Analysis and concept proposal of an artificial intelligence laboratory with thoroughly selected hardware components for research and teaching

Bernhard Haderer

Deggendorf Institute of Technology
Deggendorf, Germany
bernhard.haderer@stud.th-deg.de

***Abstract*—Today, more and more companies in all industries worldwide rely on methods of artificial intelligence to significantly improve their products, processes and services. The use of these technologies such as machine learning require modern and efficient hardware and perfectly trained specialists. The growth of artificial intelligence is counteracted by an enormous shortage of skilled personnel and the lack of key technologies in the form of a laboratory in many firms. This is where universities nowadays play a new role in providing support. Therefore, this paper describes the methodology to determine the laboratory equipment and expresses the main differences between research and teaching. After an analysis of different artificial intelligence application areas, the aim of this work is to vision a laboratory concept for teaching as well as for research projects and which is especially suitable for use at universities. Graphics processing units play a crucial role. An experiment underlines the significance of these hardware accelerators for artificial intelligence.**

Keywords— Artificial intelligence, machine learning, deep learning, laboratory design, universitiy, research and teaching, hardware, GPU

I. Introduction

Artificial intelligence (AI) as one of the most important digital future topics, is currently experiencing a huge upward trend in all fields of science, economy, technology, health care and entertainment. Despite its first peak in the 1950s, its current success can be explained by the exponential development of computer technology [1]. Thus, the mass of data (Big Data) can recently be handled effectively and efficiently by modern hardware accelerators such as graphics processing units (GPUs) [1-3]. Numerous companies around the world have already discovered the enormous potential of AI. The use of methods such as deep learning (DL) is expected to lead to competitive advantages.

However, many companies forget that the development and use of these new techniques require also perfectly trained staff [4], [5]. Especially in Germany, too few graduates are being trained for the AI field than the labour market requires. This is also the finding of a study conducted by the Leibniz Centre for European Economic Research in Mannheim, published on the 20th of March 2020 [4]. According to this study, the use of AI in the German industries is gaining momentum. The constantly growing demand for personnel, resulting from this, is countered by a considerable shortage of skilled workers. In 2019 for example, about 43% of open AI job positions remained unfilled [4]. Furthermore, many students have difficulties to follow new technical innovations in theoretical lessons without practice in a laboratory. Besides, companies often do not have the necessary resources and equipment to realize important AI research projects. As a result, many projects fail at an early stage without support from universities through modern AI laboratories. Furthermore, setting up an AI laboratory is a big challenge for companies and universities. It is often unclear, which hardware components are to be purchased. The variety of AI areas and hardware options require an AI equipment selection strategy that is missing or difficult to define.

The goal of this research paper is to propose a thoroughly selected hardware equipment for an AI laboratory for universities, after an analysis of the different applications in the field of AI. This should further support the upward trend and set the course for the future by ensuring that universities train highly professional graduates. The laboratory concept should also be designed for joint cooperation in research projects with industrial companies. All parties can benefit from this situation. This is also the challenge of the research work. This is because, in addition to teaching, the hardware equipment is also to be geared to larger research projects, but this represents a different view. Not part of the work is to suggest all manufacturers and types of every hardware device in detail, as well as to point out an exact cost plan, as this would go beyond the scope.

This paper is structured as follows: Section II outlines the methodology to design a laboratory concept. Section III describes background information and the different AI application areas. Section IV discusses the different requirements of a laboratory for research and teaching. Based on this knowledge, Section V shows a vision of the AI laboratory and the thoroughly selected hardware components. Section VI evaluates the necessity of GPUs with an experiment. Finally, Section VII concludes this paper.

II. METHODOLOGY TO DESIGN A LABORATORY CONCEPT

The laboratory requirements are determined step by step in more detail. For a thorough selection of laboratory AI hardware, the collection and evaluation of information and data are important parts of this approach. The individual steps are shown in Figure 1:

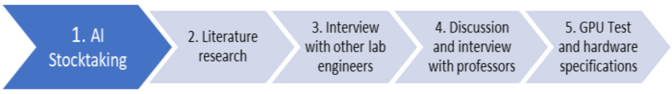

Figure 1: Procedure for the selection of laboratory equipment

The process shown in the figure above can be described as follows:

1. On the basis of a requirements analysis, it is determined, which AI resources are already in use in other departments and in the data center of the university.
2. The current state of technology is investigated in a literature analysis.
3. Hold interviews with other lab engineers to determine their approach and to agree on a possible cooperation in hardware matters.
4. After a first impression, the next step is to discuss the potential hardware equipment concept with professors, who want to use the AI lab.
5. In tests and more detailed analyses such as GPU tests, more detailed hardware specifications are set.

The result of this approach should represent a first rough concept of the AI laboratory and serve as a basis to get offers from manufacturers and suppliers.

III. RELATED WORK

In this section the scientific literature base of the work is explained. To meet the requirements for working on large AI-projects as well as for use in lectures, new approaches and principles are necessary as suggested by other preliminary studies. Many neural architectures such as support-vector-machines (SVM) or neural networks (NN) can be become enormously fast by using hardware accelerators in addition to normal processors. GPUs are especially optimized for training AI models by enabling parallel computational processing [2], [3], [6], [9]. This is necessary to keep the laboratory perfectly aligned and expandable for the future. Another aspect for the implementation of high-performance hardware is that students learn how to use such technology in practice for later safe handling.

Many hardware architectures and concepts concentrate on a single specific area. In this research paper, the laboratory should be used for two different goals. On the one hand this is teaching and on the other hand research in the field of AI in cooperation with companies. A study paper with this aim does not exist yet and is therefore described in this investigation. In this section, the theoretical basis for the laboratory concept will be briefly presented. This is one of the first steps to select the main AI hardware components and is therefore a very important part to get an overview about the current state of technology.

A. *Major Applications in AI Laboratories*

AI as a branch of computer science, covers a wide range of different methods and procedures. As a result, there are numerous possibilities that can be focused on. In many IT systems today, traces of AI can be found and have established themselves as a standard. Figure 2 shows the main areas [1], [7], [8]:

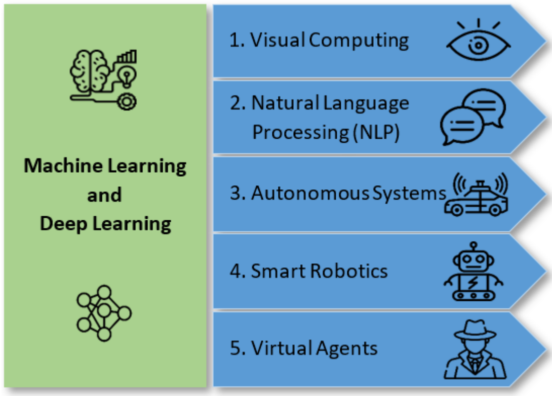

Figure 2: Main AI application areas

The figure above shows the most important fields of application areas. These areas can have different objectives:

Machine learning has developed into an important discipline in the IT industry over the last few years, located between data analytics, artificial intelligence and big data. The target of deep learning is to enable computer systems to solve complex tasks [10]. For example, cancer cells are to be detected on x-ray images [7], [8]. Many people are already using targeted intelligent assistants such as Amazon Alexa or even unconsciously hidden functionalities in webshops and search engines, which use AI to create user profiles about the user and analyze them in detail [7]. All these functionalities and topics are supported by deep learning, neural networks and AI approaches. They form the basis for further specialized subareas.

With the help of deep learning, enormous progress has been made in the field of computer vision in recent years. In addition to handwriting recognition, real time video analysis is now also possible, which can be used to monitor public places [8]. Online services such as Amazon Recognition or Microsoft Cognitive Services offer out-of-the-box solutions for the identification of people or fonts [7].

One of the most important goals of natural language processing (NLP) is to create a dialogue between humans and machines, in the complexity and form that is common among humans [8]. Here it is important not only to have the ability to recognize words and sentences, but also to have a lexicon combined with morphology. Furthermore, language assistants formulate answers themselves and have mastered a model for word formation [1]. For example, the language translator "DeepL" use neural networks as well as statistics, dictionaries, and grammars to try to create a translation model from one language to another. This analysis and transformation of millions of words in less than a second has only recently become possible thanks to new supercomputers. [7]

Autonomous systems act independently, are capable of learning, solve complex tasks and can move safely through the environment despite unforeseeable events [8]. In addition to robots, these systems also include intelligent machines and transport systems, which are used in the interest of humans or in areas which are too dangerous for living beings. Robots for defusing bombs or the Mars robots are among the most popular achievements of our past. Autonomous vehicles are currently among the most discussed and media-effective innovations. Technologies such as "Waymo" illustrate how automatic sensor fusion combined with deep learning can be used to control vehicles autonomously and intelligently. On this basis, large amounts of data, the corresponding models are trained outside the vehicle (offboard) using special learning methods. To prepare cars for rare events, training and test drives in virtual environments are essential. [7]

Smart robotics is one of the supreme disciplines within AI, since it combines many sub-disciplines of computer science including methods and techniques. Robots perform a variety of different tasks. Robots can be divided into the categories of industrial robots, service robots and humanoid robots [1], [8].

Software agents, often also called software robots, exist in countless variations and especially on the internet. They scour news sources, analyze the interest and buying behavior of users and produce personalized advertising tailored to the user. They learn the behavior of interested parties and can also react dynamically to changing environment situations. This kind of AI is the least noticed by many people today and has already integrated itself unconsciously in many areas of life. [1], [7]

Apart from the many different specialization subclasses (Figure 2, right), the basis of all of this is machine learning or deep learning (Figure 2, left). As a result, one should concentrate on machine learning and deep learning in the laboratory concept to keep it open for future potential areas.

B. GPUs for Deep Learning

Powerful graphics accelerators have brought the practical application of AI and the associated processing of large amounts of data in an effective and efficient way within reach. For example, the graphics card manufacturer Nvidia offers professional solutions for developers to perform deep learning analysis. The parallel architectures of GPUs and the associated large explicit properties for high data throughput are not only suitable for graphics calculation in the video game sector, but also for enormous AI performance. Therefore, chip manufacturers now offer their own dedicated AI platforms and numerous libraries for AI developers [9 -12].

For teaching and for many research projects in the AI field, the execution and training of applications on a central processing unit (CPU) is usually sufficient [10]. Nevertheless, there are some large projects that require a different approach. Another aspect for a GPU is that laboratories should also be designed for future projects, which need more performance to process the ever-increasing amount of data. In contrast to conventional processors, GPUs enable a more efficient processing of large amounts of data.

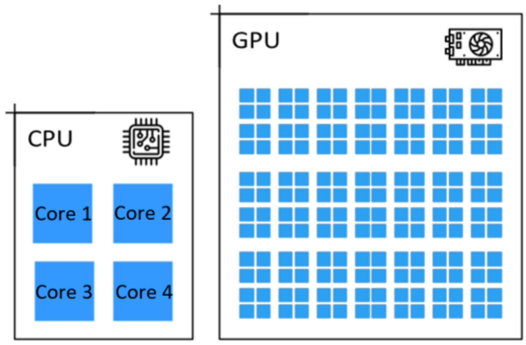

Figure 3: Difference between CPUs and GPUs

Figure 3 outlines the main difference between a CPU and a GPU. The preprocessing and training of artificial neural networks based on large amounts of data is very time-consuming and is no longer performed on CPUs, since they only consist of a few processor cores. They are designed for the processing of a large variety of different tasks, but less for simultaneous parallel calculation tasks. For example, the ability of a graphics card, with hundreds to thousands of such cores to compute complex data and operations very quickly [13]. Consequently, deep learning applications are hardly separable without implemented GPU support. A disadvantage is that graphics cards are much more expensive to purchase than usual CPUs. In addition to simple gaming graphics cards, Nvidia offers special business components for use in the machine learning sector. In addition to purchasing own GPU training servers, some service providers also offer special cloud solutions with servers for parallel data processing. These are also equipped with GPU clusters, such as several Nvidia Tesla V100 GPUs. [10-12]

IV. Requirements for the AI Laboratory

As mentioned in the previous sections, the laboratory will be used both for research and for practical lectures for students. As shown in Table I, it is important to start thinking about the differences between research and teaching at the beginning of the concept development:

TABLE I. DIFFERENCE BETWEEN RESEARCH AND TEACHING

Research	***Teaching***
Small number of projects	Many projects
Various projects in different AI areas	Almost similar projects in identical AI areas
High complexity	Small and simple teaching projects
Large file sizes or unstructured data files	Small file sizes or structured data files
Small numbers of researchers with expertise	Large number of students with less expertise
Requires more system performance and the integration of further technologies	Requires several similar workstations and AI gadgets for different lectures with less performance

Research	*Teaching*
Data security und data protection is important	Data security and data protection is less important than in research
More professional hardware components are necessary	Focus on easy handling for students with less experience

Table I shows the main differences between research and teaching in the use of an AI laboratory and therefore the abstract requirements. As can be seen, both concepts are different. These aspects were developed through interviews with professors and other laboratory engineers at the Deggendorf Institute of Technology and are based on their experience. For this purpose, an interview guideline was used to get good quality interview results. Some smaller scientific projects could also be carried out with the equipment of students, but this does not represent a claim to expandability for future large-scale projects. In summary, one can say that research projects can also be larger projects that require more powerful components. Unlike these, the main objective in teaching is cheap and easy to use hardware components.

V. Design of the Lab and Selection of the Main AI Hardware Components

To meet all requirements (research and teaching), the selection of AI laboratory equipment is divided into servers, workstations and AI gadgets.

A. Proposal for AI Workstations (in teaching)

After analysis of workstation variants, a good solution can be found in terms of price and performance. When selecting the workstations, the greatest possible value is placed on covering all requirements for teaching, student projects and smaller research projects. Best suited are tower workstations and not laptops, so that one can make own hardware adjustments over time such as replacing graphics cards. Even if the training of neural networks can still be performed on CPUs for small batch jobs, a GPU is recommended. So, there are no restrictions and the students are able to deal with this topic already to train neural networks efficiently. Depending on the offer price of these AI workstations, it may be advisable to buy separate graphics cards and assemble the workstations themselves.

Major deep learning frameworks such as TensorFlow, are already GPU-accelerated, so data scientists and researchers can get productive in minutes without GPU programming. Besides the usual gaming graphics cards, Nvidia offers expensive business GPUs, especially for training neural networks in deep learning. For example, if fifteen workstations must be set up in the laboratory, an expensive graphics card is not worthwhile per workstation. Technically it is not possible to share GPUs between several workstation without considerable effort, so each one of them needs its own card. The GeForce GTX 1080 or the GeForce RTX 2080 Ti are especially recommended in terms of price-performance ration. A recommendation and choice for Nvidia can be justified because it is not only the market leader in the deep learning area at the moment, but also provides the most powerful frameworks and developer tools in additions to the most capable GPUs.

An exemplary proposal for a potential workstation hardware is shown in Table II:

TABLE II. Example of requirements for AI Workstations

Major Requirements and specifications for the AI workstations		
Component	***Min. Requirement***	***Suggestion***
CPU	Intel Core i7-9700KF, 8 x 3.6 GHz	Intel Core i9-9900KF, 8 x 3.6 GHz
RAM	DDR-4, 2 x 8 GB	DDR-4, 2 x 16 GB
SSD	500 GB	1 TB
GPU	GeForce RTX 2080, 8 GB	GeForce RTX 2080 Ti, 11 GB

The table shows the decisive specifications with a focus on GPUs. To simplify the solutions, less relevant components such as mainboard, power supply or monitors have been omitted. Furthermore, this is only a proposal for orientation for the year 2020. Advantage of this solution is the easy handling by many students without complex environment configuration. This is especially important for lectures.

B. Proposal for a Deep Learning Server (in research)

In addition to the presented AI workstation solution, a special GPU server solution is necessary for large scientific projects in the deep learning area and for the training of large neural networks. These high-performance systems enable data scientists and researchers to parse petabyte-sized data much faster than with conventional CPUs or other server solutions. Another major advantage of this server infrastructure is that other departments of the university can also be operated by the same architecture. So, this infrastructure can be further expanded to a cluster solution with already existing AI resources.

As an example, a cost-effective and extendable basic solution would be a GPU Server with an Intel Dual CPU Xenon Processor (Silver or Gold), 512 GB DDR4-RAM, 4 TB SSD hard disc and 4 to 8 free slots for GPU accelerator cards. Initially two slots can be occupied with a GeForce RTX 2080 Ti, similar to the workstations or with one or two business GPU cards such as a Nvidia Tesla V100S (16 GB). The new Nvidia A100 Tensor Core graphics processor also offers a powerful alternative. The remaining free slots are left for future expansion. The server is not placed in the laboratory, but centrally in the data center of the university. The server can be used on top of the AI workstations, if dedicated computing power is required.

Virtual Environments with several GPUs can be created for different projects and used regardless of location. This concept is also preferred over external deep learning cloud solutions for data protection reasons and can be expanded with a network attached storage system (NAS) for storing large amounts of data.

C. AI Gadgets and Robots for Deep Learning Operation

At the beginning of the AI laboratory, expensive robots or mobile AI devices should not be purchased. Especially in the field of research and teaching, there are offers like developer boards with low costs, but almost unlimited development possibilities. This is how repair and maintenance can be carried out at low cost at the beginning of the laboratory.

Especially popular is the developer kit from Nvidia called Jetson. This is a small embedded AI board computer for autonomous machines. This low-cost and energy-efficient AI variant opens many new possibilities for embedded Internet-of-Things (IoT) applications such as drones, domestic robots or other intelligent devices in combination with AI. [14]

A somewhat more expensive alternative for the acquisition of robots in the laboratory would be the well-known humanoid robots "Pepper" or "NAO". Due to their high purchase price, these robots are less recommended for teaching but exclusively for selected research projects such as the service in the healthcare sector. It would be conceivable to purchase one to a maximum of two such robots if there is a corresponding need for research.

VI. Experiment and Evaluation of the Necessity of GPUs and the Laboratory Design

Thanks to the infrastructure concept consisting of workstations and a GPU server as shown in the previous section, both small and large projects can be processed. In addition to the simple use of personal computers, there is also the option of setting up complex virtual environments according to the needs of research projects. The entire infrastructure also allows an almost unlimited improvement or expansion due to the number of free GPU slots and the tower workstations. It is also guaranteed, that besides the complex high-performance hardware there is also an easy to use equipment for students in the laboratory.

A small experiment on a usual office laptop is intended to show how important it is to use GPUs for deep learning. Despite existing attempts, this experiment was carried out because many studies come to different conclusions [11], [13]. The experiment assumes that even low-cost GPUs hardware can achieve a large increase in performance. This test was carried out on a notebook with the following specifications:

- CPU: i7-8564U 8th Generation, 4 Cores, 1.8 GHz ClockSpeed and 4.6 GHz TurboSpeed
- RAM: 16 GB
- GPU: Nvidia GeForce MX150, 4 GB
- Test environment: Jupyter Notebook 6 and Python 3.7
- Operation System: Windows 10 (Pro)

The programming framework "CUDA" from Nvidia provides a parallel architecture for graphics processors that can be used for numeric computation. While a typical general-purpose Intel processor may have 4 or 8 cores, a GPU may have thousands of cores and a pipeline that supports parallel computations on thousands of threads, speeding up the processing considerably. This can be used to reduce training time for deep learning applications. [13]

The test case is to run a simple python script that generates two vectors with 20 million entries, that are randomly generated. A calculation is then performed out in such way, that each entry of the first vector is raised to the power of the corresponding entry in the second vector. A third vector stores the result of the calculation [13]. Table III represents five attempts of this experiment:

TABLE III. CPU vs. GPU performance comparision

Attempt	*CPU time in seconds*	*GPU time in seconds*
1	17.7108	1.2249
2	19.5245	1.3478
3	16.7102	1.2547
4	14.6875	1.1589
5	17.0189	0.9874

The result shows that even with simple parallel calculations, the runtime on a GPU is significantly shorter than on a CPU. On average, the GPU was approximately 15 times faster than the CPU in the test. The experiment was carried out several times to compensate for the influence of background tasks of the personal computer.

VII. Conclusion

This paper presents the consideration factors when selecting laboratory hardware for artificial intelligence. Great importance is attached to the use of both research and teaching. The main contributions of this paper are the following:

- The paper proposes a methodology to thoroughly select the hardware requirements for a new AI laboratory for research and teaching, especially for universities.
- The variety of different AI application areas and an analysis of current AI technologies, with a focus to laboratory equipment options, are shown.
- A concept is presented, which should serve as a guidance for setting up an AI laboratory and which points out the importance of flexibility and expandability.
- An experiment illustrates the importance of hardware accelerators, such as GPUs, for future research projects.

In future work and in addition to a precise selection of manufacturers, compatibility should be checked to integrate existing AI resources. Furthermore, Nvidia recently presented the A100 tensor graphics processor, a new and powerful high-performance-computing solution for deep learning. Further research and attempts in this direction as an alternative to a GPU server is advisable before purchasing the deep learning server.

REFERENCES

[1] S. Russell, P. Norvig, "Artificial Intelligence: A modern Approach", 4rd ed., Pearson Germany GmbH, April 2020, pp. 39-54, 59-53, 1120-1157

[2] M. Fukushi and Y. Kanbara, "A GPU Implementation Method of Deep Neural Networks Based on Data Swapping," *2019 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, YILAN, Taiwan, 2019, pp. 1-2, doi: 10.1109/ICCE-TW46550.2019.8991798

[3] K. V. Sai Sundar, L. R. Bonta, A. K. Reddy B., P. K. Baruah and S. S. Sankara, "Evaluating Training Time of Inception-v3 and Resnet-50,101 Models using TensorFlow across CPU and GPU," *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, 2018, pp. 1964-1968, doi: 10.1109/ICECA.2018.8474878

[4] "More Specialists Needed for Artificial Intelligence: Study on the Use of AI in Companies in Germany", Leibniz Centre for European Economic Research in Mannheim, 30.03.2020. [Online]. Available: https://www.zew.de/en/presse/pressearchiv/kuenstliche-intelligenz-braucht-fachkraefte

[5] Liudmila Alekseeva, José Azar, Mireia Gine, Sampsa Samila, Bledi Taska, "The demand for AI skills in the labour market", 03 May 2020. [Online]. Available: https://voxeu.org/article/demand-ai-skills-labour-market

[6] R. Dogaru and I. Dogaru, "Optimization of GPU and CPU acceleration for neural networks layers implemented in python," *2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)*, Galati, 2017, pp. 1-6, doi: 10.1109/ISEEE.2017.8170680

[7] M. Deru and A. Ndiaye, "Deep Learning with TensorFlow, Keras and TensorFlow.js", 1rd ed., Rheinwerk Computing, 2019, pp. 15-55

[8] J. Yeung, "Exploring Big Data and Artificial Intelligence: Three Major Fields of Artificial Intelligence and Their Industrial Applications", 22.02.2020. [Online]. Available: https://towardsdatascience.com/three-major-fields-of-artificial-intelligence-and-their-industrial-applications-8f67bf0c2b46

[9] H. El-Amir and M. Hamdy, "Deep Learning Pipeline: Building a Deep Learning Model with TensorFlow", Apress Berkeley CA, 2020, pp. 3-36, doi: 10.1007/978-1-4842-5349-6

[10] J. Ekström, "What is a GPU and do you need one in Deep Learning?", 25.04.2020. [Online]. Available: https://towardsdatascience.com/what-is-a-gpu-and-do-you-need-one-in-deep-learning-718b9597aa0d

[11] A. Kayid, Y. Khaled, M. Elmahdy, "Performance of CPUs/GPUs for Deep Learning workloads", 2018, researchgate.net, 10.13140/RG.2.2.22603.54563

[12] M. Madiajagan, S. Sridhar Raj, "Deep Learning and Parallel Computing Environment for Bioengineering Systems: Chapter 1 - Parallel Computing, Graphics Processing Unit (GPU) and New Hardware for Deep Learning in Computational Intelligence Research", Academic Press, 2019, pp. 1-15, https://doi.org/10.1016/B978-0-12-816718-2.00008-7, Available:http://www.sciencedirect.com/science/article/pii/B9780128167182000087

[13] A. Weeraman, "How to put that GPU to good use with Python", September 2017. [Online]. Available: https://weeraman.com/put-that-gpu-to-good-use-with-python-e5a437168c01

[14] "www.t3n.de", 20.03.2019. [Online]. Available: https://t3n.de/news/jetson-nano-nvidia-stellt-ki-entwicklerkit-fuer-99-dollar-vor-1151336/