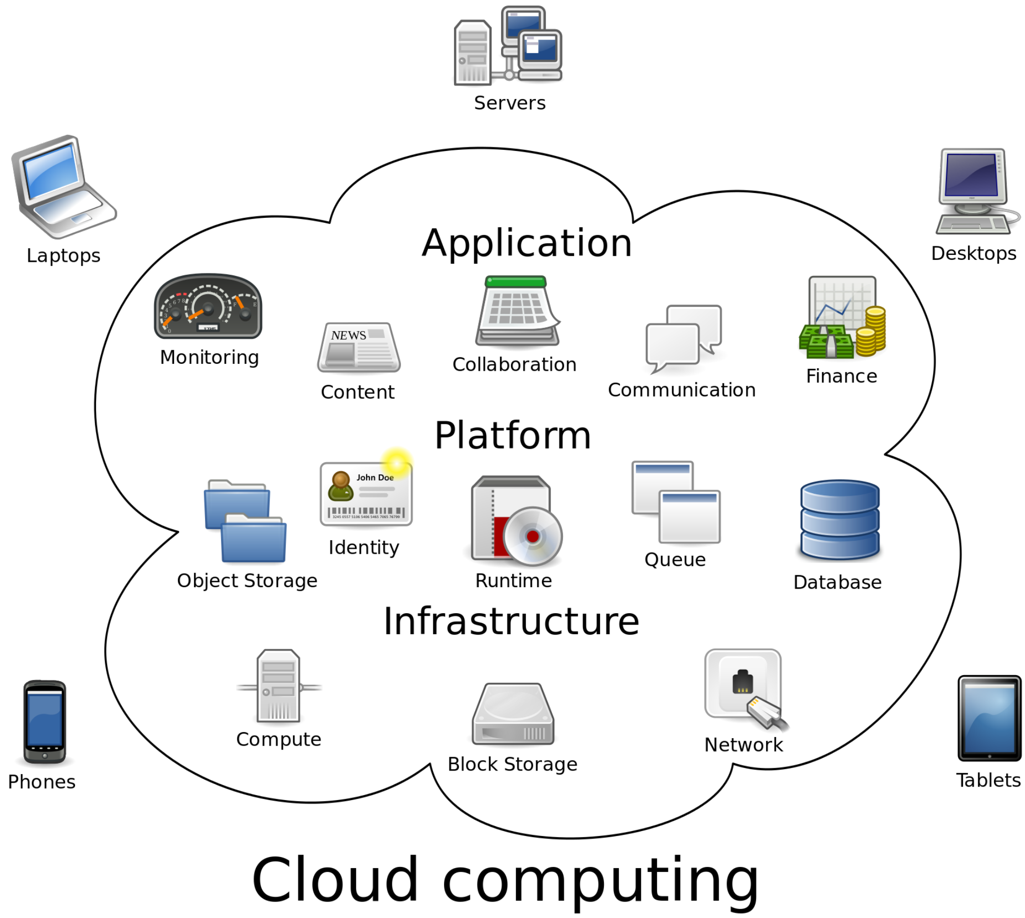


Figure 2.9: Cloud computing scheme; *source: [39]*

2.8 Future Trends

It is important to keep the past in mind while looking into the future. In the last 20 years, Moore's law seemed to last even further. However, recent studies showed that the speed of doubling transistors in an integrated circuit every two years is slowing down further [15]. In the following sections, we enlist new technologies that have already shown great potential and might help overcome the current restrictions of CPU technology.

2.8.1 Motivation

On November 17, 2008, Intel introduced the Nehalem micro-architecture, which uses transistors with a gate length for 45 nm[8]. Currently, Intel is working on 10nm micro-architecture[9]. Nevertheless, a flattening of the Moore's law can be noticed. The same flattening cannot be observed in the computational power, which shows that the transistor density is not the essential factor for the performance of a processor. To measure the computational power of a processor, the amount of floating point operations per second (Flops) are the crucial performance determination. On examining new technologies such as heterogeneous computing, quantum computing, and neuromorphic computing, we noticed that they entail advantages in fundamentally new ways of computing. Therefore, to extrapolate the computational power into the year 2040, we need a more practical approach for the performance determination. Actual benchmark rating systems determine the performance of a CPU by defining a calculation with a wide variety of requirements and measure how many calculations a CPU can handle in a given time. The resulting values make it easy to compare different computers with different hardware constellations.

According to the article "Computers Will Require More Energy Than The World Generates by 2040"[10], the energy efficiency will become a decisive criteria for the processor of the future. To tackle the proportionality of the computational power and the energy consumption, both attributes must be observed simultaneously. An easy way to reach that is to combine both into one factor. Koomey's law presented by Jonathan G. Koomey is based on the computations per kWh [21]. Here, a computation is defined by J.D. Nordhaus and is similar to the base idea of current benchmark rating systems: a time measurement needed by a CPU to handle a defined bundle of calculations [36]. In contrast to Moore's law, Koomey's law can be extrapolated into 2040, including the upcoming new technologies. In figure 2.10 on the next page Jonathan G. Koomey presents the computations per kWh of different computers and supercomputers in past 70 years. On the basis of his data, Koomey derives that "The computations per kilowatt-hour are doubled every 1.5 years" [21]. Based on this approach, the computations per kWh will be increased by a factor of 16384 in the next 21 years.

2.8.2 Quantum Computing

Still widely used, classical computers encode and manipulate information as strings with a unit of bit from information theory from Claude Shannon in 1948. A quantum computer has a great advantage because the superposition principle largely increases the states that it can encode. After a long time, with no promise of success, quantum computing is suddenly buzzing with significant achievements[11].

For the current state, only increasing the number of qubits is not enough to build quantum computers. The machine will also have to keep a low error rate. If the error rate is high, quantum computing can never perform better than a classical computer even if more qubits added. To deal with the noise, researchers work on error correcting strategies. The problem for these strategies is that most of the computing power goes to correcting errors. Thus, the length of computations that can be performed are limited. An alternative to these error correcting strategies is to avoid them or cancel out their influence.

An application of quantum computing that has leaped out of labour is quantum communication. The quantum information is carried by particles such as photons that can not be replicated exactly. Therefore, it is impossible for a hacker to listen if the data is transmitted as quantum information. The first long-distance quantum communication landline connects Peking and Shanghai over a distance of more than 1,200 miles. It is reported that several major Chinese banks are already using the link to transfer their most sensitive data[11]. China is not alone

[8] `https://www.intel.com/pressroom/archive/releases/2008/20081117comp_sm.htm#story` (accessed April 26, 2019)

[9] `https://ark.intel.com/products/136863/Intel-Core-i3-8121U-Processor-4M-Cache-up-to-3_20-GHz` (accessed April 26, 2019)

[10] `https://www.sciencealert.com/computers-will-require-more-energy-than-the-world-generates-by-2040` (accessed April 26, 2019)

[11] `https://www.insidescience.org/news/china-leader-quantum-communications` (accessed April 26, 2019)

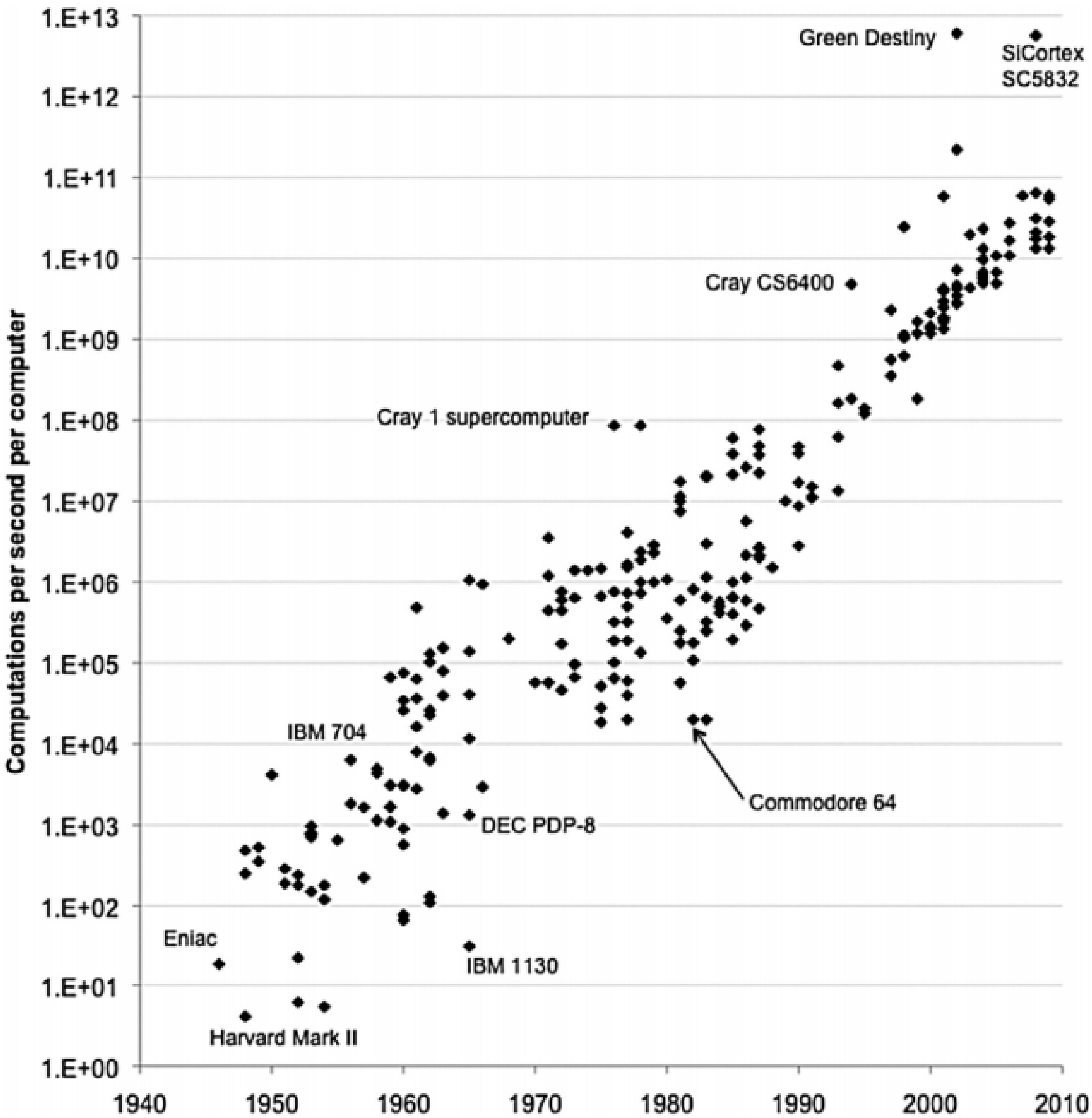


Figure 2.10: Trend of Koomey's law from 1940 to 2010; *source: [21]*

in quantum communication efforts. A team led by Harvard University had started installing a 650-kilometer line between Battelle's headquarters and its offices in Washington DC. The US and Chinese network both use the dark fibre which is not always available and can be prohibitively expensive. A research team at Toshiba Research Europa in Cambridge, UK is trying to solve this problem by piggybacking the photon streams onto the lit fibres that transform conventional telecommunication data [37].

Another application for quantum computing is quantum search. Although, currently, there has only been some success on the experimental side with a small number bits achieved. Being a trend for the future, which promises to significantly to reduce the computation time for the database search. It is also shown that it is possible for a quantum search algorithm to compute the factorization in polynomial-time [18].

2.8.3 Reversible Computing

Today's computer architecture relies on constantly overwriting and erasing the information of the temporary memory. Consequently, the energy expended to obtain the previous output gets wasted and dissipates into heat. Currently, roughly 5000 electron volts are wasted for the operation of erasing a single bit in a computer [4]. It is possible to lower the amount of energy dissipating into heat by the formula known as Landauer's principle:

$$W = k \cdot T \cdot \ln(2)$$

where k is the Boltzmann constant, T is the temperature of the heat sink and $ln(2)$ is the natural logarithm of 2.

Considering predictions about the future of rising energy consumption of computing machines as given by Dockrill[10], it may be necessary to circumvent this limit with a new approach. Reversible Computing is a subject that has been on the table for long and is receiving more attention in recent years. First brought up in a paper by Rolf Landauer in 1961 and proven to be applicable in 1973 by Charles Bennett, the implementation of the principle is still far from being well understood and being part of our daily life. However, when looking at the future development of CPUs, Reversible Computing offers big promises with regards to efficiency and power consumption; theoretically, being able to reduce the energy wasted by computers to zero [36]. The basis of this technology can be understood by observing that the most fundamental laws of physics are reversible, meaning that if one had "complete knowledge about the state of a closed system at some time, one could always run the laws of physics in reverse and determine the system's exact state at any previous time" [23]. An example to visualize this is a game of billiards without friction involved. If one hits the balls, they would bounce off each other and the bumpers endlessly. The collision physics would always remain the same and one could predict the past as well as future constellations of the balls [17].

This reversibility is also applicable to quantum-scale physics. Here, the constellations are the states of physical systems. Two different states of a system cannot become the exact same later on, meaning it is impossible to destroy information [5]. The reversibility of physics means that whenever we overwrite a bit of information with a new value, the previous information might not be practically usable anymore but is physically still there in the form of heat in the machine's thermal environment [17]. There are some projects that already show a lot of promise, e.g. researchers at Yokohama University working on superconducting electronics which dissipate less energy than stated by the Landauer principle [17].

Recently various efforts have been made to combine quantum and reversible computing, creating a new way of efficient computing. Refining this approach would result in faster, safer, and very energy efficient ways of computing without the loss of information. This is definitely something to look out for in the future [2].Assuming that Koomey's Law persists in the next 20 years, it is safe to say that Reversible Computing will play a key role in reducing the energy consumption and therefore, increasing the efficiency of computing machines.

2.8.4 Neuromorphic Computing

As [42] discusses, there is a lot of motivation to focus research on neuromorphic computing. The main motivations found in the literature were "Low Power" and "Parallelism" alongside motivations like "Scalability", "Fault Tolerance", and "Faster Computing". This is very close to the current challenges in the community of computing, which is mainly concerned about the need for more computing power and the increasing energy consumption. While the successes of today's neuromorphic computers are still limited, we are convinced that neuromorphic chips will have a more dominant role in the future as it addresses the mentioned central problems in computing. Additionally, neuromorphic hardware is likely to have a strong impact on the development of

AI. As one current trend in AI is to be inspired from the human brain, neuromorphic chips are a suitable platform for this approach. This is reasoned by the fact that neuromorphic hardware is able to simulate neuronal processes more effectively than state-of-the-art CPUs. With the increasing popularity and importance of AI in industry and society, it is imaginable that in the future, computing units with separate neuromorphic chips will be included. Eventually, neuromorphic computing could be integrated by heterogeneous computing [26].

As neuromorphic engineering grows bigger, there are several possible milestones to be achieved. A major problem is that currently, there is no simple way to design neuromorphic systems. As described earlier, each system needs to be designed on three levels of models, algorithms, and hardware. Therefore, it needs some effort to be created. This problem can be eased if standard approaches emerge in the future. Alongside a standardization of the field, a helpful development could be specialized programming languages for neuromorphic hardware [42].

Another vast field of possible future application of neuromorphic hardware is human-machine interfacing [6]. As in our future, we will probably be interacting with technical systems even more, it is very important to create simple and intuitive ways for humans to interact with machines. As neuromorphic hardware is more similar to biological neuronal systems than today's computers, it is easier to establish communication between electronic hardware and the human body.

2.8.5 Optical Computing

Another promising research area is optical computing. In optical computing, photons play a very important role. Compared to electrons, which are mostly used in conventional computers to ship data back and forth between logic circuits and memory, photons have no mass, travel at the speed of sound and do not need to flow through resistance like electrons, leading to thermal dissipation [14]. Possessing these advantages, photons provide an alternative solution to transport data in computer chips. The data transmission is highly accelerated by taking advantage of the photon speed. Also, the power consumption during the transmission is reduced [53].

Of late, applying optical computing in artificial intelligence has become a trending research topic. Besides accelerating the data transmission, photons also possess potential to accelerate the calculations. Especially, the matrix multiplication benefits from making use of photons [53]. When matrix multiplication is required, the numbers in the matrices are converted into optical signals and larger numbers represent brighter beams. Then, the whole multiplication is divided in several smaller multiplications and each one is handled by a single 'cell' of the optical chip. Imaging the cell as a water pump with two inputs and two outputs, beams from different inputs can be sped up or slowed down in the pump and mixed together before they are let out. By controlling the speed of the pump, the brightness of the output beams can be controlled as well [14].

Currently, optical computing faces a lot of difficulties. Chips that work only with optical components have not been designed yet. So, all current optical chips contain traditional transistors, which are intended to conduct electricity. However, conductivity depends on free charge carriers which can absorb light particles. This may lead to power loss [20]. If the problem of power loss is not overcome, it will be difficult to build complex circuits and enhance the signal-to-noise ratio with optical components [53]. Furthermore, designing optical memory, increasing robustness of optical computing, and realizing optical interconnection for inter-chip communication are also challenges and issues for the development of optical computing [53].

In spite of the aforementioned unresolved issues, optical computing has already been applied to accelerate calculations in deep learning [14], bringing a positive influence to other areas in artificial intelligence such as neuromorphic computing [44] and quantum computing [49]. Optical computing is expected to be further applied in artificial intelligence and taken full advantage of its strengths to reduce the power consumption, make data transmission more efficient, and speed up the computing processes.

2.9 Conclusion

Having given an overview of a wide variety of new technologies like neuromorphic systems, quantum computing, biological programming, heterogeneous computing and cloud computing, we conclude this chapter.

Nowadays, a lot of efficient neuromorphic chips are designed and applied in several areas such as image processing and bionics. In addition, neuromorphic chips promote the development of neuroscience. For quantum computing, the companies IBM and D-Wave have their own qubit system with different quantum mechanics.

For biological programming, the current problem is not the computation speed but how to extract useful information from the experiment results. However, biological computing is now widely researched because of its high potential on accelerating the computing process and reducing power consumption. Moreover, heterogeneous computing and cloud computing are two promising research areas to increase the computational capability of artificial intelligence. In addition to an overview of the current research in computation technology, we offered an insight into some of the trends believed to be most likely involved in our lives over the next 20 years.

The efforts that already have been made in the research of reversible computing show the potential of this technology. Energy consumption will be one of the big factors that limit future CPUs. When it comes to the efficiency parameter derived from Koomey's law, reversible computing has the biggest potential in saving energy and memory space. To fulfill Koomey's law. it is most likely that new technologies are needed as current ones are approaching their limits. Particularly good candidates for these new technologies seem to be approaches like neuromorphic, quantum, and optical computing. While neuromorphic hardware is still waiting for its widespread application, the current results are highly promising. Neuromorphic computing seems to give answers to central challenges of current computing and therefore, we assume it will play a major role in the future of computing.

Quantum communication is leaping out of the labor as an important application of quantum computing and on the experimental side, some success with a small number of qubits has been achieved. Increasing the number of qubits and reducing the error rate is what needs to be done to achieve a better computational performance for quantum computing.

We are convinced that computational power is a key factor in advancing artificial intelligence to a new level. As described, there are many promising technologies to improve computational power in the future. We are confident that the introduced technologies will play an important role in the development of future AI and are eager to learn about future developments.

References

[1] Christoph Adami. "Biological Programming". In: *Nature* 446 (2006), pp. 263–264.

[2] A.G. Aruna et al. "A Study on Reversible Logic Gates of Quantum Computing". In: *International Journal of Computer Science and Information Technologies*. Vol. 7. 2016, pp. 427–432.

[3] John Backus. "Can Programming Be Liberated from the Von Neumann Style?: A Functional Style and Its Algebra of Programs". In: *Commun. ACM* 21.8 (1978), pp. 613–641.

[4] Antoine Bérut et al. "Experimental verification of Landauer's principle linking information and thermodynamics". In: *Nature* 483 (2012), pp. 187–189.

[5] P. G. Bishop. "Using reversible computing to achieve fail-safety". In: *Proceedings The Eighth International Symposium on Software Reliability Engineering*. 1997, pp. 182–191.

[6] Fabio Boi et al. "A Bidirectional Brain-Machine Interface Featuring a Neuromorphic Hardware Decoder". In: *Front. Neurosci.* 2016.

[7] M. N. Bojnordi and E. Ipek. "Memristive Boltzmann machine: A hardware accelerator for combinatorial optimization and deep learning". In: *IEEE International Symposium on High Performance Computer Architecture*. 2016, pp. 1–13.

[8] J. Bornholt et al. "Toward a DNA-Based Archival Storage System". In: *IEEE Micro* 37.3 (2017), pp. 98–104.

[9] E. Chen et al. "Advances and Future Prospects of Spin-Transfer Torque Random Access Memory". In: *IEEE Transactions on Magnetics* 46.6 (2010), pp. 1873–1878.

[10] S. K. Chowdhary, A. Yadav, and N. Garg. "Cloud computing: Future prospect for e-health". In: *3rd International Conference on Electronics Computer Technology*. Vol. 3. 2011, pp. 297–299.

[11] Francesco Conti and Luca Benini. "A Ultra-Low-Energy Convolution Engine for Fast Brain-Inspired Vision in Multicore Clusters". In: *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE): 9 - 13 March 2015, Grenoble, France*. IEEE, 2015, pp. 683–688.

[12] Intel Corporation. *2015 Annual report.*

[13] R. H. Dennard et al. "Design of ion-implanted MOSFET's with very small physical dimensions". In: *IEEE Journal of Solid-State Circuits* 9.5 (1974), pp. 256–268.

[14] Jesse Dunietz. "Light-Powered Computers Brighten AI's Future". In: *Scientific American* (June 2017).

[15] L. Eeckhout. "Is Moore's Law Slowing Down? What's Next?" In: *IEEE Micro* 37.4 (2017), pp. 4–5.

[16] A. Fert. "Challenges and emerging directions in spintronics". In: *IEEE 25th International Conference on Micro Electro Mechanical Systems*. 2012, pp. 1–2.

[17] M. P. Frank. "Throwing computing into reverse". In: *IEEE Spectrum* 54.9 (2017), pp. 32–37.

[18] Pulak Ranjan Giri. "A Review on Quantum Search Algorithms". In: *Quantum Inf. Process. Vol. 16 No. 12 (2017) 1-36* (2016).

[19] C. Gordon et al. "An artificial synapse for interfacing to biological neurons". In: *IEEE International Symposium on Circuits and Systems*. 2006.

[20] Larry Hardesty. *Optoelectronic microprocessors built using existing chip manufacturing*. 2015.

[21] J. Koomey et al. "Implications of Historical Trends in the Electrical Efficiency of Computing". In: *IEEE Annals of the History of Computing* 33.3 (2011), pp. 46–54.

[22] Wang Laiyuan et al. "Controllable Multiple Depression in a Graphene Oxide Artificial Synapse". In: *Advanced Electronic Materials* 3.1 (2016).

[23] R. Landauer. "Irreversibility and Heat Generation in the Computing Process". In: *IBM Journal of Research and Development* 5.3 (1961), pp. 183–191.

[24] Yann Lecun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE*. 1998, pp. 2278–2324.

[25] Adleman Leonard. "Molecular Computation of Solutions to Combinatorial Problems". In: *Science* 226 (1994), pp. 1021–1024.

[26] X. Liu et al. "Harmonica: A Framework of Heterogeneous Computing Systems With Memristor-Based Neuromorphic Computing Accelerators". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 63.5 (2016), pp. 617–628.

[27] J. Lu et al. “Implementation of CMOS neuron for robot motion control unit”. In: *International SoC Design Conference*. 2013, pp. 9–12.

[28] Wolfgang Maass. “Networks of spiking neurons: The third generation of neural network models”. In: *Neural Networks* 10.9 (1997), pp. 1659–1671.

[29] Ilya Sinayskiy Maria Schuld and Francesco Petruccione. “An introduction to quantum machine learning”. In: *Contemporary Physics* (2014).

[30] C. Mead. “Neuromorphic electronic systems”. In: *Proceedings of the IEEE* 78.10 (1990), pp. 1629–1636.

[31] Paul A. Merolla et al. “A million spiking-neuron integrated circuit with a scalable communication network and interface”. In: *Science* 345.6197 (2014), pp. 668–673.

[32] Sparsh Mittal. “A Survey of Techniques for Architecting and Managing Asymmetric Multicore Processors”. In: *ACM Comput. Surv.* 48.3 (2016).

[33] G. E. Moore. “Progress in digital integrated electronics”. In: *IEEE Solid-State Circuits Society Newsletter* 11.3 (2006), pp. 36–37.

[34] T. Nakano. “Biological Computing Based on Living Cells and Cell Communication”. In: *13th International Conference on Network-Based Information Systems*. 2010, pp. 42–47.

[35] John von Neumann. “First Draft of a Report on the EDVAC”. In: *IEEE Ann. Hist. Comput.* 15.4 (1993), pp. 27–75.

[36] William D. Nordhaus. “Two Centuries of Productivity Growth in Computing”. In: *The Journal of Economic History* 67.1 (2007), pp. 128–159.

[37] Jane Qiu. “Quantum communications leap out of the lab”. In: *Nature* 508 (2014), pp. 441–442.

[38] S. G. Ramasubramanian et al. “SPINDLE: SPINtronic Deep Learning Engine for large-scale neuromorphic computing”. In: *IEEE/ACM International Symposium on Low Power Electronics and Design*. 2014, pp. 15–20.

[39] T. B. Rehman. “Cloud computing: A Juxtapositioning with grid computing”. In: *International Conference on Recent Innovations in Signal processing and Embedded Systems*. 2017, pp. 498–503.

[40] S. Sarkar et al. “Hardware accelerators for biocomputing: A survey”. In: *IEEE International Symposium on Circuits and Systems*. 2010, pp. 3789–3792.

[41] Jürgen Schmidhuber. “Deep Learning in Neural Networks: An Overview”. In: *CoRR* (2014).

[42] Catherine D. Schuman et al. “A Survey of Neuromorphic Computing and Neural Networks in Hardware”. In: *CoRR* (2017).

[43] M. Sharad, C. Augustine, and K. Roy. “Boolean and non-Boolean computation with spin devices”. In: *International Electron Devices Meeting*. 2012.

[44] Y. Shen et al. “On-chip optical neuromorphic computing”. In: *Conference on Lasers and Electro-Optics*. 2016, pp. 1–2.

[45] C. Shi et al. “Ultra-low latency cloud-fog computing for industrial Internet of Things”. In: *IEEE Wireless Communications and Networking Conference*. 2018, pp. 1–6.

[46] T. Shin et al. “Live demonstration: Real-time image classification on a neuromorphic computing system with zero off-chip memory access”. In: *IEEE International Symposium on Circuits and Systems*. 2016, pp. 449–449.

[47] Zhaogang Shu et al. “Cloud-Integrated Cyber-Physical Systems for Complex Industrial Applications”. In: *Mobile Networks and Applications* 21.5 (2016), pp. 865–878.

[48] M. M. Syiam et al. “Hardware implementation of neural network on FPGA for accidents diagnosis of the multi-purpose research reactor of Egypt”. In: *12th IEEE International Conference on Fuzzy Systems*. 2003, pp. 326–329.

[49] Dongyang Wang, Junjie Wu, and Xun Yi. “Optical quantum computing”. In: *11th International Conference on Natural Computation*. 2015, pp. 390–397.

[50] Lidan Wang, Meitao Duan, and Shukai Duan. “Memristive Perceptron for Combinational Logic Classification”. In: *Mathematical Problems in Engineering*. Vol. 2013. May 2013.

[51] Peter Wittek. *Quantum machine learning: what quantum computing means to data mining*. Academic Press, 2014.

[52] L. Zeng et al. “Spin wave based synapse and neuron for ultra low power neuromorphic computation system”. In: *IEEE International Symposium on Circuits and Systems*. 2016, pp. 918–921.

[53] Z. Zhao et al. “Optical computing on silicon-on-insulator-based photonic integrated circuits”. In: *IEEE 12th International Conference on ASIC*. 2017, pp. 472–475.

Chapter 3

Modularity: Building Blocks for AI

Pieter Bunnenberg
Stephan Böck
Johannes Gilg
Thomas Weinländer
Heike Wiedemann

3.1 Abstract

This article discusses modularity in the context of artificial intelligence (AI). Here, many different topics are relevant. Today, these are mainly the modularity of the human brain and methods to examine it, modular neural networks in the context of learning, and the modularization of neural networks to improve them. Research on these topics can provide us with artificial intelligence, consisting of several independent expert systems in the near future. Also, AI learning will be faster because with generative adversarial networks (GANs), big amount of real data will no longer be needed. Further, investigations on the modularity of the brain of humans and animals, like the TRACT method, can help us overcome problems like catastrophic forgetting in AI and moreover, provide a general knowledge of the functionality of the brain. This, in turn, will allow the exploration of new ways for creating intelligence.

3.2 Introduction

In the current quest for methods to achieve artificial intelligence (AI), many different techniques are explored. As presented in this book, there are several topics that have to be taken into account. Therefore, we want to present an approach that may help in this process: Modularity. Creating modules may not be a solution on its own, but it can support us by improving existing methods. Dividing a task into smaller sub-tasks (modules) is a well-known approach in computer science. The divide-and-conquer approach partitions complex problems. The smaller, easier to handle sub-problems can be seen as independent [9]. This allows a treatment in parallel, which grants benefits, e.g. reduction of computation time.

Aligning with the divide-and-conquer approach, the subsequent section shows how modularity helps handle intricate issues on artificial intelligence. To enable mankind build an AI emulating the human brain model, we need to understand the brain in the first place. Therefore, initially, the origins of modularity are briefly shown. In a next step, the current understanding of the modularity of the human brain is expounded and a method for investigating the modularity of the brain is presented. Further, reasons against highly specialized modularity explored by Baum [6] are given. In 1989, Sondak, et al introduced the importance of neural networks (NN) in the context of artificial intelligence [21]. Finally, the last section addresses the benefits of modularity for neural networks.

Section 3.4 picks up ideas previously presented in section 3.3 and gives a brief prediction on the trends that may arise. First, a general modular approach to AI development is introduced and its effects are discussed. The last sections refer to the TRACT and GANs approach and predict their influence on improving methods for developing AI.

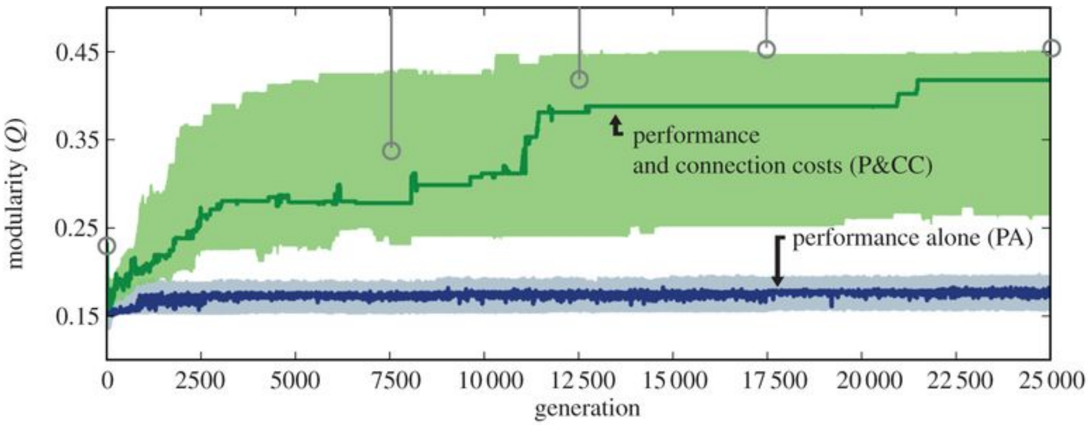


Figure 3.1: Increasing modularity with minimizing connection costs; *source: [8]*

3.3 State of the Art

This chapter shows the current State of the Art in researches on the modularity of the brain and how this can be used to develop AI. Therefore, the origins and the current understanding of the modularity of the human brain are discussed first.

3.3.1 The Evolutionary Origins of Modularity

One hypothesis is that modularity has evolved as a byproduct of selection in the evolution to reduce connection costs in a network [7]. Such costs include manufacturing and maintaining connections, energy to transmit along them, and signal delays, all of which increase as a function of connection length and number [1]. The concept of connection costs is straightforward in networks with physical connections, e.g. neural networks [8]. There are multiple studies which suggest that the summed length of the wiring diagram has been minimized, either by reducing long connections or by optimizing the placement of neurons [1]. Researchers have shown that maximizing performance and minimizing connection costs produce significantly more modular networks than only maximizing performance (see figure 3.1) [8].

3.3.2 Modularity of the Human Brain

To understand the modularity of the human brain, it is important to know its physical structure. If the brain has well defined regions or connections, then it is plausible to assign them special tasks. From this, it is possible to deduce the modularity of the human brain.

Physical structure of the embryonic brain In the vertebrate brain, it is clearly shown that the brain has morphological structures. Starting from the development of the embryonic brain, it is possible to split the brain into special modules called embryonic divisions. They are modular because the neuronal tissue of the divisions have a huge difference. Inside the embryonic divisions, the neurons and glial cells proliferate, migrate, and differentiate to distinctive structures [19]. The glial cells give structure to neurons and are also electrical insulators that support the transportation of chemical molecules.

Brain development and functional modules In the brain development the patterns of the embryonic structure have to be transformed according to the information saved in the corresponding genes. By that, the functional modularity of the brain will be introduced. To generate a functionality, more than one embryonic division needs to be in use and each division processes different information with a particular abstraction level [19]. In figure 3.2 on the next page, the visual, auditory, and motor functionality are shown, including the corresponding divisions in the human brain. Embryonic modules represent spatially separate, largely independent histogenetic fields. Each field gives rise to a coherent domain of gray matter that is later characterized by a particular way of information processing. Each domain contains several brain nuclei, or regions that are connected to nuclei or regions in other domains by fiber tracts, thus, forming neural circuits of different functional systems, e.g. blue, auditory system; red, motor system; and green, visual system.

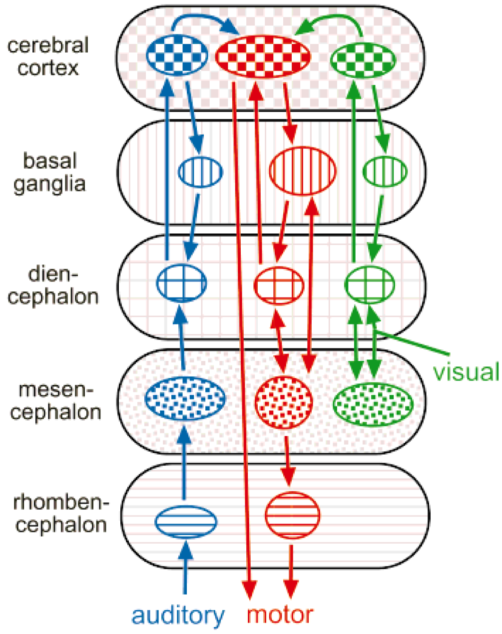


Figure 3.2: Relation between embryonic and functional modularity in the brain; *source: [19]*

Brain Functionality as a Aetwork Structure Better technologies in neuroimaging have led to the first chance to develop non-invasive functional network structures. The new imaging system, rs-fcMRI, can detect neuronal activity between different brain areas [18]. The network structure of the brain is called connectome and is structurally divided into different scaling levels. The mesocale describes the minicolumns and their connection patterns. Minicolumns consist of approximately 80 -100 neurons with a diameter of about 30-50 micrometres [22]. In 2014, a mesoscale connectome of a mouse brain was published [17]. The smallest one is the microscale, which defines the neuron to neuron connection. This level of detail is not feasible in the near future since the human brain contains an estimate of 10^{10} neurons with 10^{15} connections between them [22]. Having been published in 2013 and still available, Big Brain provides a high resolution 3D model of the human brain close to the microscale level. The model is constructed by slicing the brain into over 7000 histological brain sections [2]. Hence, this microscale is still not feasible for non-invasive methods and only provides a physical structure of the neurons instead of the functionality pattern. Latest research provides a computational approach to receive the microscale from a higher scale of a mouse brain [28]. In summary, due to the physical structure of the embryonic brain, it is possible to determine modules. However, latest research does not provide a functionality network pattern in microscale for the human brain, but provides computational algorithms to generate it for less complex brains.

3.3.3 Tracing Neuronal Circuits by Transneuronal Control of Transcription

The Lois laboratory has developed a method for tracing the flow of information across synapses called TRACT. When a neuron 'talks' - it transmits a chemical or electrical signal across a synaps. It also produces and sends along a fluorescent protein that lights up both the talking neuron and its synapses with a particular color. Any neurons 'listening' to the signal, receive this protein, binding to a so-called receptor molecule - genetically built-in by the researchers - on the receiving neuron's surface. The binding of the signal protein activates the receptor and triggers the neuron it is attached to in order to produce its own, differently colored, fluorescent protein. In this way, communication between neurons becomes visible. A semantic visualization is shown in figure 3.3 on the following page. Currently, the laboratory is using Drosophila fruit flies for their experiments. Using a type of microscope that can peer through a thin window installed on the fly's head, the researchers can observe the colorful glow of neural connections in real time as the fly grows, moves, and experiences changes in its environment [12]. If the technique gets improved in the future, one can image the connections in a human brain. Therefore, it could be proven whether or not the mind is modular and if yes, in what scope.

In figure 3.3, 'Donor' neuron is labeled in red. 'Receiver' neurons, that make synaptic contacts with the donor neurons are labeled in green. In contrast, gray labeled neurons are close to donor neurons but do not make synaptic contact.

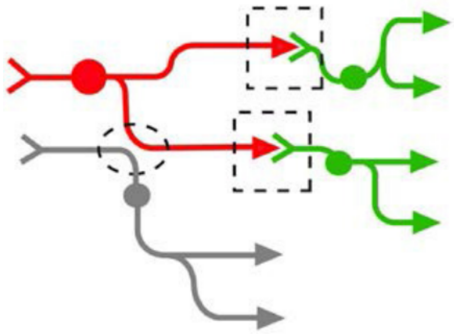


Figure 3.3: Using TRACT to detect synaptic connections between neurons; *source: [12]*

3.3.4 The Case Against Fixed Modules

There is also some evidence against the fixed and highly specialized modules described by Baum [6]. Various researchers have found that damage to brain areas, usually responsible for special functions or modules in this context, sustained early in the development of a brain can be at least partially compensated by other brain areas or modules [13][15][23]. The first example is the development of both visual maps in one hemisphere of the brain in a child born with only one hemisphere. Here, the optical nerve fibers were rerouted to the functional hemisphere, leading to the development of both visual maps [15]. The other two examples are about language development in children after a hemispherectomy and a perinatal stroke. Both show more speech-related activity in different areas like the visual cortex than compared to a control group [13][23]. While this is definitely no proof that there are no highly specialized modules in the human brain, it hints at more flexible modules and overall human brain, especially in early development phases.

3.3.5 Neural Reuse

The idea of more flexible modules gave rise to the theory of neural reuse which is thought to be a central organizational principle of the brain. According to this theory, it is quite common for neural circuits established for one purpose to be put to different use during evolution also normal development. According to this, neural circuits can continue to acquire new uses after an initial function is established [4]. There are three important findings in this context supported by data. First, a typical brain area is used by many different cognitive functions and tasks. Second, evolutionary older brain regions are highly reused and third, more recent functions will use more widely scattered brain areas [3]. This seems intuitive since newer functions have many modules, functionalities, and wiring to reuse. Researches could also show that neural reuse arises from the evolutionary pressure of varying task requirements. They showed that the agent with the highest amount of reuse is most adaptable to changing environments or tasks (see (A) in figure 3.4 on the next page).

From figure 3.4 on the following page (A) shows the fitness trajectory of the best individual in the population over time. Dashed vertical lines mark task swaps. Black-dotted horizontal line represents the 0.95 fitness threshold. The first four task swaps are labeled A1, B1, A2, and B2 respectively. Panel (B) shows the inset in panel (A).

3.3.6 Modularity Helps Against Catastrophic Forgetting

There are two types of forgetting. The first one is gradual forgetting that happens to us, humans. And the second one is catastrophic forgetting, which is a major problem while creating artificial intelligence with neural networks [14]. To learn new skills, neural network learning algorithms change the weights of neural connections, but old skills are lost because the weights that encoded old skills are changed to improve performance on new tasks [27]. The benefits of robots and artificially intelligent software agents will be extremely limited until we can solve the problem of catastrophic forgetting. In the chapter 'General AI', we have included a section on this common problem along with a solution. In the current section, we focus on the topic of modularity and its potential related to this problem. Modularity, which is widespread in biological neural networks [25], helps reduce catastrophic forgetting in artificial neural networks [10]. Modular networks are those that have many clusters of highly connected neurons that are only sparsely connected to neurons in other modules [25]. Modularity could allow learning new skills without forgetting old skills because learning can be selectively turned on only in modules learning a new task (see figure 3.5 on page 50) [10]. Researchers have shown, that networks using

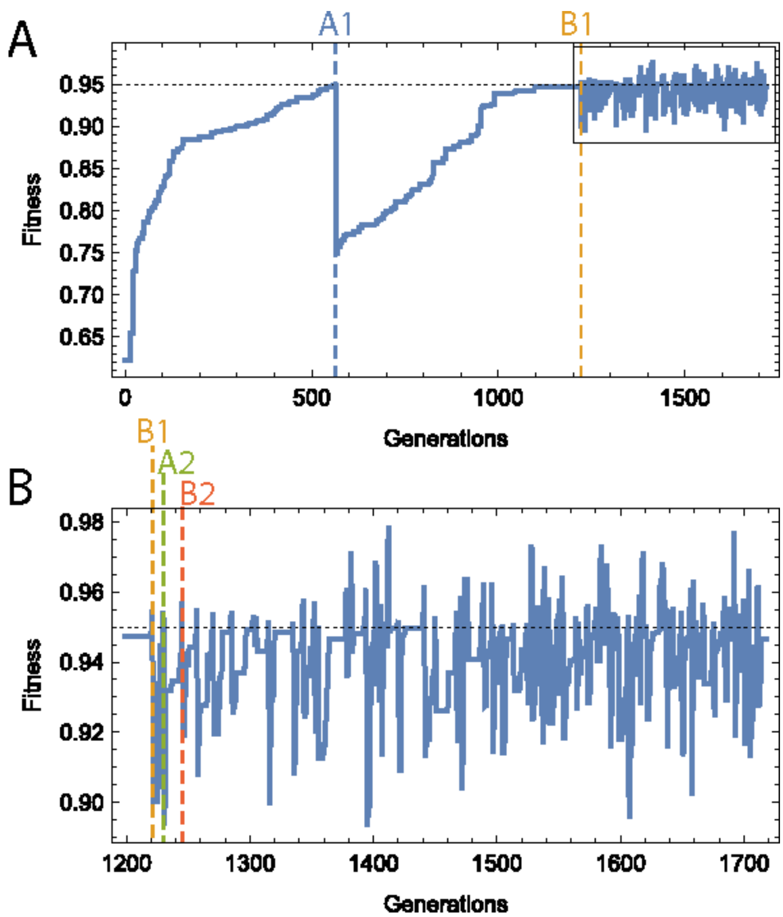


Figure 3.4: Increased adaptation over time in an evolutionary run; *source:[20]*

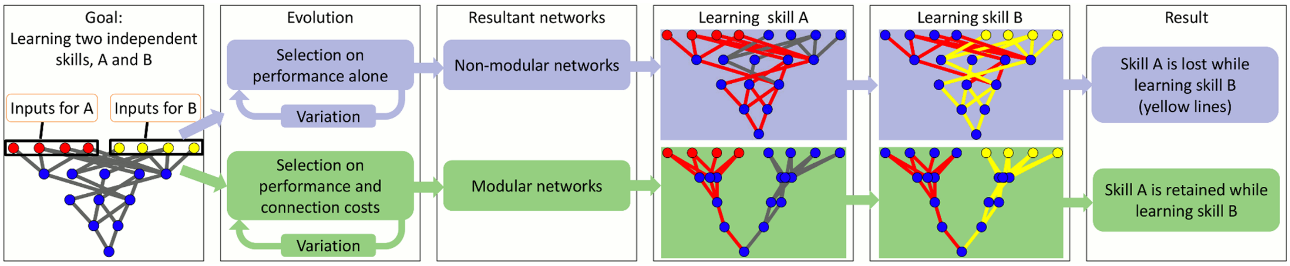


Figure 3.5: Learning in neural networks with modularity; *source: [10]*

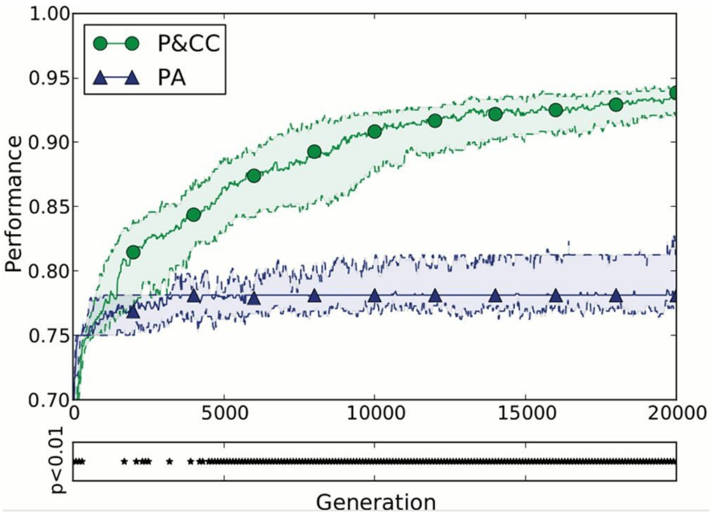


Figure 3.6: The graph shows the performance with modularity (green) and without modularity (blue); *source: [10]*

modularity are performing much better than ones without modularity (see figure 3.6). This point is further discussed in section 3.3.8.

3.3.7 Generative Adversarial Networks (GANs) as an Example of Module Interaction

One interesting and current example for modularity in machine learning is Generative Adversarial Networks. In this paper, Goodfellow et al. showed a general approach and laid the theoretical foundations. Generative Adversarial Networks are made up of two networks or modules, a generator network and a discriminator network. At the beginning, the discriminator network is only superficially trained with a relatively small number of real data compared to the traditional use of NNs, where a big amount of training data is used to reduce uncertainty. Then, the generator network creates some training data and the discriminator will determine whether the samples are 'real' or 'fake'. Therefore, the discriminator decides whether the data is real training data or if it comes from the generator. The generator can be trained on the data generated by it, labeled as real by the discriminator. When the discriminator can no longer label real and artificial data with a certain accuracy, it can also be trained further on the real data. This circle continues until the generator can produce accurate data of the desired distribution [11]. This is an interesting example of machine learning modules that help each other to get better results. It also gives some insight into the workings of classification algorithms. The fully trained generator network creates data that the discriminator, a sort of classifier, interprets as real. The created data can be examined to find certain hints as the defining properties of the classifier.

3.3.8 Modularity within Neural Networks

Many Systems are presentable as a Network. Finding structure within all different kinds of networks is a huge field of interest in science. Newman found that many of them are naturally dividable into modules or communities [16]. As learned in section 3.3.6, neural networks can be improved by a modular composition. But,

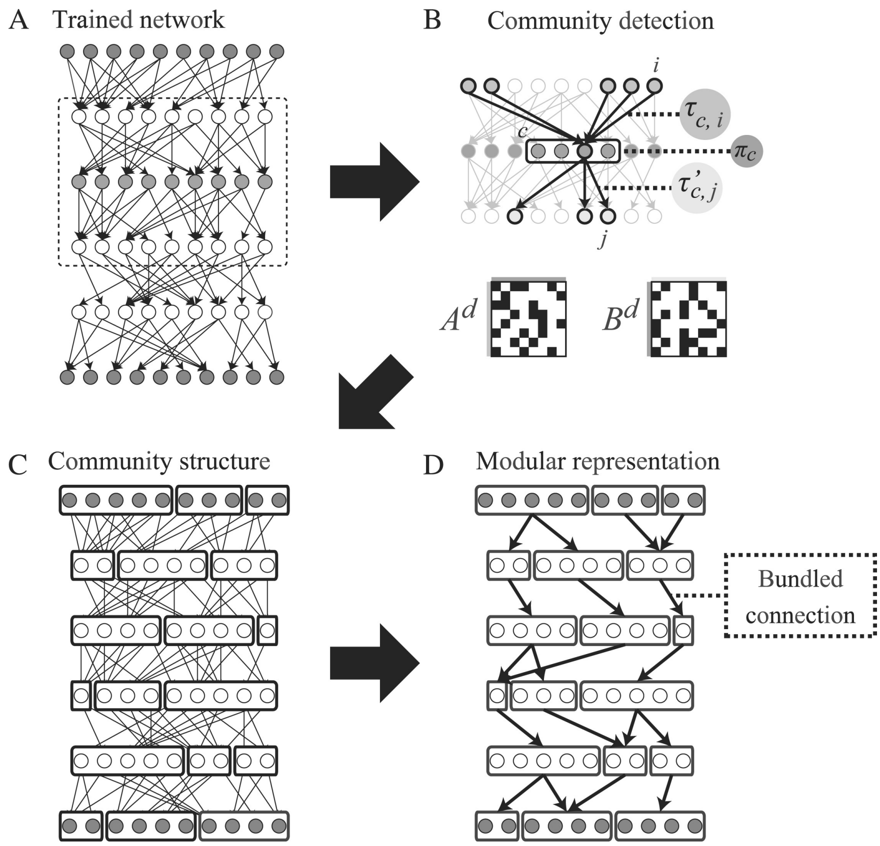


Figure 3.7: A trained layered neural network (A) is transformed into its modular representation (D). The communities in a layer are detected by evaluating the connections to the adjacent layers (B); *source: [26]*

how can non-modular neural networks profit from this? Newman's findings on structuring networks were taken by Watanabe, et al [26] and transferred to neural networks (NN). These researches led to an algorithm for dividing a NN into communities (modules). To achieve the community structure, we regard an already trained layered neural network as presented in figure 3.7 (A). Now, for every layer, the connections to adjacent layers are evaluated (B). Here, the relationships between knots are sought after. These are, e.g. many knots of layer d connected to just a few of layer $d+1$ or vice versa. As a result, the community structure is obtained (C). The edges between the different communities can be represented as bundles to further increase clarity (D) [26].

By transforming a trained neural network into its modular representation, one obtains the following benefits [26]:

- A layered neural network can be represented in form of smaller, independent networks. This allows a reduction in computation time by dividing the problem (parallelization).
- The knowledge about the modularity of a network allows estimations about the generalization error. The term 'modularity' in the context of networks means how well they are dividable into communities. A higher modularity implies a lower generalization error made by the network.
- Trained neural networks are better understandable. The community structure allows to extract connections between input, output, and hidden layers. Thus, it is easier to discover relations between data the NN was trained with.

The above points show how beneficial a transformation of a neural network into its modular representation is. Therefore, Watanabe et al. concluded that the modularity of an achieved community structure could be used as a rating to establish neural network regularization [26]. Tosh and McNally [24] did some research on the efficiency of modular and non-modular networks and discovered that the different types of networks perform better or worse depending on their size. For a big network size, a modular built up performs better and at small size, a non-modular network performs better. However, a modular network will outperform a non-modular one even at small network size, if the connective density must be kept low. This occurs e.g. if the energy needs are low [24].

Figure 3.8: Interaction of Google Duplex with a human user and with the business; *source:* [1]

This section showed that modularity is an important topic for neural networks. Finding a modular representation of a trained layered neural network improves its usability. Also, depending on the network's size and constraints, the performance can be increased.

3.4 Future Trends

As previously mentioned, modularity is sgnificant in artificial intelligence. It takes more of a supportive role by improving neural networks and helps with the understanding of the human brain. But, which direction are these technologies leading to? To answer this question, the current section takes the knowledge achieved from section 3.3 and identifies trends in the development of artificial intelligence in regard to modularity. A brief forecast on the impact of the technologies in the next 20 years along with its probability of relevance is described below.

3.4.1 Closed Domains Approach

Closed domain approach shows an AI development that sets the focus on creating a specialist. The AI is trained only in a sharply limited scope of application, thus, making it an expert in this domain. However, in other domains, it fails to produce acceptable results. Google Duplex is a prime example[1]. It is capable of making phone calls to arrange appointments at a hair saloon or reserve a table at a restaurant. In these situations, Duplex is indistinguishable from a human caller. However, it is not able to make real conversations in other contexts.

Looking into the future, the activity to create and improve such experts for special tasks will increase. This is a continuous development and for most users, the improvements of Google Duplex will be hidden in the background.

As seen in figure 3.8, a user interacts with the Google Assistant and then, as a hidden action, the needed task is selected. If there is a trained expert for this task, such as taking a reservation in a restaurant, the assignment is solved without human involvement. If there is no expert that can solve the task autonomously, a human operator will be included to manage the assignment[1]. This system has the advantages that it can schedule appointments regardless of whether the user understands the local language or the business is contactable when the user has time to make a call by himself. As a result of research and development, the trend to create a 'strong' AI consisting of specialized modules can be seen. The chapter - General AI has greater details listed regarding the classification of AI and possible requirements. In this part, the focus lies on a possible modular construction of AI. In terms of Google Duplex, more of such experts are likely to be added in the next years. For the users, the functionality of assistant systems like the Google Assistant will increase. The developers of Google Duplex want to make it possible to get business-related information like regular opening hours and deviations on holidays from companies that do not offer such content online. Google Duplex could make one phone conversation with the company and provide the received information online and available for everyone[1].

The probabilities that more experts will be developed, which are able to finish specialized tasks autonomously, is really high. As a reason for this, one can see that it is much easier and faster to solve strongly demarcated problems. Such a system consisting of a high amount of experts is able to solve tasks more efficiently than one big system that is trained to solve all problems with one algorithm. The shown example, Google Duplex, is a proof that such activities to develop experts have already been started.

As an impact of the development of specialized modules, assistant systems like Google Assistant will improve their performance with the ability of handling an increased variety of tasks. As a consequence, the use of such

[1] https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html (accessed April 26, 2019)

assistants will increase. Humans will be able to take advantage of support from experts within their specialized field of training. This could help the user to make appointments and get information easily in almost no time while the tasks are done in the background. On the other hand it is hard to distinguish between the positive impact of AI consisting of specialized modules and the continuous improvement of the environment. As an example for the usual improvement, the 'digitization' could be named, which means in this context that from smaller companies more information will be available online and the amount of businesses will increase, where online bookings are possible. Looking back into the last twenty years of development, which has been accompanied with the spread and improvement of smartphones, it has been hard to distinguish between different causes of this development.

Looking into the future raises the question - Could we achieve a strong AI in twenty years consisting of many expert modules? As an answer, we think that a strong AI without human help is not possible in this time step. It is realistic that a lot of usual tasks such as making appointments, reservations, and getting information about a company will be done autonomously by such experts. However, the assistant that selects the needed expert must be a strong AI. One could conduct a thought experiment to estimate the complexity of these problems: If a businessman has to make work-related and private appointments in different locations and some appointments need a special order, a good time management is needed. Additionally, if the overall time required to move between the different appointments should be minimized or the importance varies, the point to optimize the schedule book is quite complex. In such a situation, it is questionable whether the assistant would be able to make the appointments to the satisfaction of the businessman.

In conclusion, we believe that it is clear that the development of experts as part of a bigger AI will continue and make our life more comfortable by doing tasks for us. However, this impact is hard to distinguish from other improvements in our environment. The probability is very low that - in twenty years - a 'strong' AI consisting of specialized modules that is able to finish all tasks without human support will be built.

3.4.2 Trends TRACT

Previous methods for examining neural connections were time consuming and labor intensive, involving thousands of thin slices of a brain reconstructed into a three-dimensional structure. A laboratory using these techniques could only yield a diagram for one small piece of fruit-fly brain per year. Additionally, these approaches could not be performed on living animals, making it impossible to see how neurons communicated in real time[2]. These statements show that TRACT is very important for the future. It is the only method we have found so far that is applicable on living animals and faster than the previously used methods. Since the TRACT method is completely genetically encoded, it is ideal for use on laboratory animals such as Drosophila and Zebrafish. ultimately, the developer of TRACT hopes to implement the technique in mice to enable the neural tracing of a mammalian brain[2].

Currently, it doesn't seem that the human brain itself could be traced with the TRACT method. But tracing the brain of mice would qualify as good progress and seems possible in the near future. Perhaps, with these results, one could draw conclusions about mammal brains in general. Nevertheless, we think that with TRACT we will understand mammalian brains and possibly, human brains to a greater extent. For example, if researchers can track whether mammalian brains are modular or not and if so, in which scope they are modular, it can help building AIs. A measure to prevent catastrophic forgetting could be discovered, since humans do not face this issue. Even knowing which neurons are communicating with each other based on different activities of mice would be of huge impact on building AIs.

TRACT, however, can do more than producing wiring diagrams. The transgenic flies can be genetically engineered so that the technique prompts receiving neurons to produce proteins that have a function, rather than colorful proteins that simply trace connections[2]. "We could use functional proteins to ask, 'What happens to the fly if I silence all the neurons that receive input from this one neuron?' ", said the developer of TRACT. "Or, conversely, 'What happens if I make the neurons that are connected to this neuron hyperactive?' Our technique not only allows us to create a wiring diagram of the brain, but also to genetically modify the function of neurons in a brain circuit"[2].

So far, it works only with transgenic flies. However, it seems that this technique will soon work for mice as well. Probably, this technique could be used for designing new approaches of AI or verifying existing ones.

[2] `http://www.caltech.edu/news/new-technology-will-create-brain-wiring-diagrams-80863` (accessed April 26, 2019)

Additionally, it could be used to reconstruct the same structure and modularity of a mice brain in a microchip. This could lead to better experiments with a higher accuracy than inspecting real brains. Also, it would do no harm to animals. Instead of a microchip, the mouse brain could be simulated. In conclusion, we think that TRACT should be developed further on and it could have a huge impact on building AIs.

3.4.3 Trends GANs

A third trend in the context of modularity is in the field of Generative Adversarial Networks (GANs). In section 3.3.7 the functionality of GANs is described. With further research and development, the trend towards improving the functionality of GANs and increasing the number of possible applications can be seen. An example are the Wasserstein Generative Adversarial Networks (WGAN)[5]. It is said that the WGAN is an algorithm that offers, in contrast to traditional GAN training, a higher stability of learning and solves some other problems like mode collapse [5].

With such improvements, we think that the usage of GANs will increase and better training for NN will be possible. Also, it would suit applications where only a little amount of real data is available for training or an unrealistically high effort is necessary to record such data. This shows that the impact of GANs is high enough to increase its use for these purposes as well as new applications through improved algorithms. We believe that the probability of this impact is high. We see a continuous increase of the usage of GANs as an example of how modular systems could help to develop strong AI. A rising usage of AI in the future makes training very important. The impact of improved GANs could simplify this needed training and could therefore, help the general development of AI.

3.5 Conclusion

The previous sections showed which role modularity takes in different topics. From an evolutionary perspective, a modular structuring of the network of neurons within the human brain is a cost efficient approach to improve performance. Further, the physical structure of the brain is divisible into modules. This knowledge originates from the early stage of development, the embryonic brain. Here, the neural tissues of the individual modules physically differ from each other. For a deeper understanding of the modular functionality of the brain, the interactions between neurons must be observed. TRACT can help making connections visible, therefore, it enables further research on the modularity of the human brain. In contrast, reasons against a modular brain were discussed. Especially, the reuse of brain regions for other tasks speaks against a fixed modularity. Besides the study of the human brain, modularity plays an important role in neural networks, which are often used in AI affairs. Training a new task into a new module within the NN prevents forgetting of previously learned tasks. Also, the learning can be modularized. The GANs approach uses two networks (modules), one for data generation and one for evaluation of the generated data. The generator is trained only with generated data that fits the task (evaluated by the second network). Hence, a circle of mutual improvement arises. In addition, already trained non-modular neural networks can be transformed into their modular representation in retrospect. This grants benefits in usability and performance of the NN. Until now, research in the field of modularity has given us a good initial insight into the working of the human brain. Further, methods to improve neural networks are extensively researched, facilitating future studies on artificial intelligence.

A possible progress in the near future will be observed by the development of extremely focused expert systems which are combined under the administration of an AI. This approach on a strong AI is much easier to realize than a single system that must handle a large amount of tasks. However, an AI consisting of many subsystems does not seem to be smart because, as mentioned before, people have to intervene if the right expert could not be chosen or non-existent. At the same time, the progress in methods like TRACT and GANs has given new insights on the thinking process of living beings and is allowing better training of AIs. This combination will strongly make headway in creating an true artificial intelligence.

References

[1] Yong-Yeol Ahn, Hawoong Jeong, and Beom Jun Kim. “Wiring cost in the organization of a biological neuronal network”. In: *Physica A: Statistical Mechanics and its Applications* 367 (2006), pp. 531–537.

[2] Katrin Amunts et al. “BigBrain: an ultrahigh-resolution 3D human brain model”. In: *Science* 340 (2013), pp. 1472–1475.

[3] Michael Anderson. “Evolution of Cognitive Function via Redeployment of Brain Areas”. In: *The Neuroscientist* 13.1 (2007), pp. 13–21.

[4] Michael L. Anderson. “Neural reuse: A fundamental organizational principle of the brain”. In: *Behavioral and Brain Sciences* 33.4 (2010), pp. 245–266.

[5] Martin Arjovsky, Soumith Chintala, and Léon Bottou. “Wasserstein generative adversarial networks”. In: *International Conference on Machine Learning*. 2017, pp. 214–223.

[6] Eric B Baum. *What is thought?* MIT press, 2004.

[7] Werner Callebaut, Diego Rasskin-Gutman, and Herbert A Simon. *Modularity: understanding the development and evolution of natural complex systems*. MIT press, 2005.

[8] Jeff Clune, Jean-Baptiste Mouret, and Hod Lipson. “The evolutionary origins of modularity”. In: *Proc. R. Soc. B* 280.1755 (2013).

[9] Thomas H Cormen et al. *Introduction to algorithms*. MIT press, 2009.

[10] Kai Olav Ellefsen, Jean-Baptiste Mouret, and Jeff Clune. “Neural modularity helps organisms evolve to learn new skills without forgetting old skills”. In: *PLoS computational biology* 11.4 (2015).

[11] Ian Goodfellow et al. “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems*. 2014, pp. 2672–2680.

[12] Ting-hao Huang et al. “Tracing neuronal circuits in transgenic animals by transneuronal control of transcription (TRACT)”. In: *Elife* 6 (2017).

[13] Frédérique Liégeois et al. “Speaking with a single cerebral hemisphere: fMRI language organization after hemispherectomy in childhood”. In: *Brain and Language* 106.3 (2008), pp. 195–203.

[14] Michael McCloskey and Neal J Cohen. “Catastrophic interference in connectionist networks: The sequential learning problem”. In: *Psychology of learning and motivation*. Vol. 24. 1989, pp. 109–165.

[15] Lars Muckli, Marcus J. Naumer, and Wolf Singer. “Bilateral visual field maps in a patient with only one hemisphere”. In: *Proceedings of the National Academy of Sciences* 106.31 (2009), pp. 13034–13039.

[16] Mark EJ Newman. “Modularity and community structure in networks”. In: *Proceedings of the national academy of sciences* 103.23 (2006), pp. 8577–8582.

[17] Seung Wook Oh et al. “A mesoscale connectome of the mouse brain”. In: *Nature* 508 (2014), p. 207.

[18] Jonathan D Power et al. “Functional network organization of the human brain”. In: *Neuron* 72.4 (2011), pp. 665–678.

[19] Christoph Redies and Luis Puelles. “Modularity in vertebrate brain development and evolution”. In: *Bioessays* 23.12 (2001), pp. 1100–1111.

[20] Matthew Setzler and Eduardo J. Izquierdo. “Adaptability and Neural Reuse in Minimally Cognitive Agents”. In: *CogSci*. 2017.

[21] Norman E Sondak and Vernon K Sondak. “Neural networks and artificial intelligence”. In: *ACM SIGCSE Bulletin*. Vol. 21. 1. ACM. 1989, pp. 241–245.

[22] Olaf Sporns, Giulio Tononi, and Rolf Kötter. “The human connectome: a structural description of the human brain”. In: *PLoS computational biology* 1.4 (2005).

[23] Jan-Mendelt Tillema et al. “Cortical reorganization of language functioning following perinatal left MCA stroke”. In: *Brain and Language* 106.3 (2008), pp. 184–194.

[24] Colin R Tosh and Luke McNally. “The relative efficiency of modular and non-modular networks of different size”. In: *Proc. R. Soc. B* 282.1802 (2015).

[25] Günter P Wagner, Mihaela Pavlicev, and James M Cheverud. “The road to modularity”. In: *Nature Reviews Genetics* 8.12 (2007).

[26] Chihiro Watanabe, Kaoru Hiramatsu, and Kunio Kashino. “Modular representation of layered neural networks”. In: *Neural Networks* 97 (2018), pp. 62–73.

[27] Bernard Widrow and Michael A Lehr. "30 years of adaptive neural networks: perceptron, madaline, and backpropagation". In: *Proceedings of the IEEE* 78.9 (1990), pp. 1415–1442.

[28] Tielin Zhang, Yi Zeng, and Bo Xu. "A computational approach towards the microscale mouse brain connectome from the mesoscale". In: *Journal of integrative neuroscience* 16.3 (2017), pp. 291–306.

Chapter 4

Collaboration: Emulating Human Behaviour

Arifi Ridon
Fatma Cheikh Rouhou
Patrick Giang
Lukas Fichtner
Cengiz-Han Polat

4.1 Abstract

Since inception, human beings have developed cognitive as well as functional capabilities, evolving into the most intelligent life form on the planet. The corollary is owed to the fact that humans have always possessed an inherent capacity to communicate and cooperate. By means of language and mutual expression, we are able to combine efforts, express complex thoughts and opinions, and accelerate their learning through nurturing and education.

The following chapter encompasses the benefits of communication and collaboration for the purpose of Artificial Intelligence (AI). Initially, it touches upon the concept of Multi-Agent Systems (MAS) and Multi-Agent Learning (MAL), which provide a suitable framework for cooperating intelligent systems. Subsequently, section 4.2 contains an overview of milestones in the current field, including the major stages of development and a few popular learning algorithms. After which, the chapter is narrowed down onto two key research directions - deep reinforcement learning (DRL) and inter-agent communication in MAS.

The subsequent section 4.3 suggests certain trends for future research. It entails strategies on scaling up MAL to solve problems that are larger and more complex in nature. Being able to learn effective means of communication between several agents and via language between humans and AI would be exemplary to a successful outcome. The end of the section entails an elucidation of these trends towards the development of AI. The chapter is concluded with a short summary and the key takeaways derived from the text.

4.2 Multi-Agent Systems

The past few decades have observed decentralized approaches being increasingly adopted to solve complex real-world problems. For instance, an attempt to regulate traffic by distributed traffic light control has proven to be effective. However, it would not be possible to realize such a regulation without the communication and cooperation between different entities involved. It is for this reason that autonomous driving must achieve called communication and cooperation between various units to derive applicable solutions.

The field of study concerning multiple entities working cooperatively in order to solve problems is known as distributed systems, which in conjunction with Artificial Intelligence (AI), is commonly referred to as Distributed Artificial Intelligence (DAI). In general, DAI can be split into two distinct areas, namely distributed

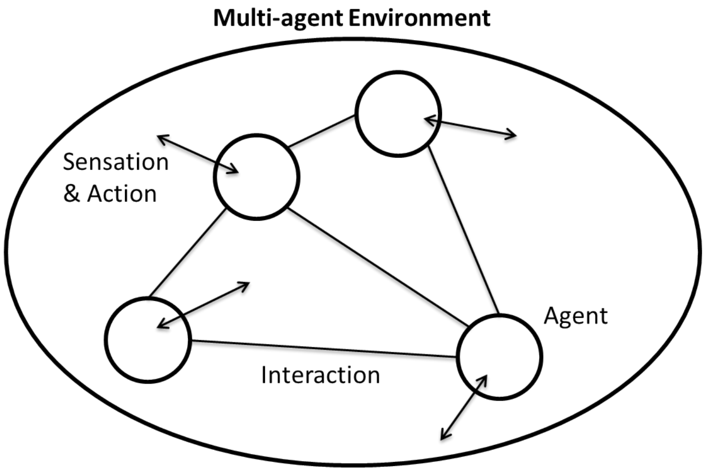


Figure 4.1: Example of a Multi-agent System; *source: own illustration*

problem-solving and multi-agent systems (MAS). The former inquires into the decomposition of a problem-solving process and its distribution among multiple nodes as well as the collective construction of a solution to the problem. The second area, which is discussed to a greater extent, underscores the joint behaviour of agents with a certain degree of autonomy and the complexity arising from their interactions.

A significant proportion of the research intends to apply Machine Learning (ML) concepts on multi-agent Systems. The underlying rationale being that an MAS is typically designed to tackle problems in large, complex, dynamic, and unpredictable environments, where it becomes nearly impossible to assign the agents' behaviour a priori. Learning provides a solution by allowing the agent - through interaction with its environment - to acquire new behaviour without human intervention.

4.2.1 Basics

The following section aims to delineate the basics of a multi-agent system and multi-agent learning (MAL). Essentially, it stems from the papers of Luke and Panait [15] and Weiss and Tuyl [24].

An agent is a computational mechanism that shows a high degree of autonomy, performing actions in the same environment that it receives information from. An environment having multiple agents interacting with each other with added constraints is called a multi-agent environment (see figure 4.1). At any given time, called constraints ensure that agents do not know everything about their surroundings including the inner state of other agents. Otherwise, the whole setup may be represented as a master-slave interaction. The 'agents' are able to know the exact behaviour of the other 'agents' and act synchronously, rendering them as the extended arms of a master controller. Additionally, if the domain requires no interaction at all, it can be decomposed into separate, stand-alone tasks, each of which are solvable by a single agent.

Multi-Agent Learning Multi-agent learning is the application of ML to problems involving multiple agents. MAL is defined by the following two distinct features: First, the involvement of several agents introduces an unusually large search and action space. One learning agent discovers a set of behaviours for a group of agents. The new learnt behaviour of this particular agent may result in unpredictable changes in the learnt behaviour of all other agents.Secondly, MAL may involve multiple learners, each learning and adapting in the context of others. This introduces game-theoretic issues to the learning process. Both features result in a dynamic environment in which the agents interact with each other.

Machine Learning Methods There are three fundamental approaches to learning in a single-agent environment, viz., supervised learning, unsupervised learning, and reward-based learning. The primary difference between these approaches is the kind of feedback the learner is provided with.During supervised learning, the critic provides the correct input, whereas in unsupervised learning, the critic provides no feedback at all. In reward-based learning, the critic conducts a quality assessment of the learner's output.

Supervised learning involves the agent to learn from a training set with labeled examples. Each example provides a certain situation and specifies, by means of the label, the expected outcome. This allows the agent to extrapolate or generalize, i.e., be able to determine the labels for data samples that it is not trained on. Despite creating significant value, this kind of learning is not sufficient for learning from interaction with the outer environment; where an agent learns from its own experience instead of given examples.

Unsupervised learning typically consists of finding an underlying structure in given data. Contrary to supervised learning, the data that the agent learns from is not labeled beforehand. Therefore, the agent is not provided with a feedback as a result of its actions. Such reward-based learning, does not provide examples of correct behaviour. Rather, it aims to maximize a reward signal, whose value directly depends on the learner's actions [21].

Majorly, research carried out in the field of MAL adopts reward-based methods, especially reinforcement learning (RL).In single-RL, an agent learns by interacting with its static environment. At each time step, the agent perceives the state of the environment and takes an action, which causes the environment to change its state. A predefined reward signal evaluates the quality of each action, and the agent collects the maximum cumulative reward possible through its interaction with the environment. The reward is a weighted sum of the current value, the reinforcement obtained when performing an action, and the expected value of future actions originating from current state [11].

In a nutshell, single-RL is described within the framework of Markov decision processes (MDPs).While providing a solid mathematical framework for single-agent learning, MDPs do not offer the same theoretical grounding for MAL. MDPs assume that the agent's environment is stationary and contains no other adaptive agents, which is in direct contrast to the dynamic environment of an MAL explained in section 4.2.1. Currently, extension of the MDP framework such as Markov games and joint action learners have been developed. Both approaches allow the introduction of multiple adaptive agents with interacting or competing goals [4][11].

4.2.2 Past Milestones

In the past years, MAL has developed along with techniques that have significantly impacted the research. This process of development is characterized by two periods, which we will further elaborate in the following section: The former period began in the 1980s, when several concepts and realization methods regarding MAL were explored. This first step towards MAL was of the breadth-first exploration type. Initial investigations known as adaptive parallel computations dominated that period. Inspired by biological examples found in eusocial insects such as termites, wasps, bees, and swarming vertebrates, these investigations involved developing nature-inspired MAS. The collective behaviours of swarms observed in nature provided striking proof that such self-organized systems composed of simple, unintelligent and purely reactive agents can accomplish sophisticated tasks and achieve a close to intelligent behavior [3]. These investigations paved the way to what is known as swarm intelligence, today. Other techniques examined were social learning, evolutionary computation, neural networks, and interactive and imitation learning.

The analysis and examination of called techniques led to the development of the artificial life field, for which first conferences such as the PPSN (parallel problem solving from nature) and A-life were organized [24].

Stemming from the field of artificial life, several techniques seem to play a suitable role for MAL purposes. Their relevance is highlighted by recent research showing a formal connection to RL. One such link, for instance, exists between multi-agent Q-learning and co-evolutionary algorithms [16]. Another one is described by the relation between swarm intelligence and RL [10].

Early multi-agent reinforcement learning (MARL) research was published by Whitehead, Tan, and Littman. Exemplary, Tan's 1993 paper demonstrated how RL agents could learn cooperative behaviour in a simulated social environment [23]. Subsequently, for the first time, MAL workshops were held and journal special issues appeared (e.g. [9][25]).This mushrooming phase resulted in a general understanding on the role of learning in multi-agent environments [24]. The expertise gained in these years formed the basis for the second period.

Since the onset of the 21st century, researchers have gained a better understanding of MAS and its applications. Their research has matured from being largely exploratory to focusing on certain MAS techniques, especially MARL, and establishing the theoretical foundations of MAS [24]. A general overview of MAL techniques and methods developed during the second phase is given by Busoniu, Babuška, and De Schutter and Shoham, Powers, and Grenager [2][18]. In the following, illustrative state-of-the art algorithms of both periods will be presented. Due to the large amount of existing algorithms and their variations in this field, we will describe only a few of them.

A demonstrative state-of-the-art algorithm of the first period is called Joint Action Learning and was introduced by Claus and Boutilier in 1998 [4]. In context of cooperative repeated games, a joint action learner (agent) learns so-called Q-values for joint behaviors (Q-values represent 'quality' of an action taken in a given state). This is in direct contrast to the Q-learning of individual learners for their individual actions [24].

One of the two algorithms of the second period is termed Nash-Q Learning. It is an algorithm introduced by Hu and Wellman and extends the independent Q-Learning algorithm to the multi-agent case [8].This is realized by using the Markov game framework. To find the optimal Nash Q-values (Q-values received in a Nash equilibrium) each agent maintains a model of other agents' Q-values, which are used to update their own Q-values. The goal is to find the best policy for each agent, relative to how the other agents behave [24].

Next, are a family of algorithms called Gradient Ascent Algorithms. One landmark member of this type is the infinitesimal gradient ascent (IGA) introduced by Singh, Kearns, and Mansour [19]. This algorithm is based on the limit of infinitesimal learning rates, where each agent updates its policy, such that it follows the gradient of expected payoffs. IGA is limited to two-player, two-action matrix games and is shown to converge to Nash equilibrium, however not in all cases. In order to overcome the limitation of the binary-action space, the generalized infinitesimal gradient ascent (GIGA) was introduced.In 2003, its inventor, Zinkevich, defined the 'regret measure'. Regret gauges how bad an algorithm performs as compared to the best static strategy, with the aim to guarantee at least zero average regret [27].As mentioned earlier, IGA does not converge in all cases. This led to an improvement of the IGA algorithm resulting in the IGA-WoLF (win or learn fast) [1].The principle idea of this algorithm is to make the learning rate large when WoLF is losing and kept small when a good strategy is found [24].

4.2.3 Current Research

As mentioned earlier, MAS are a highly active field of research, mainly due to their applicability to a variety of technological fields. Moreover,learning in MAS has been a widespread field of research as well, which may not be surprising, considering the current popularity of ML in general and the clear benefits that it carries when being applied to MAS.

Previous investigations have shown that learning in a multi-agent environment is more complex than in a single-agent case because agents simultaneously interact with their environment and each other [2]. The key challenge in MAL systems is learning how to best respond to the actions of other agents. Since all agents are learning simultaneously and thus, changing their behaviour, this may become a non-stationary problem. The problem of such systems is that actions performed by any agent influence the environment of all others. Thereby, also their goals and objectives. This implies that each agent is faced with a moving-target learning problem [24] since the environment that he perceives changes with every change in the other agents' behaviour. Simultaneous learning and adapting behaviour has worked well in a single-agent case, but most attempts break down when several agents share an environment and should learn to adapt their behaviour concurrently [7].

Non-stationarity is still a prevalent problem and with no effective solution [7]. This is one of the reasons for most developed MAL algorithms to have used a small number of agents in their environments. On the other hand, a higher-dimensional multi-agent domain could cause the learning algorithms to become intractable. Recently, a few work-around techniques have been developed with the goal of overcoming the non-stationarity problem. For instance, some researches attempted to generalize the system to a population level by modeling the agents as classes rather than individuals. Other researches went towards determining the degree of interaction among agents. This way, the environment of an agent could be reduced, by dropping the agents with a low degree of interaction.

Another problem is that most MAL algorithms assume that agents have access to all other agents' knowledge, which is not the case in real-life situations. Therefore, it has to be additionally learned. This task is taxing due to the non-stationarity of these agents [7]. A remedy for this impediment, which we will elaborate on later, would be to use deep learning techniques to correctly learn the models of the other agents.

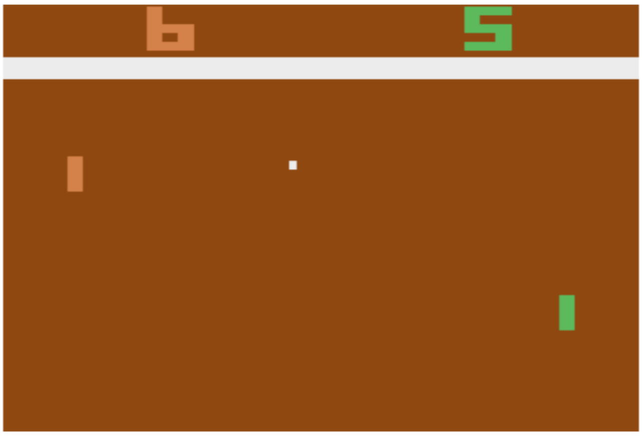


Figure 4.2: The Pong game. Each agent is represented by a paddle; *source: [22]*

In the light of these existing challenges and difficulties that MAL techniques and algorithms still encounter, an integral part of the current research aims at finding work-around solutions to them.

Multi-Agent Deep Reinforcement Learning Deep reinforcement learning (DRL) builds upon the conventional reinforcement learning methods. Since most of them require a function to be derived which is applicable to each state-action pair [21], the scalability for complex domains can only be provided by extra engineering work such as state abstraction. The idea of DRL is to use a deep NN to approximate those functions by letting the NN learn the state abstraction by itself over the raw sensory data.

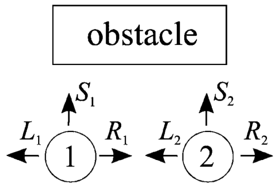
However, previous attempts have exhibited undesired behaviour like divergence. In order to overcome these, the first breakthrough approach added two countermeasures, first one being experience replay and second one being periodic updates of the target action-value function. The result is a so called Deep Q-Network (DQN) and was first introduced by Google DeepMind [14]. This allowed the team of Google DeepMind to achieve very promising results in many single agent scenarios, e.g. Atari Breakout. Ever since the success of DQNs, an important number of efforts have been put into developing even more powerful algorithms based on DRL [12].

Having observed that DRL might be a potential way to overcome the problem of high dimensionality, which holds more value in the multi-agent case, researchers are currently trying to apply DRL methods on MAS. An early approach from 2015 by Tampuu et al. dealt with the case of two agents playing the computer game Pong [22].

They used an autonomous learning algorithm for each agent. The only source of interaction was through their shared environment. For each agent, one deep Q-Network was trained. The results suggested that deep Q-Networks might also be a practical tool for the multi-agent case, as the two agents were able to develop different strategies for a game under different rewarding schemes through either being cooperative or competitive. In order to achieve this, they only relied on raw sensory data. This particular feature differentiates this approach from past MARL studies, which were mostly conducted in simple grid worlds or with agents having an abstraction for their perception.

Considering the fact that MARL suffers from non-stationarity, additional problems arise when transferring the knowledge of the single-agent case to the multi-agent case. In order to find a solution to this issue, many researchers are dedicating themselves to stabilizing DRL approaches for MAL. An example for a mechanism that is beneficial for DRL is experience replay [13], a method that stabilizes the training of the neural net and improves sample efficiency by randomizing over the data. This is achieved by storing an agent's experience for multiple time steps and performing learning updates on samples randomly drawn from the pool of stored experiences. Unfortunately, problems arise when using experience replay in a non-stationary multi-agent environment as the dynamics that generated the stored experiences no longer resemble the current learning dynamics.

A group of researchers were able to stabilize experience replay for the multi-agent case by creating mechanisms to lower the usage of obsolete data and thereby, were able to play scenarios in Starcraft, a computer strategy game with rather high complexity that is commonly used to test AI with multiple agents [5]. Observing the



Q	L_2	S_2	R_2
L_1	10	-5	0
S_1	-5	-10	-5
R_1	-10	-5	10

Figure 4.3: (Left) Two mobile agents trying to overcome an overcome a joint obstacle. (Right) The common Q-values of the agents for the state depicted to the left; *source: [2]*

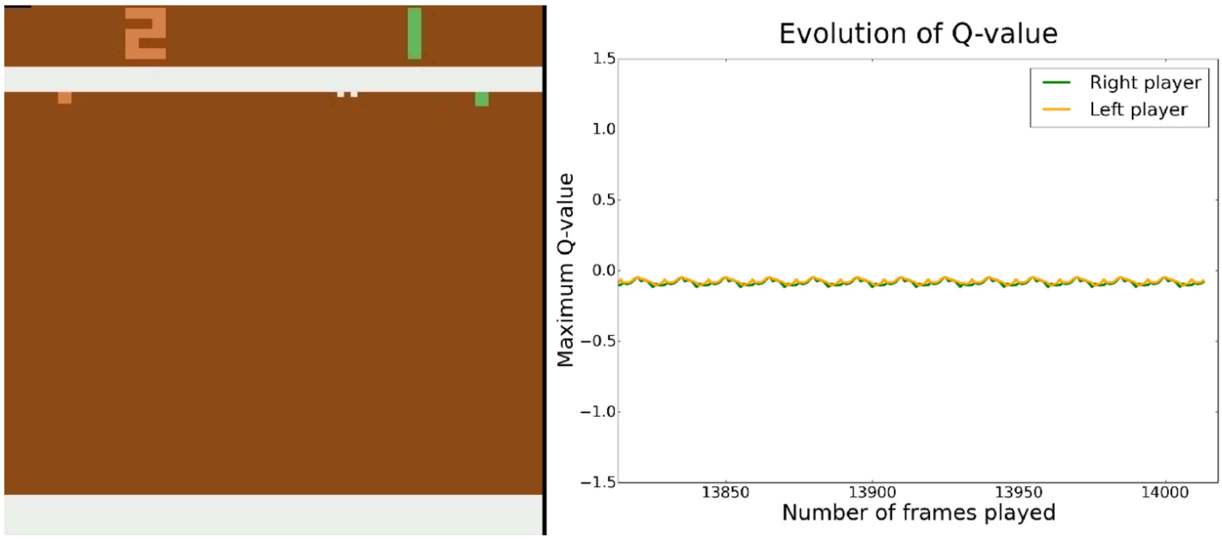


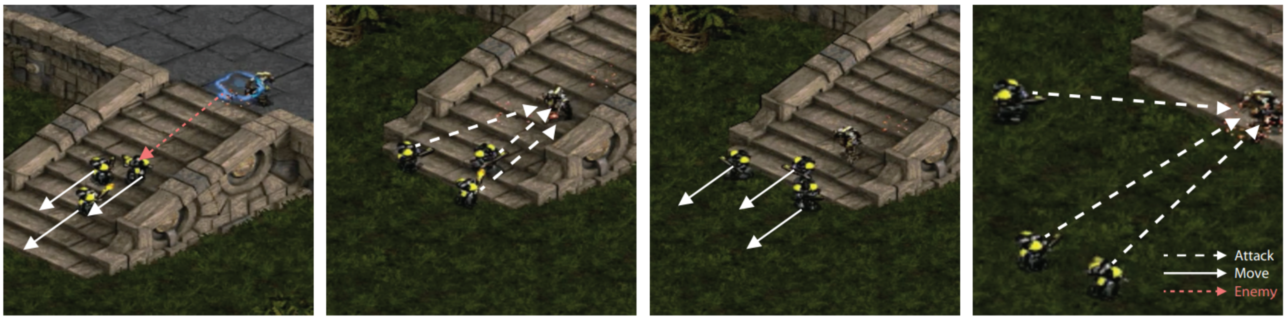
Figure 4.4: In the cooperative setting the agents sometimes reach a strategy that allows them to keep the ball in the game for a very long time; *source: [22]*

progress in DRL and the fact that it has already emerged as a go-to approach for MARL, it remains interesting to see how quickly and to upto what extent it can be made utilizable for the multi-agent case.

Communication between Agents Another interesting field of research is the learning of communication between agents. To illustrate the need for coordination, let us imagine a simple scenario of two agents passing by an obstacle while remaining in a certain formation. Since the formation has to be preserved, the only best option to fulfill this task would be to collectively either move left or right around the obstacle. Since these two possible actions can be seen as equally good, somehow, the agents need to perform the same action. Without an appropriate coordination mechanism, the chosen actions may be mutually inconsistent, which would cause the agents to collide or violate their formation [2].

Coordination in MARL necessitates some communication between agents. In the past, the communication has often been predefined by humans but recently, the focus has been shifted to letting agents learn the communication by themselves. Most of the time, the communication infrastructure is provided but no restrictions or instructions on how the communication should be performed are given. The goal is to let the agents develop their own collaboration strategies. However, there are quite different approaches to achieve this. Back to the previous example from the work of Tampuu et al. [22], where two agents were supposed to play Pong, communication did only take place through the environment, meaning there was no direct interaction between the agents. Still, the agents were able to develop cooperative or competitive strategies depending on the predefined reward, e.g. playing the ball horizontally to achieve a long phase where the ball kept moving between them without them needing to take large actions.

Furthermore, multiple approaches involving DRL as a tool for emergence of communication have surfaced in the past couple of years. In 2016, a group of researchers from the University of Oxford introduced the differentiable inter-agent learning (DIAL) [6], where each agent had a recurrent NN that, apart from approximating the function for taking actions, also outputs a message to transfer for each time-step. This message is then passed to other agents, who use it as additional input for the next time-step. The approach was designed for



(a) time step 1 (b) time step 2 (c) time step 3 (d) time step 4

Figure 4.5: : Hit and Run tactics in combat 3 Marines (ours) vs. 1 Zealot (enemy); *source: [17]*

partly observable settings. Results proved that agents trained with this approach were able to discover elegant communication protocols along the way of solving their task despite the experiment having not been complex. The authors describe their work as 'a first attempt at learning communication and language with deep learning approaches'. ([6], p.8) Nevertheless, they also admit that there is still a long way to go for understanding communication and language to as a whole.

In the same year, another team of researchers presented an approach for joint action learners in a fully observable setting called CommNet [20]. Assuming fully cooperative agents, there is no difference between having one controller for every agent or having one for all. Therefore, CommNet uses a single NN which has access to a communication channel carrying a continuous vector. Each agent receives the averaged message of all other agents. Since the communication is continuous, the model can be trained using back propagation. Combining this with reinforcement learning, communication can be learned. Results showed that this simple controller with a broadcast channel was able to learn communication for a varying set of agents. They observed that the model learned communication protocols which delivered meaningful information between agents. The model also outperformed models a range of other models. However, there were also limitations as the model was not able to handle heterogeneous agents and was not examined in terms of scalability for large amount of agents.

Building upon the two aforementioned approaches, in 2017, a bi-directional approach using recurrent NNs was developed and called Multiagent Bidirectionally-Coordinated Nets (BiCNet) [17]. Due to the bi-directionality of the approach, communication was not fully symmetric allowing different priority among agents, which could prevent ties between optimal actions. By using bi-directional recurrent NN structure which not only serves as communication channel but also as local memory saver, each agent can maintain its internal states and also share its information with collaborators. Afterwards, using a policy network that processes a shared observation and a local view, the action for each agent can be determined. The approach was tested in some Starcraft scenarios and proved itself to be able to master diverse settings. The agents even developed some human-level coordination strategies, e.g. the hit and run strategy where the agents repeatedly lure the enemy, run away and then attack him from a safe distance.

Summing up, all approaches hint that learning collaboration and communication is possible and an interesting field to pursue further. Nevertheless, every presented approach only resembles a little step forward to understanding communication in its full splendour.

4.3 Future Trends

Electricity distribution, control of traffic lights and coordination of robots [16] are some of the numerous promising applications of MAL. However, research in this area is still in an early stage and various problems need to be solved until more complex tasks are feasible through MAS.

This chapter suggests some directions for further investigation in MAS, which will most probably enable these applications. In the end, it will also describe how the development of general AI and thought can benefit from progress in these fields.

4.3.1 Large-Scale Problems for Multi-Agent Reinforcement Learning

Taking a look back at the past years and particularly at the development of MARL, one can observe a clear tendency towards increasing complexity of addressed problems. Starting out from the early attempts, there was a long period of stagnation regarding the complexity of problems. Most approaches dealt with rather simple settings like grid world and simple games. This originated namely, from the curse of dimensionality present in reinforcement learning, which becomes even more severe in MARL as the complexity grows rapidly with an increasing number of agents. Additionally, challenges like the non-stationarity of the learning problem for every single agent and the necessity of coordination further complicated the process. As mentioned earlier it was the success of DRL that set the impulse for MARL to slowly overcome the aforementioned phase of stagnation, enabling researchers to move in a decent pace from grid worlds to Atari-like games to slowly venturing on complex domains like Starcraft.

The trend of transitioning from small, simple problems to more complex and larger scales is highly dependent on the ability to handle large state-spaces and therefore, will highly profit from research that enables us to provide better state abstractions as it already was the case with neural networks leading to DRL. Furthermore, RL for the multi-agent case requires even more calculations than the single agent case which could limit its possibilities. But, observing the evolution of the computational power over the past years as well as the high efforts that are invested into providing more computational power, it is not unlikely that the numerous calculations necessary for MARL could be handled appropriately in the future.

Considering that large companies like Google and Facebook try to push their research forward in this domain, one can probably expect even more complex problem settings in the future, which could eventually open doors to large scale problems or even real world applications. There could be various applications ranging from actually solving complex meaningful simulation tasks to robots and devices working together to solve complicated tasks. Thus, having a more concrete application than science-fiction media. However, there are a lot of problems to be solved until we reach the point where MARL will have a significant impact on human life rather than merely being a research topic.

4.3.2 Self-Developed Communication among Agents

Revisiting the current research focus which tackles communication between agents, we may notice that a recent shift from predetermining communication strategies to initiating learning mechanisms enabling agents to develop their own strategies. This gives researchers not only the opportunity to address the necessity for coordination in a new way but also to observe how collaboration and communication emerges. Taking into account the boom of machine learning over the past couple of years, it is only a matter of time for someone to start applying those concepts to this domain in hope of overcoming the problems of MARL.

Since the topic of communication and especially, the emergence of communication/collaboration is not restricted to being explored by computer scientists, there might be an opportunity for more interdisciplinary work and approaches in the future, combining different viewpoints leading to a widespread investigation of this domain. Furthermore, as the results hint that communication and collaboration strategies learned by agents themselves have a high potential of being superior to predefined ones, one can expect more to come. Combined with the noticeable tendency towards more complex and sophisticated approaches to learn communication, this field of research could turn into a hot topic in the future, providing insights and tools applicable to various domains of science. As for now, a lot of approaches have profited from research in neural networks and a large amount of approaches have used them to learn communication. The involvement of neural networks seems to deliver promising results, becoming prevalent in research nowadays. Novel techniques to apply or utilize them might arise, which could also be applicable to learning communication. However, other machine learning methods might also bring success to this field.

For until now, the work done only resembles a small step towards understanding this complex topic, it might also take some time for this domain to eventually make an impact on our daily lives. Still, combining this with the trend of MARL moving towards large-scale problems, some day, devices might be equipped with some kind of AI to establish communication from scratch, without the need for human intervention. Heterogeneous devices could then, perhaps, based on their self-established communication, collaborate using MARL methods to solve complex tasks which would be currently unfeasible. This would bring a humongous number of possibilities of technological usage.

4.3.3 Language Development

Language and translation has always been a topic researchers have laid emphasis upon. An increasing number of research groups dealing with emergence of communication are trying to let agents establish a language along with the communication strategy. Profiting from the progress in machine learning, especially, deep learning, neural networks and reinforcement, first steps in development of language were made. Moreover, it became evident that agents are able to develop simple language. Many approaches involved some kind of MARL or self learned communication between agents. Therefore, development of language can highly benefit from progress made in the domain of communication and MARL.

The majority of approaches still involve a decent amount of human supervision. However, as researchers also try to understand how the language is developed, the trend goes towards lowering the degree of supervision. In the future, apart from developing simple word associations, agents should be able to have more complex conversations. In an ideal world, the agents would also have an understanding for the language and words that they are using. If humans would be able to still interpret the language developed by agents, this could lead to new opportunities for machine-human interaction. This however turns out to be quite complex, as [26] points out. Due to their different way of 'thinking', the developed languages of AI systems are not understandable for human readers, although being more efficient, just as human languages are complex to understand for machine learning systems. This is a huge problem since the decision process of modern AI systems like neural networks is also not comprehensible to humans. Because of this, both an efficient inter-agent language understandable for humans and an AI system being able to fully understand a human language with all its semantics would be a huge success.

4.3.4 Benefits for General AI

First of all, it is quite obvious that for the development of a general AI system with something similar to human thought, the problem space needs to be extended. The simplified environments used for MAL until now do not suffice here, but once this step is done it can provide several advantages.

A truly intelligent system cannot simply be designed and programmed by a human. Learning needs to be involved in the process and there is a lot to learn. Multiple agents can be of use here as they are able to learn simultaneously and maybe even specialize in a certain field and share their experiences with each other. This is also valuable when new systems need to be integrated. Through the MAS framework, a new agent can quickly learn from the other agents and benefit from their knowledge. Additionally, by learning to communicate at the same time, the information and knowledge can be exchanged efficiently. The ability to communicate is always necessary for general AI because the system needs to be able to interact with its environment and other systems to be considered as intelligent. This ability is also convenient for tasks that demand collaboration of several systems because often, the collective coordination of several entities is necessary to find a solution to a problem at hand.

Nevertheless, the comprehensibility of AI communication is still a problem. Understanding it or being able to communicate with an AI system is necessary for humans to give instructions and evaluate its thoughts and understanding. Intelligence is not provable without being able to understand a system's decision process.

4.4 Conclusion

Since inception, traditional AI has been concerned with the means of designing an intelligent agent to tackle real world problems. However, many real-life situations require intelligent systems not to work in isolation, but rather to be part of a broader environment, where they typically have to interact with other intelligent agents. In most dynamic domains where several agents coexist, the designer cannot foresee all situations that a single agent would encounter and thus, the agents should be able to learn new behaviors through interacting with their environment. This motivated the application of machine learning approaches on MAS so that the system does not have to be designed a priori since every single agent is given the possibility to learn and adapt to its environment. This combination with ML makes such systems suitable to a broader class of problems.

Multi-agent learning is a young but highly active and rapidly evolving field of research that have taken proactive strides to witness a number of considerable developments in a considerably short period of time. While so much progress has been achieved over the past decades, the field still suffers from several problems that impede its scalability and applicability to more complex and more realistic applications. Most problems directly result from the non-stationarity of the agents' environment and the high-dimensional state-action space. A large number of current researchers have tried to find work-around solutions to overcome these issues, through applying deep reinforcement learning and agent communication to MAL problems. The observed results are promising, along with the many research efforts dedicated to this topic nowadays.

Since multi-agent systems have always benefited from efforts in various disciplines (such as self-organizing biological systems and swarm intelligence, machine learning, communication and language, game theory), this field should follow a broader and more interdisciplinary approach, in order to be able to implement efficient multi-agent learning in complex applications.

References

[1] Michael H. Bowling and Manuela M. Veloso. "Multiagent learning using a variable learning rate". In: *Artif. Intell.* (2002), pp. 215–250.

[2] L. Busoniu, R. Babuska, and B. De Schutter. "A Comprehensive Survey of Multiagent Reinforcement Learning". In: *Transactions on Systems, Man, and Cybernetics* (2008), pp. 156–172.

[3] Y. Uny Cao, Alex S. Fukunaga, and Andrew Kahng. "Cooperative Mobile Robotics: Antecedents and Directions". In: *Auton. Robots* 4.1 (1997), pp. 7–27.

[4] Caroline Claus and Craig Boutilier. "The Dynamics of Reinforcement Learning in Cooperative Multiagent Systems". In: *Fifteenth National Conference on Artificial Intelligence/ Tenth Conference on Innovative Applications of Artificial Intelligence.* 1998, pp. 746–752.

[5] Jakob N. Foerster et al. "Stabilising Experience Replay for Deep Multi-Agent Reinforcement Learning". In: *CoRR* (2017).

[6] Jakob Foerster et al. "Learning to communicate with deep multi-agent reinforcement learning". In: *Advances in Neural Information Processing Systems.* 2016, pp. 2137–2145.

[7] P. Hernandez-Leal et al. "A Survey of Learning in Multiagent Environments: Dealing with Non-Stationarity". In: *arXiv e-prints* (2017).

[8] Junling Hu and Michael P. Wellman. "Nash Q-learning for General-sum Stochastic Games". In: *J. Mach. Learn. Res.* (2003), pp. 1039–1069.

[9] Michael Huhns and Gerhard Weiss. "Special Issue on Multiagent Learning". In: *Machine Learning* (1998), pp. 123–128.

[10] Nyree Lemmens and Karl Tuyls. "Stigmergic landmark foraging". In: *8th International Joint Conference on Autonomous Agents and Multiagent Systems.* 2009, pp. 497–504.

[11] Michael L. Littman. "Markov Games as a Framework for Multi-Agent Reinforcement Learning". In: *In Proceedings of the Eleventh International Conference on Machine Learning.* 1994, pp. 157–163.

[12] Volodymyr Mnih et al. "Asynchronous Methods for Deep Reinforcement Learning". In: *CoRR* (2016).

[13] Volodymyr Mnih et al. "Human-level control through deep reinforcement learning". In: *Nature* 518.7540 (2015), pp. 529–533.

[14] Volodymyr Mnih et al. "Playing Atari with Deep Reinforcement Learning". In: *CoRR* (2013).

[15] Liviu Panait and Sean Luke. "Cooperative Multi-Agent Learning: The State of the Art". In: *Autonomous Agents and Multi-Agent Systems* (2005), pp. 387–434.

[16] Liviu Panait, Karl Tuyls, and Sean Luke. "Theoretical Advantages of Lenient Learners: An Evolutionary Game Theoretic Perspective". In: *Journal of Machine Learning Research* (2008), pp. 423–457.

[17] Peng Peng et al. "Multiagent Bidirectionally-Coordinated Nets for Learning to Play StarCraft Combat Games". In: *CoRR* (2017).

[18] Yoav Shoham, Rob Powers, and Trond Grenager. "If Multi-agent Learning is the Answer, What is the Question?" In: *Artif. Intell.* (2007), pp. 365–377.

[19] Satinder P. Singh, Michael J. Kearns, and Yishay Mansour. "Nash Convergence of Gradient Dynamics in Iterated General-Sum Games". In: *CoRR* (2013), pp. 541–548.

[20] Sainbayar Sukhbaatar, arthur szlam arthur, and Rob Fergus. "Learning Multiagent Communication with Backpropagation". In: *Advances in Neural Information Processing Systems.* 2016, pp. 2244–2252.

[21] Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning.* 1st. MIT Press, 1998.

[22] Ardi Tampuu et al. "Multiagent cooperation and competition with deep reinforcement learning". In: *PLoS One* 12.4 (2017).

[23] Ming Tan. "Multi-agent reinforcement learning: Independent vs. cooperative agents". In: *10th International conference on machine learning.* 1993, pp. 330–337.

[24] Karl Tuyls and Gerhard Weiss. "Multiagent Learning: Basics, Challenges, and Prospects". In: *AI Magazine* (2012), pp. 41–52.

[25] Gerhard Weiss. *Multiagent systems: a modern approach to distributed artificial intelligence.* MIT Press, 1999.

[26] Mark Wilson. “AI is inventing languages humans can’t understand. Should we stop it”. In: *Fast Company* (2017).

[27] Martin Zinkevich. “Online Convex Programming and Generalized Infinitesimal Gradient Ascent”. In: *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*. 2003, pp. 928–935.

Chapter 5

Nature as a Source of Inspiration

Edgar Ricardo Chavez Rosas
Johannes Groll
Carla Pregel Hoderlein
Michael Schagginger

5.1 Abstract

The following article outlines the current development of two fields of research, namely DNA research and embodied intelligence. Both topics are strongly influenced by nature and offer a great potential to enable the implementation of artificial intelligent systems in the future. Therefore, this article presents not only important milestones and the current state-of-the-art but also gives a valuable insight into the ongoing trends as well as the future potential of the mentioned research.

5.2 Introduction

Humans have always tried to emulate nature by means of technology. For instance, the structure of a duck's skin has been synthetically reproduced to have waterproof coats. Also, observing bird flight helped us develop aeronautical systems. In the past decades, trying to understand and imitate the functionality of the human brain has become a vast field of research. Algorithms in the artificial neural networks are based on the brain's neurons. Learning how the brain works has rendered the possibility to implement the knowledge into artificially intelligent systems.

Another part of the human body that is currently being researched is DNA. Initiatives such as the Human Genome Project aim to sequence and map all of the genes[1]. The field of DNA Research is growing with the investigation of DNA as a possible storage medium for data.

Furthermore, research in fields like robotics and autonomous systems propose to adapt a new understanding of intelligence. The concept of embodied intelligence claims that the intelligence of a system is composed of the brain, the body, and the environment. This has opened up completely new fields of research and possible embodied-intelligence analysis. The main goal of this posture is to consolidate a legitimate image of the human cognitive system. It is then important to integrate bottom-up strategies, which represent the connection of the agent's actions to the environment, and top-down strategies, which care for the abstract thought. Some of the current projects dealing with embodied intelligence are being carried out in virtual worlds, where the environmental input is controlled and softened so that the solutions of the AI system can be better analyzed.

5.3 DNA Research

This section presents the important milestones that have been achieved in DNA research over the past two decades and provides an overview about current research topics.

[1] `https://www.genome.gov/10001772/all-about-the--human-genome-project-hgp/` (accessed April 26, 2019)

5.3.1 Past Milestones

In the past 20 years, several big explorations have been made in the area of DNA research. We have listed the most notable ones as follows:

Human Genome Project The Human Genome Project (HGP) was one of the biggest scientific investigations ever. For the first time, the human genome (the sum of human genes) was mapped[1]. Scientists were able to sequence and map all of the genes and henceforth construct a first draft of the human genome. Having began in 1990, this project was an international and collaborative research program. The results were published in 2003[1], which revealed that there are about 20,500 human genes that provide instructions for making proteins. These only account for roughly 1% of the entire genome. The remaining parts were labeled 'junk DNA'.

ENCODE Project The ENCODE Project is a worldwide effort that involves more than 400 scientists and 30 research groups. Planned as a follow-up to the HGP, the ENCODE Project (Encyclopedia of DNA Elements) was started in 2003. This is currently an on-going project that aims to build a comprehensive list of parts of the functional elements in the human genome. This includes elements acting at lower levels such as the protein and RNA levels. Furthermore, it investigates regulatory elements that control cells and circumstances in which a gene is active[2]. Later, it was discovered that more than 80% of what was once considered 'junk DNA'[2], actually serves a purpose in the gene expression. Gene expression describes the regulation of activity of particular genes. Researchers have been constantly working to gain further insights into the functioning of the human genome in order to understand its effect on human health.

1000 Genomes Project Being conducted from 2008 to 2015, the goal of this project was to create a public catalogue of human variation and genotype data. After the project ended, the International Genome Sample Resource (IGSR) was founded with three primary tasks[3]:

- Ensure future access to and usability of the 1000 Genomes reference data
- Incorporate additional published genomic data on the 1000 Genomes samples
- Expand the data collection to include new populations not represented in the 1000 Genomes Project.

During the 1000 Genomes Project, several developments in sequencing technology took place. Therefore, the cost of sequencing a single human genome was drastically reduced as shown in figure 5.1 on the following page. The overall costs for the sequencing during the HGP is estimated to be between 500M and 1,000M US-$. Nowadays, the cost is about 1.000 US-$.

5.3.2 Current Genomic-Research

Next stages of genomic research are directed towards deriving meaningful knowledge from the DNA sequence. Several research projects based on the HGP are being pursued worldwide. New technologies to study genes and DNA are being developed. Furthermore, the focus in DNA research has shifted further towards early detection, diagnosis, and treatment of disease. Scientists are trying to find variations in the DNA sequence among different people and determine their significance. Hence, allowing us to predict risks of specific diseases more accurately as well as individual response to certain medical treatments[4].

Also, genomes of other organisms are being sequenced in order to compare similarities. Among others, rats, cows, and chimpanzees have been successfully sequenced[4]. However, the complex 3-dimensional structures of proteins and their functions are still to be fully discovered.

5.3.3 CRISPR-Cas9

CRISPR-Cas9 is a revolutionary technology trying to introduce genome editing. CRISPR stands for clustered regularly inter-spaced short palindromic repeats. It is a system that guides a protein called Cas 9 to cut DNA strands [10]. The editing of the genome occurs in a three step process. First, the address is provided, where the faulty or malign DNA sequence can be found. In the second step, Cas 9 uses that information to find the desired sequence in the DNA. In the third and last step, the sequence is cut out from the DNA strand. From

[2] `https://www.encodeproject.org` (accessed April 26, 2019)
[3] `http://www.internationalgenome.org/about/` (accessed April 26, 2019)
[4] `https://ghr.nlm.nih.gov` (accessed April 26, 2019)

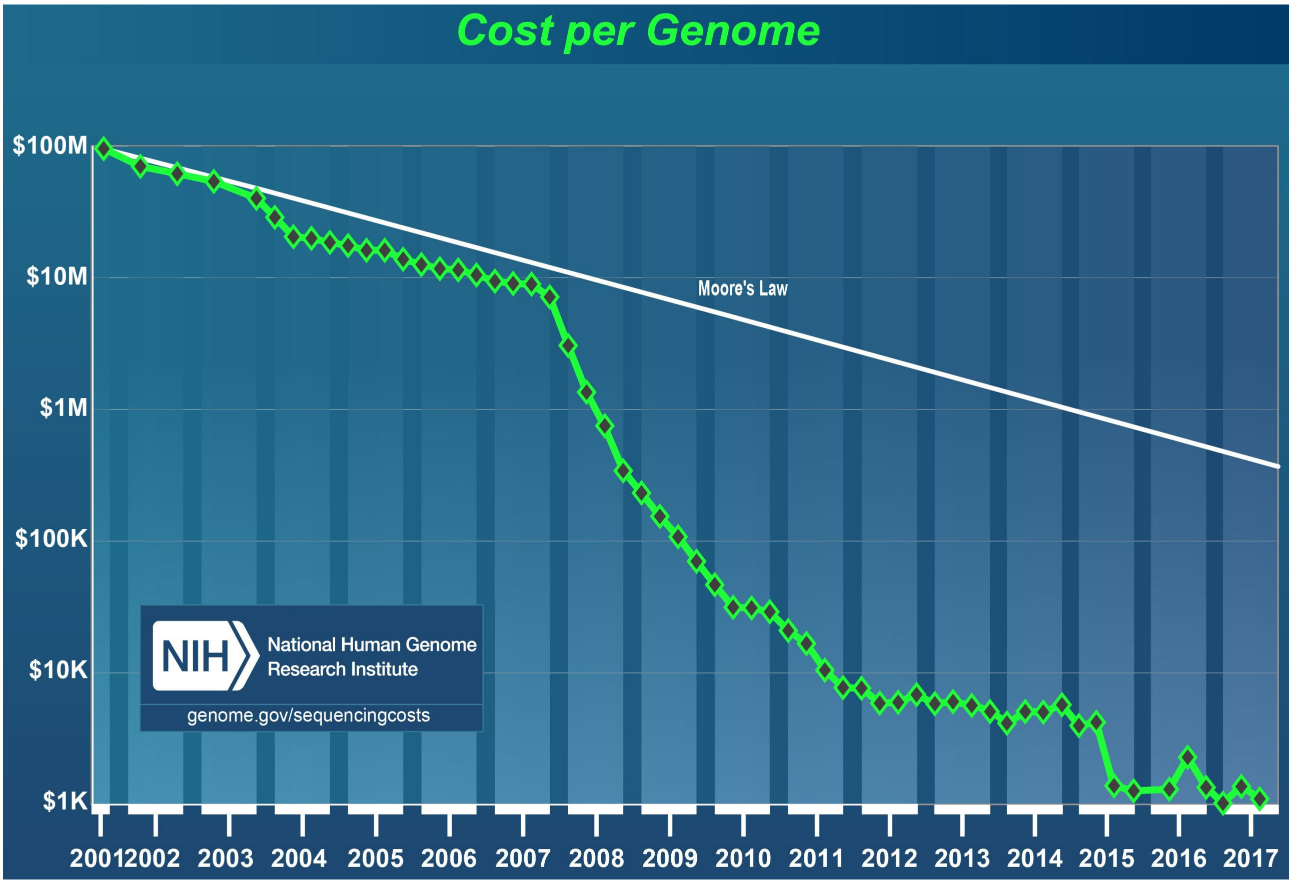


Figure 5.1: Cost per Genome - The cost of sequencing a human-sized genome; *source:* [5]

.

this point, there are two possibilities: either to let the DNA bind together with the missing sequence or to edit and introduce a new base sequence into the created hole. This technology can be used to treat genetic disorders caused by mutations in genes, among other applications.

5.3.4 DNA as a Digital Storage Device

Exponential increase of the global data collection with a prognosis of reaching the $3*10^{24}$ Bits in 2040 [14] is revolutionizing research fields like digital storage. This section will expand upon the possibility of using DNA to change the paradigm of data storage.

Silicon Silicon is the material used to manufacture microprocessors in computers. Nowadays, our data storage is being covered by silicon chips. This material has disadvantages on multiple levels. First of all, extremely harmful chemicals are used in the manufacturing process. Without a doubt, this is an important issue for our environment. In addition, these chemicals and their by products also compromise human health. One third of the workers in this industry are suffering from illnesses provoked by silicon pollution [14]. Every month, a typical factory produces 20 million silicon chips. Each chip requires ten times the amount of water. This means, that to produce 20 million silicon chips, 200 million gallons of water have to be used [14]. Furthermore, the reserves of silicon are limited. On top of this, as technology advances quickly, this kind of data storage has a limited lifespan. Also, after some time, it becomes unreliable.

Human DNA Every deployed human cell contains 23 pairs of chromosomes. These chromosomes are made up of proteins, called histones, supporting their structure. The strings of DNA are coiled many times around these proteins. The DNA is composed of 4 bases represented with the letters A, C, T and G and store data for different kinds of proteins. Sequences of DNA can store more information than binary storing mechanisms. For a character string of length X, the DNA with its four bases, A, T, C and G, has the possibility of 4X representations for it. In contrast, a binary system can only contain that information in 2X [14]. Moreover, a gram of DNA is formed by 10^{21} of these 4 bases and can therefore store 108 terabytes of binary data [14]. This means, that with just one kilogram of DNA we could store the whole $3*10^{24}$ Bits, which are predicted for 2040.

Chances of DNA As a Digital Storage Device Taking into account the disadvantages of silicon and the possibilities offered by human DNA, it has become of common interest to investigate whether DNA could be suited as a data storage medium. First of all, DNA is environmentally friendly and does not create pollutants in its production. This is very important for the reasons stated above. One very interesting point about DNA is that it is the most stable form of data storage in nature. In addition to this, it has a very high longevity and once the data is stored, it is very easy to reproduce. Via polymerase chain reactions techniques, the data can be easily replicated [14]. Nowadays, regarding data privacy discussions, DNA offers a very interesting positive aspect as it is possible to encrypt data. The data can be read as a code and is encrypted through the four bases. Furthermore, since DNA is invisible to humans, it cannot be destroyed as easily as a silicon chip [14]. Last but not least, DNA offers an aspect which is not trivial in today's world; DNA will be readable in the future. Since the reading and writing of the DNA occurs naturally in the human body, it can be stated that while humans exist, there will be mechanisms to decode data from DNA strands.

Challenges of DNA As a Digital Storage Device The major challenges that DNA faces as a digital storage device are the encoding of information and its information. This process is time consuming [14]. The synthesizing DNA, meaning writing or encoding DNA, takes even more time than other technologies used for data storage. In addition to this adversity, sequencing errors are also a problem for DNA replacing conventional methods. DNA is very susceptible to extreme conditions and creating mutations. To correct this, the human body created enzymatic proofreading mechanisms. These look for errors that may have occurred in the sequencing process and correct them. In the artificial sequencing of DNA, these mechanisms are unavailable [14], which means that data can be wrongly encoded and therefore, lost or corrupted.

5.3.5 Conclusion

The renowned mathematician and genome scientist, Nick Goldman once said,"As DNA is the basis of life on Earth, methods for working with it, storing it, and retrieving it will remain the subject of continual technological innovation" [14]. DNA research has made significant progress over the last 20 years. Mankind has been able

[5] `https://www.genome.gov/sequencingcostsdata/` (accessed April 26, 2019)

to sequence the human genome. When it comes to understanding this sequence, scientists have made proactive strides. Looking into the future, it can be expected that the human genome will be completely understood in an acceptable time period. On the other hand, DNA as a storage medium is a very interesting opportunity that has to overcome a lot of challenges in the near future, to be established. The main problems are in the en- and decoding of data. Regarding the cost, it can be expected that the cost of writing on DNA and reading will come down with the advancement of technology in the coming years.

5.4 Embodiment

Sometimes, just an abstract, off-line cognitive process with no relation to the exterior world does not suffice to call an agent intelligent. It is important to remember that the environment is constantly changing and any intelligent system that wants to perform any action to the world must have a prior input from it. If the input is not updated at the same time as the environment changes, then the action cannot be performed and the system results will be obsolete. For this reason, the intelligent agent must have a physical body that constantly interacts with the environment.

5.4.1 General Definition

The renowned Swiss author and professor Rolf Pfeifer has been one of the leading exponents of the embodiment intelligence hypothesis. He co-authored the book '*How the body shapes the way we think*', where he pointed out the importance of the connection between mind, body, and environment in order to have an actually intelligent agent [15].

Pfeifer makes several statements in his book that are worth discussing in the current section. In several chapters, he defines a necessary characteristic every intelligent agent must have: an ambivalence between *compliance* and *diversity*. The first term is used to describe the importance of the interaction between the agent and its niche. If the agent is capable enough to smartly exploit the resources given by its niche, then it is classified as intelligent. *Diversity* is also necessary for the agent's behaviour. If the agent acts monotonously, it will not last long, as the real world is always changing and thus, adaptation is necessary.

It may also be that the intelligent agent acts by following simple reflexes. These are called low-level reflexes and are present in common human behavior. In programming, the implementation of such behavior is called the 'Zen of robot programming'. Derived from the compliance-diversity concept, it is expected that the interaction with the world is complex and hard to encode in just a finite number of programming code. The robot Kismet proves this point otherwise [15]. Kismet's facial expressions change due to low-level reflexes, such as sound localization of a certain external stimulus, smooth pursuit (which is the act of slowly following moving objects with the eyes), and habituation. It may sound complex, but it is not. For this, it is important to remember about the frame of reference problem. An agent's act is intelligent, if the observer defines it. This could be despite the act being a mere cause-consequence effect unforeseen by the agent. The book concludes this statement at the end with "perhaps we are much more driven by low-level reflexes rather than our high-level rational thoughts" [15].

5.4.2 Analysis of Embodied Cognition

About a decade ago, professor Margaret Wilson published an article exposing a novel insight into possible different views of embodied cognition [17]. While it is true that these represent only simple perspectives of human intelligence, which can be insufficient to completely explain the cognitive process, they are also powerful analysis tools.

The first constellation states that cognition is situated in a certain time-space scenario, where task-relevant inputs and outputs flow from the agent to the environment and vice-versa. This perspective explains how humans work simultaneously on several tasks of different complexities. On the one hand, there are unconscious decisions, such as breathing, blinking and digesting. On the other hand, there are conscious decisions such as grasping a glass of water, solving an equation or even writing this article. This view, however, represents only a fraction of the human cognition work as it does not include the ability to create mental images or certain social skills, e.g. gossiping (it can be argued that this act may have had an important influence in the evolution of human intelligence [7]).

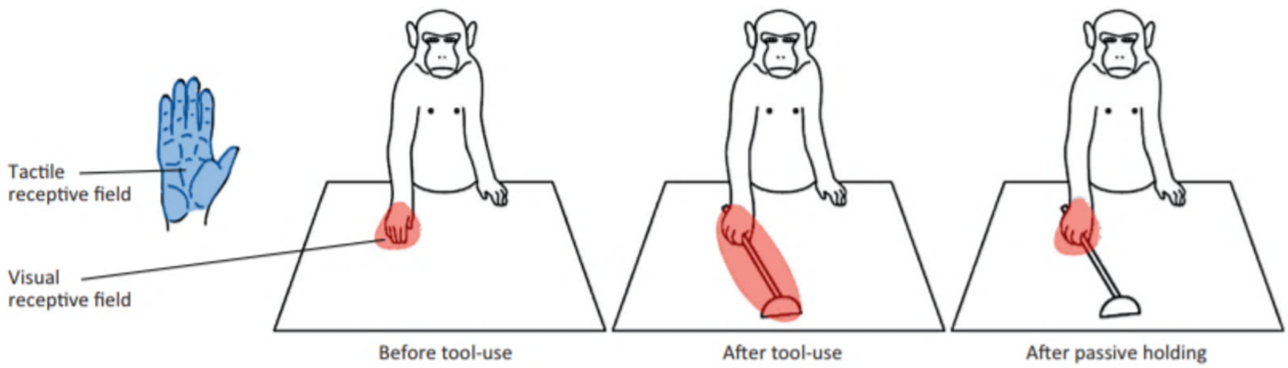


Figure 5.2: Changes in the visual receptive field of a monkey in different situations; *source: [5]*

The second view is built on top of the first one. The agent has to deal with time constraints in real-world scenarios. When the agent is under time pressure, it is forced to develop cheap and efficient tricks that manage to solve the problem. This is defined as the 'representational bottleneck' [17]. Although this may be seen as advantageous, it is important to remember that constant pressure on the agent results in a decay of efficiency. Furthermore, it is erroneous to generalize all realistic problems as time-constrained.

The third perspective can be explained using a simple example. In the well-known video game, Tetris, the player has to deal with falling blocks, which have to be configured to the correct position and orientation before they touch the bottom of the playing field. Most players define the blocks' properties while playing the video game rather than thinking of a specific attribute each time, i.e., humans off-load some of the cognitive work onto the environment in order to ease the cognitive process.

Embodied intelligence parts from the idea that truly intelligent agents have a constant connection to the environment. The next view of embodied cognition affirms that in order to properly analyze the cognitive process, it is necessary to include mind, body, and environment as a whole, creating a situational analysis. The body's actions, which are connected to the mind, result from the constant interaction with the external world. It can be pointed out, however, that it is not necessary to understand the causes in order to understand the agent's actions. It is more important to understand the fundamental principles of organization and function of the agent's cognitive processes.

The fifth view considers and analyses cognition in terms of 'their function in serving adaptive activity' [17]. Figure 5.2 shows an example of an experiment with a monkey. At the beginning of the experiment, the monkey's main visual focus was placed upon its hand. After it reaches the tool and uses it, the visual receptive field expands until it covers the whole tool. However, if the monkey can only grasp the object without using it, then the visual field remains in the hand.

Overall, this view of cognition can be seen as incomplete. As well as the first perspective, this represents only a fraction of the usual cognitive process. Mental images are not always created in response to physical stimuli or based on any specific physical attributes. For example, imagining a sunset can be a mental image with general attributes and does involve any physical action.

The last set of embodied cognition describes the processes that occur mainly in the mind, that is, off-line. Even in these situations, when the mind is not present in the situation, its activity is based on interactions between body and environment. In such cases, the body serves the mind with input (verbal and visuospatial information), which is later used to construct an internal simulation of external situations.

5.4.3 Recent Projects and Strategies

Following the embodied cognition theory, it is expected that the agent builds a simulation of the surroundings based on its own motor actions and the associated sensory simulations. Prior to this, there is no model of the environment that helps the agent to navigate or interact with the environment. This is called the Theory of Sensorimotor Contingencies (SMC) [12].

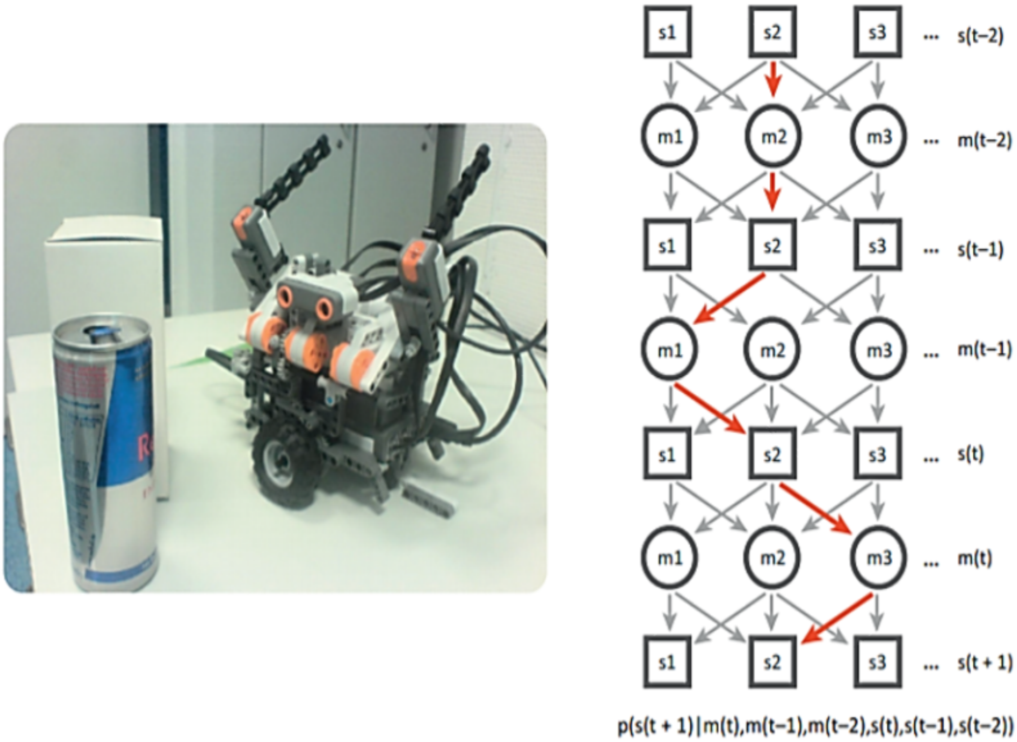


Figure 5.3: The goal of this robot is to learn to distinguish two different object classes by implementing different SMCs and Markov chains; *source: [5]*

Activation of different SMCs lead to a possible classification. Figure 5.3 shows a Lego-based robot that learns to distinguish two different object classes using SMCs and Markov chains. The Markov model helps to predict a current state based on a given current movement and a context. The context sums the past states of the robot.

The SMC theory suggests a bottom-up strategy which embraces the essence of embodied cognition. The counterpart is the well known top-down strategy. This approach suggests the iterative break-down of a complex problem into sub-problems, which can be easier to understand. While the bottom-up solution relies completely on the environment, the top-down redirects the cognitive work to the abstract thought. A more rich cognitive architecture integrates both strategies. The Distributed Adaptive Control for Embodied Machine Learning (DAC-EAI), as shown in figure 5.4 on the next page, proposes a novel integration of both strategies [13].

The direct interaction with the environment is placed on the bottom part of the figure 5.4 on the following page, described by the effectors and sensors of the agent. The abstract thought lies on the top part. The architecture also contains a heterogeneous and functional organization of several current AI strategies. For example, the DRL algorithm can be described using most of the Adaptive layer of the diagram. The bottom-up strategies, such as behaviour-based robotics, relies rather in the Somatic and Reactive layer.

Different strategies have been tested in recent years. Testing the early and novel embodied-based algorithms in the real world may be overwhelming. For this reason, most of the scientists opt to test them using simulations inside a 3D-virtual world. Inside this world, EC algorithms are tested using a virtual body. The open-source project MALMO, for example, is built on top of the famous game Minecraft and offers an AI experimentation platform [11]. Currently, there are bots capable of speaking to each other and with humans. Another ongoing project is DeepMind Lab [3]. Following this idea of a 3D virtual environment, the project allows users to test reinforcement and unsupervised learning algorithms. Current achievements include bots being capable of walking and learning how to interpret language in a semantic manner [8]. OpenAI Gym is also an open source toolkit for comparing and testing different RL algorithms [4].

The EC vision can be also seen as an interdisciplinary project, where not only robotic engineering plays an important role but also psychology and neuroscience. In the recent ROSSI project that ended on 2011, several scientists and engineers worked together in order to analyze the influence of the body (and thus, the personal

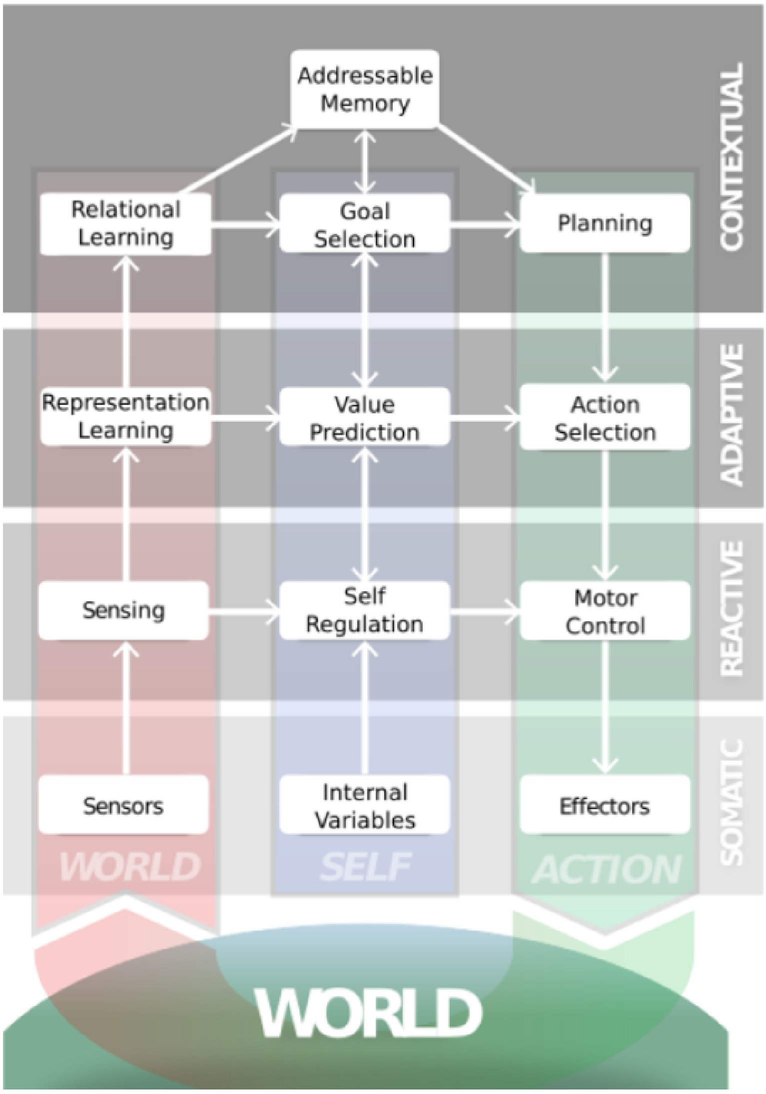


Figure 5.4: This architecture combines the top-down and bottom-up strategies into one novel architecture; *source: [13]*

perspective of the agent of the world) on the sensorimotor competences and the communication capacities of the agent [18]. The project analyzed the communication performance between both robots and robots, and robots and humans. In order to understand the content of the message, it is crucial that both agents have the same or at least similar knowledge on the context. In case that both agents' perspective is similar, the length of the message can be greatly reduced and the content, easily understood.

5.4.4 Conclusion

The definition of intelligence resides in the eye of the observer. The renowned Turing Test, taken sometimes naively as the ultimate test for intelligent agents, does not define an algorithm or mathematically proven theorem which delimits the borders of intelligence. The test is based on human intelligence and derives conclusions only meaningful to the humans observer. Now, the human constantly interacts with the environment and thus, it is expected that the intelligent agent does this too. Otherwise, the behavior would be seem odd and discarded by the human observer as unintelligent.

The bottom-up strategy relies entirely on this behavior, building a model of the world based on the agent's actions and its effects on the environment. This is what babies do. However, it is not possible to entirely remove the abstract thought from the cognitive analysis, as it constitutes a vast part of the human psyche. This corresponds to the top-down strategy. For this reason, a more rich and natural solution is to combine both strategies into one novel cognitive architecture. In order to test this solution and make simple but meaningful statements about its performance, the model must be tested first in a virtual environment. This way, the amount of interaction with the agent can be controlled and the AI process can be better understood.

5.5 Future Trends

We have observed that taking nature as a model to conceive former cognition is not always straightforward, which is understandable because nature had billions of years for trial-and-error procedures, while science has tried to do it in a couple of hundred years. A known example is the flying machine, which is modelled after the bird's anatomy. Between Da Vinci's sketches and the first practical aircraft of the Wright brothers, there were at least four hundred years of trial-and-error engineering.

In this section, based on the state-of-the-art, possible trends for future research will be presented and their potential impacts will be analyzed.

5.5.1 DNA Research

We expect that DNA research will become more significant to human lives in the next 20 years. In this section, we will focus on some trends that we believe will prove to be inevitable over the years as DNA research progresses. According to current trends, Genome Sequencing will be greatly improved and lead to new opportunities in several fields of our lives. Furthermore, we predict that DNA will overcome its challenges as a medium for data storage and its importance will be further increased.

Genome Sequencing As mentioned in section 5.3, Genome Sequencing has come a long way. Due to developments in sequencing technology, the HGP[1] costs have dropped considerably over the past decade. One trend that can be predicted is further improvement and therefore, a continuing decrease in costs. In [16], researchers have predicted costs for genome sequencing that have already been reached in less time. Although researchers are not completely sure about the future development and its speed, we make the assumption that the price will drop to a segment, where genome sequencing will be available to everyone. This will have immediate influences on several areas of our lives. Medicine could be personalized and be precisely adjusted to one's own genotype. Humans will be judged by their sequenced genome and future disease prediction may be improved. Tailored interventions could be made early on, such as adjustments in nutrition and life-style or environment. Creation of large databases, cf. to the 1000 Genome Project[3], could lead to easy comparability. Also, one could benefit from another person's data and medical history. On the contrary, this could also influence employment or insurance options on a large scale. This may raise ethical and privacy issues, but will be disregarded in this article.

Genome Editing With the CRISPR-Cas9, as mentioned in section 5.3.3, a technology to edit the human genome has been invented. One of the main visions of this technology is to cure diseases that have no cure till date. Diseases like cancer or inherited disorders such as sickle-cell anaemia and muscular dystrophy, where the root cause can be found in the person's DNA. The idea is to extract the sequence provoking a certain disease and to introduce the healthy DNA sequence instead. Trials on animals are already being carried out and soon enough there will be first experiments on humans. This technology could shift the paradigm of incurable genetic diseases. Furthermore, through projects such as the Encode Project (see section 5.3.1) that are trying to build a comprehensive list of parts of the functional elements in the human genome, diseases can be prevented by extracting the possible disease causing sequences even before the symptoms are exhibited. This technology also leads to other future trends. CRISPR-Cas9 can not only be used to cure and prevent diseases but also could contribute to the creation of super-babies. It is a possibility that when all the regions on the genome responsible for intelligence are found, one could start to edit them in-vitro, before inserting the fertilized egg into the mother. Hence, the intelligence of newborns could be increased beforehand. Apart from intelligence, parents could start deciding their child's eye-colour and the extent of their muscle development. Once we know which parts of the genome are responsible for specific features, the use cases of CRISPR-Cas 9 will be endless. For sure, this can bring along a positive impact to our society. However, it will raise many ethical discussions as well.

DNA Data Storage In section 5.3.4, we introduced the multitude of possibilities of DNA research and the challenges currently preventing DNA from becoming an acknowledged storage medium. Judging by the development over the past two decades, one can arrive at the conclusion that in acceptable time, its challenges will be overcome. If prices decrease and errors in decoding are improved, new possibilities will arise. Microsoft is already planning a DNA-based computer, that will make use of DNA data storage[6]. The device is expected to work as some kind of boutique application for storing only immensely important information.

5.5.2 Embodiment

Embodied cognition conceives intelligence as the result of a complex and dynamic interaction between agent and environment. Like a human infant, an agent perceives the surrounding environment and at the same time, has to learn how its interactions change and influence the world. Therefore, intelligent behavior cannot be assessed without considering the given environment.

Bridge Between Top-Down and Bottom-Up Learning Strategies Part of the abstract thought (top-down strategy) is based on the semantic meaning or conceptualization [1]. For language or even emotional attachment (say, social interactions), semantics play a complex but necessary role in the human cognition. Even in the most simple forms (e.g. how the ants communicate), a true intelligent agent has to perceive the direct or indirect meaning of objects or situations and act accordingly. Building an intelligent agent from scratch, letting it perceive the world in its own way supposes an underlying drawback compared to living beings. There are no past generations of agents, whose knowledge had been encoded and passed to the present. For example, it has proven that the human "DNA contains huge constraints on the form of language that enable us to learn the rest of grammar [after we have learned it from our parents as infants] very quickly." [2] As such, certain fundamental properties cannot be learned in just one lifetime via bottom-up strategies, but have to be processed in other ways. For this reason, it is necessary to build a bridge between the EC and abstract thought and test it. The DAC-EAI presented in the previous section, proposes a novel architecture that solves this problem. Nonetheless, it still has to be tested in the near future.

Evolve from Virtual to Real World Some current projects were presented in the last section. All of them evaluate several EC-hypothesis in 3D-virtual scenarios. However, in the future, it is expected to evolve to function in the real world as well. After the first hypothesis has been successfully tested (for example, communication between agents in Minecraft as part of Project Malmo), the next step is to test it using real sensors and effectors. Naturally, this means more complex and possibly distorted input from the environment, but it is necessary to prove the EC hypothesis in the real world. It is unknown if the EC algorithms will lead to abstract thought in the end (as it may happen with humans) or they will just work in a more simple, weak intelligence. Furthermore, it is important to test the embodied agents for possible interaction with each other and humans to acquire a better understanding of semantics [1]. The tests in the real world do not have to be constrained using humanoids. It is also possible to test EC hypothesis using various bodies, possessing different

[6] `https://bigthink.com/philip-perry/microsoft-plans-to-have-a-dna-based-computer-by-2020` (accessed April 26, 2019)

numbers and kinds of limbs. After all, the 'brain' or rather the cognitive processor of the agent has to be flexible enough to adapt itself in different situations. Using the bottom-up strategy, the adaptation should not suppose a problem. Another possible trend is to test different approaches using a rich, complex, and natural scenario. The DARPA Robotics Challenge was one of the most ambitious tests by these means. Filled with massive input and unclear task goals, the agents were often overwhelmed and could not achieve all the tasks. However, a human could have easily performed them. Therefore, it is important to overcome this difficulty and the agent must have diverse actions that comply to its environment in order to succeed.

Robot-Robot and Robot-Human Communication Based on the premise of project ROSSI, it is important that the agent is capable of analyzing the human and other robots' state as well as their range of capabilities in order to work together with them. Future research can also cover this scope of the EC hypothesis, where the human verbally dictates the action to the robot, without providing him with an example. The communication in a teamwork between robots may not be verbal or even explicit, although necessary. The robot has to act based on previous self-experiences in order to succeed in a natural environment.

5.6 Conclusion

For centuries, humans have tried to imitate nature and have used it as a source of inspiration for new technological accomplishments. Today, one of the greatest challenges is to understand and create artificial intelligence. To build such intelligent systems, once again, nature serves a great source of inspiration. Therefore, a comprehensive understanding of the emergence of intelligent behaviour and the functionality of our brain is crucial. Oftentimes, abstract ideas and concepts of different sciences like humanities or psychology and also our personal imagination as well as perception of intelligence have to be merged. Finally, these theories have to be matched with findings and discoveries in biology regarding our body, living beings in general, and our environment. In this article, we mentioned examples of DNA research and the concept EC to present the current state-of-the-art in these fields of research. Also, we have given an overview of the ongoing trends and how these results might help in the future.

Over the past 20 years DNA research has made significant progress. It is possible to sequence the human genome and the cost for this process is declining. The meaning of some DNA sequences is understood and scientists expect to further decode our genome in the future. Our DNA and the related working mechanism, storage and retrieval of information are the basis of human life. Researchers are convinced that the adoption of these processes can be the underlying structure of future of technological innovations towards artificial intelligence [9].

The concept of embodiment is actually based on the 'mind-body problem', concerning the relationship between the human mind and body. Having been discussed since the time of Greek philosophers, this topic has become relevant again, with evolving technology in the field of robotics. Conventional robotic systems strictly rely on one central control system and often experience limitations regarding their ability to interact with a complex and real-life environment. Today's researchers opine that complex cognitive abilities only emerge when forward-planning mechanisms become decoupled from the actual system prediction. Therefore, there are several approaches in the field of robotics which try to split-up and distribute cognitive abilities and information processing between a central control system and peripheral parts of a robot to enable intelligent interactions with an environment [6].

This article has picked out two fields of research which demonstrate how nature might inspire researchers to create and re-engineer real intelligence. As done in countless cases before, the approach of analyzing and understanding natural phenomena might once again trigger new technological innovations. Although the presented state-of-the-art research is still in an early stage, previous scientific achievements substantiate the validity of this approach.

References

[1] Michael L Anderson. "Embodied cognition: a field guide". In: *Computing Reviews* 45.5 (2004).

[2] Eric B Baum. *What is thought?* MIT press, 2004.

[3] Charles Beattie et al. "Deepmind lab". In: *arXiv preprint arXiv:1612.03801* (2016).

[4] Greg Brockman et al. "Openai gym". In: *arXiv preprint arXiv:1606.01540* (2016).

[5] Andreas K Engel et al. "Where's the action? The pragmatic turn in cognitive science". In: *Trends in cognitive sciences* 17.5 (2013), pp. 202–209.

[6] Hoffman Guy. "Embodied Cognition for Autonomous Interactive Robots". In: *Topics in Cognitive Science* 4.4 (2012), pp. 759–772.

[7] Y. N. Harari. *Sapiens: A Brief History of Humankind*. Harvill Secker, 2014.

[8] Karl Moritz Hermann et al. "Grounded language learning in a simulated 3D world". In: *arXiv preprint arXiv:1706.06551* (2017).

[9] Matej Hoffmann and Rolf Pfeifer. "The implications of embodiment for behavior and cognition: animal and robotic case studies". In: *CoRR* (2012).

[10] Marten H Hofker, Jingyuan Fu, and Cisca Wijmenga. "The genome revolution and its role in understanding complex diseases". In: *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1842.10 (2014), pp. 1889–1895.

[11] Matthew Johnson et al. "The Malmo Platform for Artificial Intelligence Experimentation". In: *International Joint Conference on Artificial Intelligence*. 2016, pp. 4246–4247.

[12] P. König et al. "Embodied cognition". In: *6th International Conference on Brain-Computer Interface*. 2018, pp. 1–4.

[13] C. Moulin-Frier et al. "Embodied artificial intelligence through distributed adaptive control: An integrated framework". In: *Joint IEEE International Conference on Development and Learning and Epigenetic Robotics*. 2017, pp. 324–330.

[14] Darshan Panda et al. "DNA as a digital information storage device: hope or hype?" In: *3 Biotech* 8.5 (2018).

[15] Rolf Pfeifer and Josh Bongard. *How the body shapes the way we think: a new view of intelligence*. MIT press, 2006.

[16] Deirdre Weymann et al. "The cost and cost trajectory of whole-genome analysis guiding treatment of patients with advanced cancers". In: *Molecular genetics & genomic medicine* 5.3 (2017), pp. 251–260.

[17] Margaret Wilson. "Six views of embodied cognition". In: *Psychonomic bulletin & review* 9.4 (2002), pp. 625–636.

[18] Tom Ziemke et al. "ROSSI: Emergence of communication in Robots through Sensorimotor and Social Interaction". conference poster. 2008.

Chapter 6

The Unchartered Waters of Strong AI

Sven Gronauer
Fabian Kreutmayr
Dominik Krockenberger
Alex Ostertag
Killian Pfeiffer

6.1 Abstract

Artificial General Intelligence (AGI) investigates the question of building intelligent machines similarly complex to the human being. Current approaches mostly consider the narrow view of the research domain with narrow artificial intelligence, but overlook the global prospective of strong artificial intelligence.

In this work, we present past developments in the domain of AGI, examine current challenges, investigate open research questions, and outline future trends with a special focus on the next two decades.

6.2 Introduction

"Only a small community has concentrated on general intelligence. No one has tried to make a thinking machine. The bottom line is that we really haven't progressed too far towards a truly intelligent machine. We have collections of dumb specialists in small domains; the true majesty of general intelligence still awaits our attack. We have got to get back to the deepest questions of AI and general intelligence..."

– Marvin Minsky (2000), as interviewed in Hal's Legacy

All currently available artificial intelligence (AI) systems constitute a narrow AI landscape. This means that they are designed, trained, and optimized to solve a single, very specific task. In such specialized domains, algorithms outperform humans in their established skill-set. However, current AIs are not able to extend their capabilities and generalize to new domains [13]. Rather, they appear to be unintelligent in areas that lie outside their design template. Highly specialized expert systems are called weak or narrow AIs.

Nonetheless, there is an ongoing effort to develop a general or strong artificial intelligence. Such systems should not just perform well on one domain; their functionality should at least be as general as humans. In the long run, mankind wants to design a general AI to have a mind, consciousness, and the abilities to abstract and learn. Although there is no known AI system available today that fulfills these requirements, on-going research has revealed astonishing results in AI performance in the past years. We want to introduce interesting topics in the following pages and explain how researchers all over the world try to overcome the stage of weak AI.

6.3 Past Milestones

Throughout the past two decades, there has been steady progress in the area of AI research. Very often, these have been demonstrated by machines beating humans in direct competitions (figure 6.1 on the next page shows

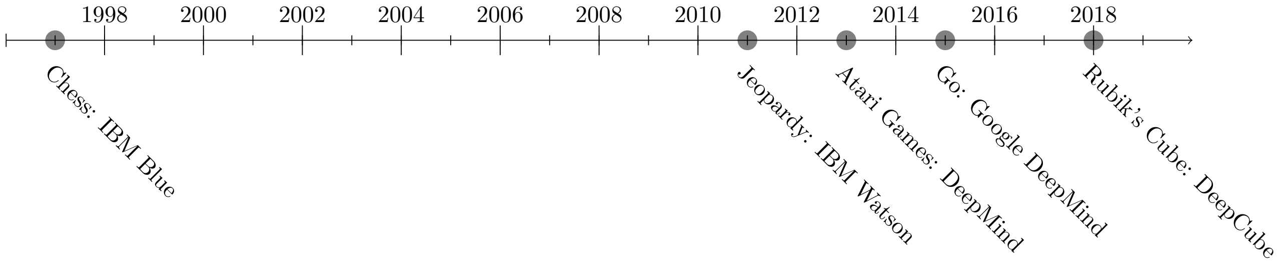


Figure 6.1: Past AI milestones; *source: own image*

an overview).

One of the first big successes from machines against humans was IBM's chess computer - Deep Blue [6]. While the former world champion of chess, Garry Kasparov, was able to beat the computer in a first match in 1996 by a score of 4:2, the artificial intelligence system won the rematch in 1997 with 3.5:2.5. According to Kasparov, some moves of Deep Blue showed deep intelligence and creativity [22]. This was perhaps the first time that an AI was assigned such attributes. Nevertheless, we know that this success was very much related to the combination of a large amount of data and computing resources.

Another big success was achieved by IBM's Watson in the year 2011. Watson is a specially trained AI for processing natural language and competed against two former quiz champions in Jeopardy - an American quiz show in question and answer style. To their astonishment, Watson outperformed the human competitors by 77,147 USD over 24,000 USD and 21,600 USD as a contestant, live on TV and in real time[1] [12]. IBM's Watson is technically based on an idea called DeepQA network. For building Watson, the researchers combined more than 100 techniques for analyzing natural language, content filtering, hypotheses generation, and evidence scoring. As emphasized in [12], not a particular technique is crucial, moreover, the combination of many techniques yields the successful results.

Two years later, in 2013, the company DeepMind showed that it is possible to "learn successfully control policies directly from high-dimensional sensory input" [30]. To achieve this, they used a variant of Q-learning, which is a reinforcement learning (RL) algorithm, to train a deep learning (DL) model. Hereby, the convolutional neural network (CNN) uses pixels as input, extracts the important features and outputs a value function that estimates future rewards. From this value function, the action to take in the next time step can be concluded. The capability of this approach, called deep reinforcement learning (DRL), got evaluated by learning strategies for seven different Atari 2600 games. Without adapting the algorithm's structure, DRL not only outperformed existing approaches but also exceeded human performance in three of the games.

The results of this experiment laid the groundwork for one of the most respected successes in AI research since Deep Blue. So, DeepMind introduced a system called AlphaGo in 2015 [38]. As the name suggests, this AI was designed to play the famous Chinese board game, Go. This was assumed to be the most challenging classic board game to master, since a standard 19×19 board allows to have about 2.08×10^{170} states. In contrast to this, the state space of chess has 10^{43} states and is, thus, much smaller (in comparison, the universe is said to consist of $\sim 10^{80}$ atoms). Therefore, it was a remarkable success when AlphaGo was able to win a full match against Fan Hui, the European Go Champion at the time.

Recently, the DeepCube algorithm managed to solve a 3x3 Rubik's Cube completely on its own, by using a technique called autodidactic iteration [27]. The algorithm works back from the fully solved cube to a similar configuration using a self defined reward function and a search tree while learning which moves are generally beneficial. It is claimed that the "algorithm is able to solve 100% of randomly scrambled cubes while achieving a median solve length of 30 moves — less than or equal to solvers that employ human domain knowledge" [27].

Looking at these milestones, some interesting developments can be observed. While DeepBlue solved chess by using a smart implementation of a search tree, this method is almost unusable for more advanced games and problems. The number of states is very often too big to efficiently implement a search tree solution, even if the growth of computational power by Moore's Law is considered. Thus, IBM's Watson used a quite complex and

[1] `https://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy` (accessed April 26, 2019)

modular structure to solve Jeopardy. This certain architecture is called DeepQA and combines several machine learning techniques to solve different tasks that occur throughout the game. Even if DeepQA already makes use of some neural networks, the vast usage of deep neural networks came up later, driven for example, by the success of DeepMind. Especially, the combination of reinforcement learning and deep learning is perhaps the biggest enhancement of the AI community in the past years. Hereby, it should be mentioned that the structure of these DRL approaches is quite flexible and therefore, the method is not restricted to solve Atari games or a board game like Go. DRL can also be applied to continuous control tasks like cartpole swing-up and legged locomotion [25]. Moreover, today's methods can not only solve more advanced problems but also have a broader application area than their predecessors.

Comparing past and state-of-the-art learning approaches, there is one major similarity worth being mentioned: All these intelligent systems are based on an incredible amount of data. The dependence on comprehensive databases has always been key to create an AI. Of course, the algorithms have gotten more efficient and powerful over time, leading to the capability of being able to handle a greater amount of data. Without these large amounts of data, neither DeepBlue nor AlphaGo would have been able to beat humans in some of the most challenging games.

Overall, it can be stated that there have been a lot of interesting achievements in AI research during the past years. Especially, the development of more advanced methods to model and train deep neural networks has been an accelerator for AI research. Thus, besides these big successes, there has been a lot more significant work in fields like vision performance and language processing [37]. Although these AIs are able to achieve superhuman performance in a very specific field, there has been no system with a general intelligence up to this point. So, all of the systems mentioned in this section are only qualified as weak AIs that can act intelligently only in their small area of application.

6.4 Current Research

As mentioned before, there are no AI systems yet that are qualified as an artificial general intelligence (AGI). Instead, many expert systems exist that act very well in one narrow, specific area of application; these systems are not applicable for tasks outside their design purpose. Consequently, the development of an AGI remains one of the main goals of current AI research. After a brief introduction of the differences between weak and strong (or general) AI, there will be an overview of on-going research in this section.

6.4.1 Difference between Weak and Strong AI

Since a long time, philosophers and researchers have been trying to solve the big questions related to AI: Can machines think? Do they have a mind and thoughts? According to the terminology in the literature [35], machines that act as if they were intelligent are counted as weak AI, whereas machines that are actually thinking and not just pretending are called strong AI.

6.4.2 Theoretical Approaches

In order to create a solid base for future developments, researchers and philosophers have thought about theoretical frameworks for artificial general intelligence. Frameworks grasp topics like requirements, properties or the building of artificial beings. The following paragraphs introduce some of the theoretical key concepts that should be employed to AGI systems.

Core Requirements for AGI

General intelligence, as described above, demands a number of irreducible features and capabilities. In order to proactively accumulate knowledge from various complex and dynamic environments, it requires [15]:

- Senses to obtain features from 'the world' (virtual or real)
- Coherent means for storing knowledge
- Adaptive output/actuation mechanisms (both static and dynamic)

These are core requirements to interact with an environment in a proper way. What purpose does an artificial agent have, if it is just intelligent but has no abilities to interact with its environment? Senses are necessary to perceive the current state of the environment, whereas actuation mechanisms allow agents to actively change the state of the environment.

Properties of AGI

Beside the core requirements for a general agent, there are many other properties that an agent must fulfill to be accounted as a thinking and general machine [15]:

- the ability to solve general problems in a non-domain-restricted way, in the same sense as that of a human
- the ability to solve problems in a particular domain and context with a particular efficiency
- the ability to use its generalized and specialized intelligence capabilities in a unified way
- the ability to learn from its environment, other intelligent systems, and teachers
- the ability to become better at solving new types of problems with experience

How to Build and Educate AGI

Another open question: How to master the education of a general intelligence? Currently, conventional algorithms learn the appreciated task from scratch. In contrast, algorithms for AGI have to master complex tasks and must generalize well beyond a specialized domain. To accomplish this generalization, new learning techniques can be applied to the agent. One instance for learning complex tasks is curriculum learning [3]. The learner begins with coping a relatively simple task, which gradually evolves in complexity over time. Even though, the complexity of the agent's tasks becomes very high, the agent can exploit its knowledge from previous, simpler learning tasks. The agent's problem-solving abilities gradually increase along with the complexity of new tasks. Hence, an agent that is capable of mastering complex tasks must feature several of the following key points [34]:

- Exhibit a very large repertoire of ideally general skills
- Skills as building blocks towards general AI
- Intrinsic (apriori knowledge) and learned skills
- General agent architecture and an optimal learning curriculum

A conceptual approach for building a general agent is shown in figure 6.2 on the following page: an optimization process for creating general artificial intelligence. A framework defines principles, ideas, and methodologies that are combined in different ways to form skills. Skills can be ordered hierarchically or in a chain to output a road map. In particular, the road map describes the set of skills that are either learned by experience or in prior coded into the agent (intrinsic skills). Many road maps are evaluated in the entire environment (AI landscape) to retrieve feedback of its performance. This feedback is then used to build a better set of skills for the next learning curriculum.

6.5 Current Challenges

Past milestones in the field of AI belong to the definition of weak artificial intelligence. Generating more sophisticated and general AIs requires new approaches. To overcome the stage of weak intelligence, we have to face new challenges, namely:

- **Agents:** According to strong AI, agents are not only able to perform well on a certain domain but also transfer their current knowledge to new domains and generalize well. Here, limited resources reveal the major difficulty in knowledge representation and decision-making taken by a single agent; either the amount of memory is restricted or the available computational power is limited. An agent has to be carefully designed in order to accomplish the trade-off between restricted resources and fast decision making in real-time. In addition, agents are not programmed with a prior goal. The agents' goals are adaptive and change over time to cater to the current needs of the agent.

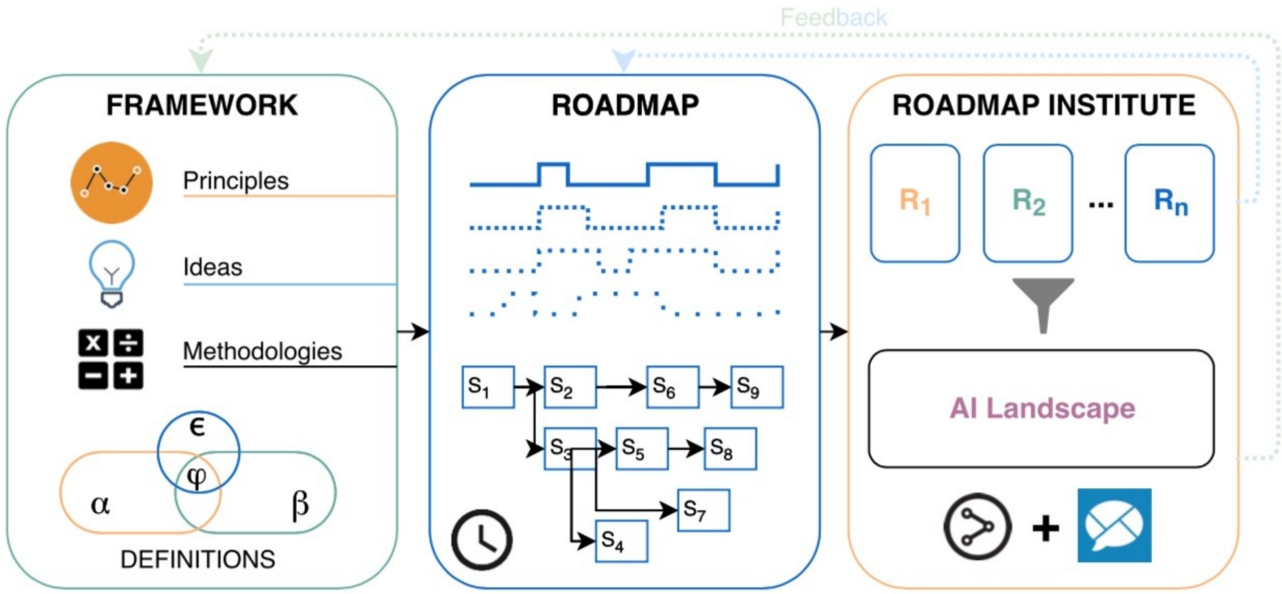


Figure 6.2: Conceptual approach for creating AGI; *source: [34]*

- **Environments:** In order to develop intelligence that is far superior compared to yet known intelligence, very complex environments have to be created. Several key properties are deduced from the requirement that these environments must be beyond the capabilities of current AI. The first properties of such complex environments are dynamic and open: agents interact with the environment and affect the state of the environment as well as the performance of other acting agents.
- **Tasks:** The environment does not only provide tasks for which the agent was trained. The environment challenges the agent also with tasks that are completely new to it. Mastering these novel tasks forces the agent to transfer its knowledge. In addition, the tasks must be diverse and complex. This implies that agents must consider their decision-making very well and make use of their general knowledge.

6.5.1 Critical View on Deep Learning

While Deep Learning is the reason for most recent milestones in AI technology (see figure 6.1 on page 82), a magnitude of problems arises if one tries to build a general AI with it. First, Deep Learning is very data hungry. Since it is essentially a statistical analysis of a preferably huge amount of data, Deep Learning depends heavily on the availability of the said data. This results in a problem if no valid training data can be provided, which is especially prominent in the case of abstract concepts like ‘democracy’ [26].

These abstract concepts provide another challenge for Deep Learning Nets, since the nets right now can’t really grasp concepts of for example objects like a cup. If an image recognition algorithm was trained solely with cups shown from a specific side, it would probably not recognize a cup if it was turned upside down. The net only recognizes similarities in the image, but doesn’t really understand the concept of the object “cup” and can therefore not extrapolate the seen information into the three-dimensional space[2]. Another problem lies in the recognition of hierarchy. For example, the hierarchy of language. When we as humans read a compound sentence, it is intuitively clear to us which parts of the sentence contain the main information and which parts add extra information. In case of a neural net, this intuition is not present. On the contrary, since the sentence is basically just a flat hierarchy of words, extracting the hierarchy, while manageable with enough effort, is still a complicated task for the machine [26].

Furthermore, it is especially complicated to program predefined knowledge into a neural net. Currently, most neural nets are purposely trained without predefined knowledge, since the machines should learn by themselves. However, even if this bias was set aside, the task of programming that specific knowledge into the machine is not that simple. Since we do not fully understand the inner workings of nets at this point, there still has not been

[2] `https://www.nytimes.com/2017/11/28/technology/artificial-intelligence-research-toronto.html` (accessed April 26, 2019)

1+1 = 2	
2+1 = 3	1+2 = 3
3+1 = 4	1+3 = 4
…	…
9+1 = 10	1+9 = 10

Figure 6.3: Addition with digit 1

1+2 = 3	2+1 = 3
2+2 = 4	
3+2 = 5	2+3 = 5
…	…
9+2 = 11	2+9 = 11

Figure 6.4: Addition with digit 2

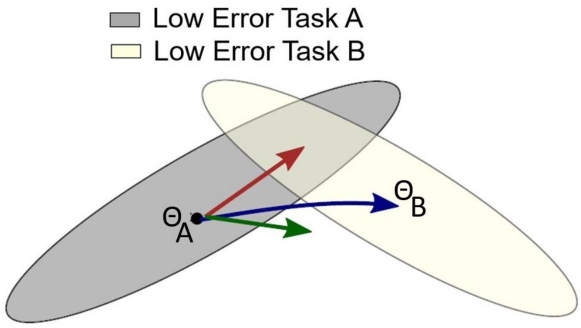


Figure 6.5: Overview Catastrophic Forgetting; *source: [23]*

a clear working method of alternating net weights to achieve predefined knowledge [26]. Additionally, without preset experience or intuition, Deep Learning algorithms struggle to differentiate between correlation and causality. There are tons of examples of data that seem correlated, but to assume a causality there would be ridiculous.

Lastly, probably the most challenging aspect of Deep Neural Nets is their virtual appearance: a black box that we struggle to look inside. This makes it hard to justify an algorithm's decisions since we can not say for sure what its thought process was. This lack of explainability is especially dangerous while applying Deep Learning in sensitive fields like the Military or Investment[3]. In addition, since summer 2018, operating Article 22 of the EU GDPR states – "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her" [17]. This will additionally increase the significance for comprehensible algorithms in the future. The lack of transparency also makes Deep Neural Nets hard to work with because it aggravates debugging and designing, apart from making the guarantee of function complicated, which is already a problem in autonomous driving.

6.5.2 Catastrophic Forgetting

The ability to learn sequentially is one key element for strong AI. Learning sequentially means to acquire one skill after another without losing the previously learned skill. Although this may be quite natural for humans, it is a big challenge for neural networks in AI. In 1989, McCloskey and Cohen first stated the term catastrophic inference or catastrophic forgetting [28].

Problem Description

At first, they trained a neural network the addition of the digit 1 with all other digits which are 17 combinations in total, see figure 6.3. To approximate these 17 additions is an easy task for a neural network, so the training for the first sequence was no problem. Subsequently, they used the same neural network and trained only the addition with the digit 2 in the same manner until the results were satisfying (see figure 6.4).

Thirdly, after training 2 was finished, they tested the ability of addition with the digit 1 again and found that this skill was completely lost. This is the effect of catastrophic forgetting. Figure 6.5 visualizes the effect of catastrophic forgetting in the parameter space.

θ_A are the learned parameters for task A in the example above – the addition with the digit 1. The grey area represents good parameters for task A and the beige area represents good parameters for task B. The blue arrow shows the development of the training for task B without restrictions. One would end up in an area where the

[3]https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (accessed April 26, 2019)

skill for task A is completely forgotten. Taking the mean of θ_A and θ_B would lead to parameters where neither task A nor task B are solvable, which is depicted by the green arrow. Only the parameters in the intersection between the grey and the beige area cope with both tasks. Therefore, during training of task B, one has to change the parameters in the direction of the red arrow. A conventional training set up for a neural network is not able to do this and has the following drawbacks with respect to sequential learning:

- In order to solve different tasks, the training data has to be s uncorrelated as possible. Therefore, the training data has to be complete right from the start of the training. Also, leading to a high storage requirement
- The expandability of new skills later is not possible without forgetting others
- Non-sequential learning does not resemble the way humans learn

However, there are several approaches to reduce the effect of catastrophic forgetting, which are presented in the next section.

Overcoming Catastrophic Forgetting

Elastic Weight Consolidation (EWC) is a way to acquire new skills for a fully connected neural network without forgetting previously learned ones [23]. The idea is to first identify the importance of weights associated with the previously learned tasks with a probabilistic approach. Then, according to their importance, the learning rate for each weight is calculated. This can be imagined as a spring, anchoring in the previous solution. The restoring force of the spring is high in the directions of important weights for the previously learned tasks and low in the directions of unimportant weights. Now, under these constraints, the new skill is learned and the algorithm tries to find an intersection for all tasks (see figure 6.5 on the previous page).

EWC was tested in an RL environment by sequentially learning 10 different Atari games and compared with gradient descent. While the neural network with gradient descent was only able to play one game, the EWC algorithm prevented catastrophic forgetting and could play all 10 games. However, 10 separate deep-learning-agents, which are only trained for their specific games, outperform the EWC-trained agent. Moreover, EWC-based learning also enabled to learn hand-written digits, sequentially [23].

Compared to EWC which was a rather mathematical approach to avoid catastrophic forgetting, a lot of research is going on in biologically inspired learning. For example, in [9] researchers introduced costs for connections between neurons. This idea was transferred from the human role model to the neural network as maintaining and building those connections cost resources from the human body. Compared to EWC, this approach directly influences the structure of neural networks. This approach was tested in an artificial biological environment. The setup was as follows: There are animals which try to survive as long as possible. A year has two seasons: winter and summer. During each season, the animals see different kinds of food: some are poisonous and some healthy. In order to test the effect of catastrophic forgetting, they checked if the animals are able to remember healthy food from last year or if they eat food randomly. With use of the costs on the connections, the neuronal network developed concentrated modules implicitly. Each module was responsible for a specific task. In the example above, one module found healthy food during winter while the other module could find healthy food during summer. Instead of changing weights in the hidden layer, this approach directly changes connections between neurons and creates new modules to learn a new skill. However, this approach struggles to acquire skills which have the same input neurons.

Till date, we wait for the big breakthrough to avoid catastrophic forgetting. Current approaches may work fine in their specific domains, but are far from being applicable to real world scenarios.

6.6 State-of-the-Art

Artificial Intelligence has come a long way in recent years. Projects like AlphaGo have shown that with state of the art Deep Learning Neural Nets, problems that were believed to be unsolvable by a machine, until very recently , can now be overcome. Also, the use cases for AI have grown enormously in the last decade, from video processing to first prototypes of autonomous vehicles. However, as already stated, all these milestones have been achieved with weak AIs. Despite recent progress, great challenges like catastrophic forgetting and the lack of transparency in Deep Neural Nets, still keep researchers from creating strong AI. Experts in the field,

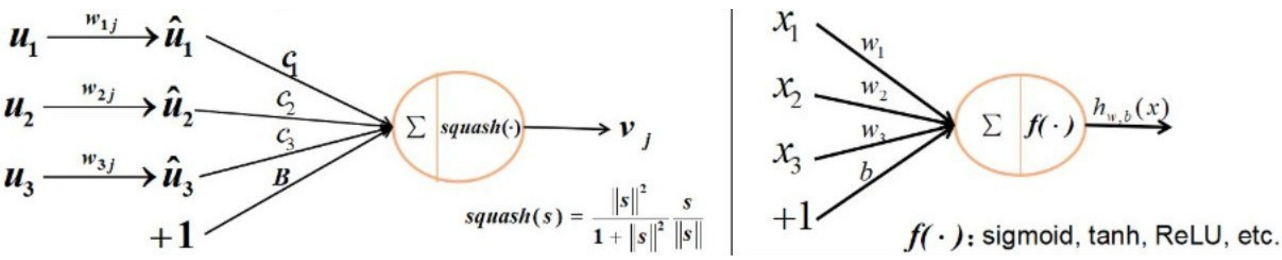


Figure 6.6: Analogy between capsules (left) and neurons (right); *source: https://goo.gl/SVKjwZ*

like Geoffrey Hinton, have started looking for new and improved approaches to AI[2]. For now, the thought of a real, human-like or, as some fear, super-human artificial intelligence, remains part of science fiction.

6.7 Future Trends

We expect that general AI will, for the most part, only play a role in theoretical concerns during the next two decades. Philosophers and researchers will rack their brains about theoretical methodologies, ethical and legal questions rather than about concrete concepts and approaches for creating a thinking strong AI. However, the real technological progress will take place in specialized domains of AI. Over the next years, AI is going to continue improving steadily. In this chapter, we will examine future trends in AI research and development with a special focus on the next two decades; wanting to reveal insights into the encountering domains.

6.7.1 Improvement of Deep Learning

The most newsworthy research trend in recent years is the continuous improvement of deep learning, which we suspect will carry on over the foreseeable future, either through inventing new methods like capsule networks [36] or improving contemporary methods with new techniques.

Improvement of Deep Learning Techniques

Contemporary deep neural networks such as CNNs and RCNNs have shown tremendous success in domains like image recognition, natural language processing, and reinforcement learning. By employing advanced techniques, deep learning networks are able to perform even better on their specialized domain. Such techniques include dropout [39], residual blocks [18] or long short-term memory [20]. In fact, contemporary deep neural networks can be pushed even further beyond their limits in the future. We expect that new techniques will be applied to reach the full potential of current deep learning methods.

Capsule Networks: A new Method for Deep Learning

As already mentioned in section 6.5.1, deep learning has several drawbacks. One drawback is that deep learning is rather a brute force method which does not incorporate spatial relations between objects. For instance, in facial recognition, the pure existence of a nose, ears, eyes, and mouth is a high indicator for deep learning to detect a face regardless of its components' spatial order. Another drawback is that training data has to comprise all possible incidences to work reliably. Thus, the information of the training data is not used efficiently. So, deep learning's approach is rather focused on mass of data instead of proper information exploitation.

A new method introduced by Hinton et al. [36] is capsule networks, which deal with the problems mentioned above. Capsules represent modules that detect present features or objects in the input by expressing the probability of that object by an activity vector. They consider spatial relations between different objects to classify. By connecting capsules in a hierarchical manner, spatial relations between objects are maintained throughout the whole network. In contrast, these precise positions are lost in conventional CNNs because of the max pooling operation [36]. In capsule networks, the presence of a nose, ear, eyes, and mouth along with their spatial relation is considered to detect a face. In addition, information about the spatial relations of objects (rotation, relative distance) are used for reasoning. Capsule networks can be regarded as a new method that extends the idea of conventional CNNs with vectorized neurons.

Figure 6.6 on the preceding page shows the analogy between capsules on the left and neurons on the right. The input vectors u_i are outputs of capsules from the previous layer. Let us assume that u_1 depicts the detection of the mouth. Each input u_i is multiplied with the corresponding weight matrix W_{ij} which can be interpreted as the relation between mouth and face. Thus, $\widehat{u}_1$ represents the position of the face given the position of the mouth. Then each vector $\widehat{u}_i$ is weighted with a scalar c_i which describes the importance of $\widehat{u}_i$ to detect a face. All these vectors are then summed up and put into a squashing function which leads to the output vector v_j of the capsule. Due to the squashing function, the length of v_j ranges from 0 to 1 and correlates with the probability of a detected face. The elements of v_j encode the internal states of the detected face such as pose, lighting, and deformation.

The special thing about this approach is – instead of aiming at invariance to rotations and positions by using max pooling – capsule networks focus on equivariance. Equivariance means that the internal states of v_j or the activity of the capsules adapts, if the face moves or rotates in the image but the length (probability) of v_j stays the same. This allows capsule networks to even extrapolate relations between objects to new viewpoints. The weights in this architecture are updated with a combination of the back-propagation algorithm and a dynamic routing algorithm between capsules, explained in [36] or [19]. The dynamic routing algorithm plays a central role because it implicitly determines the weights c_i. Therefore, with its use, each capsule can 'agree' with the output of the low level capsules. Hence, this algorithm is called dynamic routing by agreement.

Until now, the performance of capsule networks was mainly tested on classifying handwritten digits (MNIST dataset). The capsule network reached state-of-the-art results and performed especially well in recognizing overlapping digits [36]. Due to the number of weights and the high dimensionality, the training of capsule networks is time consuming and computationally intense. But, compared to traditional approaches, it is implicitly robust to affine transformations of objects due to its architecture. This makes capsule networks much smarter than the brute force method of CNNs and is definitely a big step towards general AI. Also, training data is used more efficiently, which is more human like. Another advantage is that it is built in a modular way which simplifies the understanding of the inner structure.

We believe that this approach has high potential in the future, when the computational resources are available. Especially, the field of object recognition might be revolutionized and since object recognition is the very foundation for many applications, capsule networks are worth investigating in detail.

6.7.2 Combining Specialized Agents into More General Ones

State-of-the-art algorithms have a limited capability. Thus, they only perform well in a very special domain. So, it seems natural that a connection of specialized agents is going to be an important trend during the upcoming years. Since this approach is already discussed in the chapters – Module and Collaboration, the interested reader is invited to look there for further details.

6.7.3 Research Environments

To enable general intelligence in the long run, AI systems must be created that perform beyond the capabilities of contemporary AI systems. Therefore, very complex environments must be created that allow training and teaching of AI systems. At the moment, 45 research and development projects tackle the research question – General AI. Across 30 countries in 6 continents, these projects are commonly based in major corporations and academic institutions, which are heavily funded [2]. The three most known and biggest projects are: OpenAI, DeepMind, and the Human Brain Project.

OpenAI

OpenAI is a non-profit organization focusing on research in artificial intelligence. The overall goal of the organization is to develop intelligent systems on an open-source basis for the benefit of mankind.

One field of their research focuses on building environments with feedback to generalize learning (see section 6.5). The aim is to reduce the apriori given information by humans. Therefore, the algorithms tend to be trained in a model-free manner. An example environment presented by OpenAI is a simulation of a robotic arm, which can interact and manipulate its environment. Goals currently are to move the end effector to the desired position; to hit a puck such that it slides to a desired goal; to pick up an object and place it somewhere. A classical approach would be to teach every goal manually by an engineer. Instead, the robot is on its own to explore its environment introduced by Lillicrap [25] and is rewarded for his actions.

For fast and generalized learning, the right type of feedback is important. Consider a robot which is only rewarded for achieving the end goal. We know that humans learn from failure, sometimes, even more than from success. If a human is supposed to hit a target with a puck, he also learns from unsuccessful trials. Information like "how is the friction of the surface?", "do I have to push harder?" or "is there a spin?" are acknowledged. At the time, the research focuses on adapting this kind of feedback for robotic simulations. In a paper by Andrychowicz [1], an algorithm called Hindsight Experience Replay (HER) is introduced, where a robot is rewarded for virtual goals. If the robot pushes the puck in the wrong direction the trajectory is re-examined with a different goal. With HER it can be shown that with the presented algorithm the learning speed of the robot can be improved. The recent trend in research tends towards improving and further generalizing the feedback.

DeepMind

DeepMind is a technological research company founded in 2010. After proposing stunning results, DeepMind was acquired by Google in 2014[4]. The company's focus lies on developing algorithms for solving and building Artificial General Intelligence. They develop solutions to fundamental problems in machine learning, healthcare, and ethics. With big successes like AlphaGo, DeepMind has shown its capability to be a major player in AI research in the coming years. Their newest accomplishment – teaching multiple agents to play the multiplayer game 'Quake 3' together and in an advanced team-play fashion, eventually beating human players [21] – fuels the hope for multiple cooperating AI systems in the future.

AI can be of extraordinary benefit to the world, but only if held to the highest ethical standards. With careful application, these technologies may help humanity solve some of its toughest challenges, like climate change and the universal provision of healthcare, paving the way for a better future. With its recent successes and Alphabet's funding to back up its research, we expect great advancements in AI development coming out of DeepMind.

Human Brain Project

The Human Brain Project (HBP) is a scientific research project with a focus to further understand and research the human brain activity[5]. The project is scheduled for 10 years (2013-2023) and has a budget of 1.19 Billion Euros. One research aspect is the simulation of the human brain. This task turns out to be quite complex since the human brain has about 86 billion brain cells, and today's computational power is not sufficient enough for this task. To reduce the complexity, the brain is modeled on a more abstract level.

An interesting approach on the path towards general AI is to use these simulation models and to embed them into an artificial environment. This is described in a paper by Falotico et. al. [10]. Due to the complexity of the brain model, the environment cannot deal with real-time constraints so far. The environment provides a realistic experience with physical elements like gravity and friction. Since this approach is fairly new, the future will show if this is a promising way.

6.7.4 Domains of Utilization

There are several domains where humans will recognize the impact of smart algorithms. However, it can be expected that, especially in these particular domains, the progress will be most dramatic:

- medical sector
- politics
- financial market

Of course, there might be many other areas in which AI is going to have a big impact, e.g. logistics and transportation. To keep this chapter concise, we only focus on the above mentioned ones. Following, we will elaborate on the potential innovations in the mentioned areas.

[4] `https://deepmind.com/about/` (accessed April 26, 2019)
[5] `https://www.humanbrainproject.eu/en/science/overview/` (accessed April 26, 2019)

Medical Sector

We expect mankind's biggest profit from artificial intelligence to be in the medical sector. By applying AI based systems, the doctors' decision making will be supported in order to make more accurate prognosis. Firstly, image based methods such as CNNs are used to segment a medical image into different classes. Each class is an indicator for the existence of a particular disease. Further approaches include the V-net for analysis of magnetic resonance imaging volumes [29] or the U-Net for bio-medical image segmentation [33]. Concrete examples for these applied techniques will be found in disease prognostics such as cancer detection [4][24].

Besides image based systems, artificial intelligence-based medical diagnosis support systems as described in [11] enhance the doctors' inference. Support systems such as IBM's Watson are adapted to serve as a clinical decision process [7]. By relying on such AI based medical systems, we expect a tremendous increase in the prediction accuracy of diseases.

Politics

Another sector, in which AI can have dramatic impact, might be politics. During the last years, the Brexit poll [5] as well es the election of Donald Trump in the USA [16] were highly influenced by intelligently using big data. Here, people were directly influenced by intelligent algorithms that where able to predict their behavior and, thus, how they would react to certain personalized (fake) news. These were often spread by AI-controlled social bots [8]. However, when written fake news already seems to be somehow normal, another way to influence people might happen via videos. While an authentic fake-video is still hard to create, AI systems could make this much easier in the future. Some research based on deep learning has already been done in this area [40]. Since there are already promising results and influencing people can be of high interest, it seems very clear that such systems will be further developed throughout the next decades.

Overall, influencing the public opinion might become a driving force in some parts of AI research. Although, there is also the possibility that society can use these systems in a more supportive way. If there is the possibility to create fake content via AI, it is very likely that there might be AIs that can detect such fake news. Moreover, it is imaginable that intelligent systems can support politicians in their decision making process, e.g. by analyzing complex data. Hence, politics may be an appealing sector for development and usage of AI.

Financial Market

The introduction of Bitcoin and the blockchain technology in 2008 can now be seen as one of the big steps towards the digitalization of currencies and the whole financial market [31]. Since that time, the interest in new technologies has grown even more. As explained in [41], blockchain in combination with AI has the potential to change the way financial compliance and regulation work. Thereby, companies would have to store and process all of their data in a centralized database that is secure from any manipulation. Then, authorities could offer all of their handbooks as an intelligent digital system such as an AI. This AI could not only give advice to match the regulations but also automatically check whether all requirements are fulfilled. As a consequence, a lot of workforce could be saved in addition to the whole process becoming more secure and transparent.

Moreover, the financial market is subject to change throughout the next decade. Especially, the prediction of the financial market is a promising area for AI research, since the analysis there is quite hard and relies very much on the recognition of already described patterns - a field of application in which intelligent algorithms perform traditionally well. Thus, there are efforts to use techniques like deep learning to do long term prediction on, e.g. the S&P 500 [14] from historical financial time series data. However, there are also indicators to predict stock prices that are not as obvious as pure financial data. As it can be seen in [32], the analysis of postings on social media platforms can help to predict changes in the financial market.

While these methods seem attractive because humans cannot process as much data as AI systems, there is one major disadvantage to this day: At the moment, all of these prediction methods only work reliably on historical data, i.e. on events that lie in the past. Hereby, especially deep learning methods seem to work very robustly. But in the future, banks, investment companies or even private investors might want to effectively use such systems in their daily business. Thus, AIs in this area can be assumed to get further developed during the upcoming years, since more reliable predictions mean more return on investment for the AI users.

6.8 Final Summary

Like today's progress in artificial intelligence, the future and cutting edge research also focuses on incremental improvements of rather weak AI. Research is still driven by specific problem statements and solutions are provided for very specialized engineering or academic problems. Nevertheless, there is a trend towards generalizing knowledge and finding new ways to make machine learning more efficient and explainable. Weak AI tends to get a bit stronger and able to extrapolate at least a little bit beyond their given training set. Due to optimized feedback during training, the learning process will be accelerated. While neural networks are still the major focus of current research, new and improved approaches like Capsule Networks are also being explored and will hopefully be able to sort out some of the current deep learning problems.

The next decade of algorithms will be used to solve real world problems. They are still very likely to be designed with the help of human background knowledge to tackle a specific problem. We will see even more astonishing AI accomplishments, but these AIs will probably all still be considered weak AIs. A new radical approach – to overcome these limitations and generate real, strong and thinking AI – is yet to come.

References

[1] Marcin Andrychowicz et al. "Hindsight Experience Replay". In: *CoRR* (2017).

[2] Seth Baum. "A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy". In: *Global Catastrophic Risk Institute Working Paper* 17-1 (2017).

[3] Yoshua Bengio et al. "Curriculum learning". In: *P26th annual international conference on machine learning*. 2009, pp. 41–48.

[4] Dmitrii Bychkov et al. "Deep learning based tissue analysis predicts outcome in colorectal cancer". In: *Scientific Reports* 8.1 (2018).

[5] Carole Cadwalladr. "The great British Brexit robbery: how our democracy was hijacked". In: *The Guardian* 20 (2017).

[6] Murray Campbell, A Joseph Hoane Jr, and Feng-hsiung Hsu. "Deep blue". In: *Artificial intelligence* 134.1-2 (2002), pp. 57–83.

[7] Christian Castaneda et al. "Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine". In: *Journal of clinical bioinformatics* 5.1 (2015), p. 4.

[8] Dominic DiFranzo and Kristine Gloria-Garcia. "Filter bubbles and fake news". In: *XRDS: Crossroads, The ACM Magazine for Students* 23.3 (2017), pp. 32–35.

[9] Kai Olav Ellefsen, Jean-Baptiste Mouret, and Jeff Clune. "Neural Modularity Helps Organisms Evolve to Learn New Skills without Forgetting Old Skills". In: *PLOS Computational Biology* 11.4 (Apr. 2015), pp. 1–24.

[10] Egidio Falotico et al. "Connecting Artificial Brains to Robots in a Comprehensive Simulation Framework: The Neurorobotics Platform". In: *Frontiers in Neurorobotics* 11 (2017).

[11] Wenjuan Fan et al. "Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS)". In: *Annals of Operations Research* (2018), pp. 1–26.

[12] David A. Ferrucci et al. "Building Watson: An Overview of the DeepQA Project." In: *AI Magazine* 31.3 (2010), pp. 59–79.

[13] Jan Feyereisl et al. "General AI Challenge - Round One: Gradual Learning". In: *CoRR* (2017).

[14] Thomas Fischer and Christopher Krauss. "Deep learning with long short-term memory networks for financial market predictions". In: *European Journal of Operational Research* 270.2 (2018), pp. 654–669.

[15] Ben Goertzel and Cassio Pennachin, eds. *Artificial General Intelligence*. Springer, 2007.

[16] Roberto J Gonzalez. "Hacking the citizenry?: Personality profiling, 'big data' and the election of Donald Trump". In: *Anthropology Today* 33.3 (2017), pp. 9–12.

[17] Bryce Goodman and Seth Flaxman. "European Union Regulations on Algorithmic Decision Making and a" Right to Explanation"". In: *AI Magazine* 38.3 (2017), pp. 50–57.

[18] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *CoRR* (2015).

[19] Geoffrey E Hinton, Sara Sabour, and Nicholas Frosst. "Matrix capsules with EM routing". In: *International Conference on Learning Representations*. 2018.

[20] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[21] Max Jaderberg et al. "Human-level performance in first-person multiplayer games with population-based deep reinforcement learning". In: *arXiv preprint arXiv:1807.01281* (2018).

[22] Garry Kasparov. *Deep thinking: where machine intelligence ends and human creativity begins*. Hachette UK, 2017.

[23] James Kirkpatrick et al. "Overcoming catastrophic forgetting in neural networks". In: *Proceedings of the National Academy of Sciences* (2017).

[24] Jinsa Kuruvilla and K. Gunavathi. "Lung cancer classification using neural networks for CT images". In: *Computer Methods and Programs in Biomedicine* 113.1 (2014), pp. 202–209.

[25] Timothy P Lillicrap et al. "Continuous control with deep reinforcement learning". In: *arXiv preprint arXiv:1509.02971* (2015).

[26] Gary Marcus. "Deep learning: A critical appraisal". In: *arXiv preprint arXiv:1801.00631* (2018).

[27] Stephen McAleer et al. “Solving the Rubik’s Cube Without Human Knowledge”. In: *arXiv preprint arXiv:1805.07470* (2018).

[28] Michael McCloskey and Neal J Cohen. “Catastrophic interference in connectionist networks: The sequential learning problem”. In: *Psychology of learning and motivation*. Vol. 24. 1989, pp. 109–165.

[29] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. “V-net: Fully convolutional neural networks for volumetric medical image segmentation”. In: *4th International Conference on 3D Vision*. 2016, pp. 565–571.

[30] Volodymyr Mnih et al. “Playing Atari with Deep Reinforcement Learning”. In: *CoRR* (2013).

[31] Satoshi Nakamoto. “Bitcoin: A peer-to-peer electronic cash system”. bitcoin.org. 2008.

[32] Nuno Oliveira, Paulo Cortez, and Nelson Areal. “The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices”. In: *Expert Systems with Applications* 73 (2017), pp. 125–144.

[33] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *CoRR* (2015).

[34] Marek Rosa and Jan Feyereisl. “A Framework for Searching for General Artificial Intelligence”. In: *CoRR* (2016).

[35] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 2nd ed. Pearson Education, 2003.

[36] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. “Dynamic Routing Between Capsules”. In: *Advances in Neural Information Processing Systems 30*. 2017, pp. 3856–3866.

[37] Jürgen Schmidhuber. “Deep learning in neural networks: An overview”. In: *Neural networks* 61 (2015), pp. 85–117.

[38] David Silver et al. “Mastering the Game of Go with Deep Neural Networks and Tree Search”. In: *Nature* 529.7587 (2016), pp. 484–489.

[39] Nitish Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15 (2014), pp. 1929–1958.

[40] J. Thies et al. “Face2Face: Real-time Face Capture and Reenactment of RGB Videos”. In: *Computer Vision and Pattern Recognition, IEEE*. 2016.

[41] Philip Treleaven, Bogdan Batrinca, et al. “Algorithmic regulation: automating financial compliance monitoring and regulation using AI and blockchain”. In: *Journal of Financial Transformation* 45 (2017), pp. 14–21.