h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Hochschule Darmstadt

– Fachbereich Mathematik und Naturwissenschaften–

Sustainability of Artificial Intelligence in Corporations: Status Quo and Trends

Abschlussarbeit zur Erlangung des akademischen Grades

Master of Science (M.Sc.)

vorgelegt von

Luise Wiesalla

Matrikelnummer: 738810

Referent	:	Prof. Dr. Bernhard Humm
Korreferent	:	Prof. Dr. Horst Zisgen

DECLARATION

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht.

Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen.

Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

Darmstadt, 14. Juli 2022

Luise Wiesalla

ABSTRACT

Sustainability and artificial intelligence are two trending topics for corporations, which are brought together in this work. Due to the rapid growth of state-of-the-art models and the steady spread in companies, the footprint that AI applications produce is no longer negligible. While in research the topic of *Green AI* is slowly gaining a foothold, its relevance in enterprises has not yet been scientifically captured. This is done in the course of the work alongside the derivation of beneficial measures to improve the sustainability of artificial intelligence.

In discussions with 12 experts from German companies with different levels of maturity in artificial intelligence (AI), the current status of the topic and future development, including hurdles and drivers, were determined qualitatively. Here it becomes apparent that the topic has not yet arrived in general and that there are only isolated efforts for a sustainable design of AI. In particular, the issue of transparency with regard to the emissions caused and effective measures comes up frequently here.

In order to not only illustrate the development, but also to actively influence it, two contributions on the resource-conscious design of AI can be found in this thesis. One contribution is a comprehensive collection of best practices and measures for less emissions via a systematic literature search. Since enterprises are specifically addressed here, these were evaluated for applicability in terms of (A) time and resource requirements (B) impact on model performance (C) typagnosis and (D) constraint of application. A total of 27 measures were derived from the areas of hardware and resources, model selection, model training, model operation, and organization. Many of these are relatively easy to apply. Others can be well integrated if the AI domain is still being built up.

Finally, the second contribution includes an evaluation of model types of classical machine learning from an environmental sustainability perspective. Here, a benchmark of 9 different classification algorithms was made with respect to their energy consumption in training and prediction.

ZUSAMMENFASSUNG

Nachhaltigkeit und Künstliche Intelligenz sind für Unternehmen zwei Trendthemen, die in dieser Arbeit zusammengeführt werden. Durch den rasanten Wachstum der state-of-the-art Modellen und die stetige Verbreitung in Unternehmen, ist der Fußabdruck den KI Anwendungen produzieren nicht mehr vernachlässigbar. Während in der Forschung das Thema der *Green AI* langsam Fuß fasst, ist die Relevanz in Unternehmen noch nicht wissenschaftlich erfasst. Dies geschieht im Verlauf der Arbeit neben der Ableitung von förderlichen Maßnahmen zur Verbesserung der Nachhaltigkeit von Künstlicher Intelligenz.

Im Gespräch mit 12 Experten aus deutschen Unternehmen mit unterschiedlichen Reifegrad der Künstliche Intelligenz (KI) wurde der aktuelle Stand des Themas und die zukünftige Entwicklung samt Hürden und Treiber qualitativ bestimmt. Hier zeigt sich, dass das Thema im Allgemeinen noch nicht angekommen ist und nur vereinzelt Bestrebungen für eine nachhatlige Gestaltung von KI existieren. Insbesondere das Thema Transparenz hinsichtlich der verursachten Emissionen und wirksamen Maßnahmen kommt hier immer wieder auf.

Damit die Entwicklung nicht nur abgebildet, sondern auch aktiv beeinflusst werden kann, finden sich zwei Beiträge zur ressourcenbewussten Gestaltung von KI in dieser Thesis. Ein Beitrag ist eine umfassende Sammlung an Best Practices und Maßnahmen für weniger Emissionen über eine systematische Literatursuche. Da hier insbesondere Unternehmen angesprochen werden, wurden diese nach der Anwendbarkeit bezüglich (A) Zeit- und Ressourcenanforderungen (B) Auswirkungen auf die Modellperformance (C) Typagnosität und (D) Einschränkung der Anwendung bewertet. Insgesamt wurden 27 Maßnahmen aus den Bereichen Hardware und Ressourcen, Modellauswahl, Modelltraining, Modellbetrieb und Organisation abgeleitet. Viele davon sind relativ einfach anwendbar. Andere sind gut integrierbar, wenn der KI Bereich noch aufgebaut wird.

Der zweite Beitrag umfasst schließlich eine Bewertung von Modelltypen des klassichen Machinen Learnings aus der Perspektive von ökologischer Nachhaltigkeit. Hier wurde ein Benchmark von 9 verschiedenen Klassifikationsalgorithmen hinsichtlich ihres Energieverbrauches bei dem Training und der Vorhersage angefertigt.

CONTENTS

I THESIS
1 INTRODUCTION 2
2 THEORETICAL BACKGROUND 5
2.1 Artificial Intelligence 5
2.1.1 Machine learning models 5
2.1.2 Technology adoption in companies 7
2.2 Sustainability in IT 8
2.3 Green AI 9
2.3.1 Quanitfying Sustainability 10
2.3.2 Available libraries and tools for carbon accounting for AI 11
2.3.3 Current hurdles of Green AI in companies 12
3 METHODOLOGY 13
3.1 Literature review 13
3.2 Expert interview 14
3.2.1 Qualitative Research 15
3.2.2 Preparation of the interview 16
3.2.3 Evaluation 18
3.3 Ranking of the energy consumption of classification algorithms 19
4 POSSIBLE ACTIONS FROM RESEARCH 22
4.1 Hardware and resources 22
4.2 Model selection 27
4.3 Model training and automated machine learning 32
4.4 Model Operations 36
4.5 Organizational 38
4.6 Discussion 40
5 RELEVANCE OF SUSTAINABILITY IN COMPANIES 41
5.1 Results of the gained data 42
5.1.1 Reasons for sustainability in general 43
5.1.2 Relevance of the topic 45
5.1.3 Hurdles for more sustainability in AI 47
5.1.4 Necessary steps for more sustainability 49
5.1.5 Forecast development 52
5.2 Discussion 54
6 RANKING OF CLASSIFICATION ALGORITHMS 57
6.1 Results of the benchmark 57
6.1.1 Training of the models 57
6.1.2 Prediction of the models 58
6.2 Discussion 58
7 DISCUSSION OF THEORETICAL AND PRACTICAL RESULTS 60
8 CONCLUSION AND FUTURE WORK 62

II APPENDIX
A EXPERT INTERVIEW MATERIAL 64
A.1 Interview guide 64
A.2 Maturity of AI in the interviewed companies 66
A.3 Results of qualitative content analysis 69
A.3.1 Reasons for sustainability in companies 69
A.3.2 Relevance of sustainability in the data science department 73
A.3.3 Hurdles to a more sustainable AI 78
A.3.4 Necessary steps for a sustainable AI 83
A.3.5 Future of sustainable AI 87
B MODEL RANKING MATERIAL 90

BIBLIOGRAPHY 91

LIST OF FIGURES

Figure 1.1 Computational cost of AI models measured by petaflop/s-day: 1 petaflop/s-day means performing 10^{15} neural net operations per second to run the experiment in one day. Taken from [Amo+19] . . . 2
Figure 2.1 schematic structure of a neural network with two hidden layers . . . 7
Figure 2.2 Focus of 60 randomly selected conference papers of top AI conferences determined by contents; taken from [Sch+19] . . . 9
Figure 3.1 7 Steps of qualitative content analysis according to Mayring . . . 19
Figure 4.1 Measurement of the PUE of a data center daily average in a year (A) and hourly average in the summer time (B). Taken from [McD+22] . . . 25
Figure 4.2 methods for compact architecture design. Taken from [Xu+21] . . . 29
Figure 4.3 Comparison of the compressing tactics hybrid matrix factorization (HBF), pruning (P), low rank matrix factorization (LMF) and train a smaller baseline model (SB). The accuracy and speed up depending on the compression rate is displayed. Taken from [Tha+20] . . . 38
Figure 5.1 amount of interviewed companies in a specific maturity level . . . 42
Figure 5.2 categorized reasons for companies to look deeper in sustainability . . . 43
Figure 5.3 categorized relevance of sustainability of AI in companies . . . 45
Figure 5.4 categorized hurdles of sustainable AI . . . 48
Figure 5.5 next steps for more sustainability in AI . . . 50
Figure 5.6 Experts' opinions on the development of environmental sustainability by and with AI . . . 52
Figure 6.1 logarithmic normative energy consumption of models in training . . . 57
Figure 6.2 logarithmic normative energy consumption of models in prediction . . . 58
Figure B.1 logarithmic mean normative energy consumption of models in training . . . 90
Figure B.2 logarithmic mean normative energy consumption of models in prediction . . . 90

LIST OF TABLES

Table 2.1 Artificial Intelligence Maturity Model taken from [ACM19] 7
Table 2.2 Overview measurements for quantifying the footprint of an model based on [Tor+21] 11
Table 3.1 models for energy consumption benchmark. 21
Table 4.1 Collected measures for hardware and infrastructure selection 23
Table 4.2 Collected measures for model selection 27
Table 4.3 Collected measures for model training and automated machine learning 32
Table 4.4 Collected measures for model operations 36
Table 4.5 Collected measures on an organizational level 38
Table 5.1 Example lines for the categorization of reasons 44
Table 5.2 Example lines for the categorization of relevance 47
Table 5.3 Example lines for the categorization of hurdles 49
Table 5.4 Example lines for the categorization of next steps . . . 51
Table 5.5 Example lines for the categorization of the topic development 53
Table A.1 coded passages reasons for categorize the interviewee in the maturity model 66
Table A.2 document statistics reasons sustainability 69
Table A.3 coded passages reasons for sustainability 69
Table A.4 document statistics relevance sustainability 73
Table A.5 coded passages relevance of sustainability 73
Table A.6 document statistics hurdles sustainability 78
Table A.7 coded passages hurdles sustainability 78
Table A.8 document statistics necessary steps 83
Table A.9 coded passages necessary steps 83
Table A.10 document statistics outlook 87
Table A.11 coded passages outlook 87

ABBREVIATIONS

AI Artifical Intelligence

ML Machine Learning

CPU Central Processing Unit

GPU Graphical Processing Unit

TPU Tensor Processing Unit

FPGA Field-Programmable Gate Arrays

ASIC Application specific Integrated Circuits

PUE Power usage effectiveness

GLaM Generalist Language Model

AutoML Automated Machine Learning

Part I

THESIS

1

INTRODUCTION

MOTIVATION

Due to the ongoing global climate crisis, the issue of sustainability is gaining in importance socially, politically, and economically. Although Artifical Intelligence (AI) is seen as a possible solution to fight global warming and improve the overall efficiency of companies worldwide, the costly model training and a large amount of data can lead to a high energy consumption which is a negative effect for sustainability. And as AI advances, the ecological impact of resource-intensive computing becomes increasingly significant.

Figure 1.1 shows an analysis of the OpenAI research lab in 2018. Here, the computational cost of selected AI models over the years is displayed. Over the course of AlexNet to AlphaGoZero, there was an increase by the factor of 300,000 regarding computational effort. This 3.4 months doublings[1] served as an alarm signal to bring sustainability of artificial intelligence itself into research.

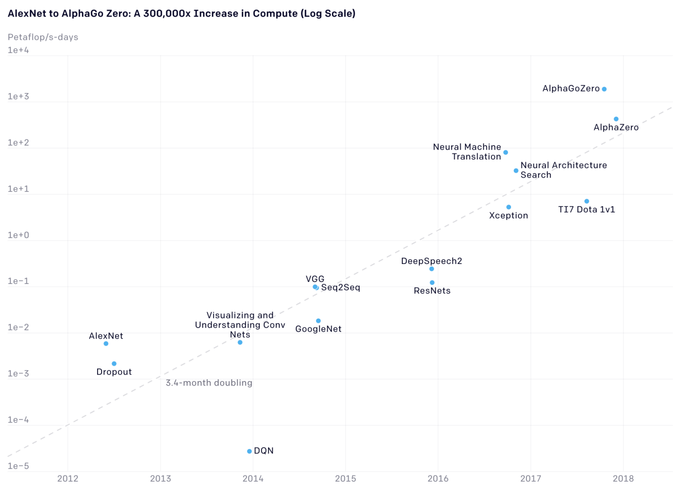

Figure 1.1: Computational cost of AI models measured by petaflop/s-day: 1 petaflop/s-day means performing 10^{15} neural net operations per second to run the experiment in one day. Taken from [Amo+19]

If the computational effort is brought to a CO2 equivalent, the necessity of consideration becomes even more clear. The training of a former state-of-

1 The exponential growth might not continue though. The training of the recent GPT-3 model (May 2020) needed 3000 petaflop/s-day [Bro+20]. This is above the 1850 petaflop/s-day [Amo+19] period of computational cost of AlphaGoZero (October 2017), but below the predicted factor of ca. 2^{10}.

the-art transformer model with 213 million parameter in the field of natural language processing (NLP) and its related architecture search is estimated to produced CO2 equivalent to 5 times the lifetime of an American car including the overall gasoline consumption. [SGM19]

However, artificial intelligence is not reserved for research anymore. While the resource consumption of the development and benchmark of state-of-the-art models is of primary importance in research, the operation of the models is the main focus in the business application. According to Amazon, 90% [Bar19] of the cloud resources are used for the interference. A resource-saving design of the implementation therefore raises special requirements compared to sustainability of AI in research.

In companies, this technology has been rising for years. In a 2021 survey conducted by the digital association Bitkom 8% (previous year 6%) of the surveyed German companies said they were already using AI and a further 30% (previous year 22%) were already making concrete plans to use it. [SU21] Thus, the number of applications will increase sharply over time.[2]

In summary, due to the size of the models in more complex use cases and continuous growth of applications, artificial intelligence will soon become a non-negligible factor for climate change. Here, in particular, the question of a resource-friendly design of AI in companies still needs to be addressed, which will be covered in this thesis.

RESEARCH GOAL

The goal of this thesis is to investigate the status of sustainability of Artificial Intelligence applications in enterprises and to develop an applicable set of actions to promote a sustainable AI. The present work is dedicated to the central research question:

What is the awareness and status of sustainability of artificial intelligence in companies and how can it be improved?

Since the terms Artificial Intelligence and sustainability are broad concepts, they are more narrowly defined for this work in chapter 2. The work itself is an intersection between literature work regarding existing measures from research and best practices from the AI community,a conversation with AI responsible persons in companies to collect empirical data for representing the current status and a practical benchmark for machine learning algorithms. The subordinate research questions on which the method selection is based can be found in 3.

2 The survey covers 603 companies with 20 or more employees, which were interviewed via telephone calls

In the discussions with companies, the relevance of sustainability is worked out and in particular, hurdles and future developments are collected. Based on this, the prioritized need for action for this topic in research will be created.

STRUCTURE OF THE THESIS

In chapter 2 the theoretical background for this thesis will be laid out. Here the measurement of ecological sustainability and the influencing factors will be presented. Following up on this information, chapter 3 will present the methodology used for this thesis. chapter 4 then shows the collected measures that lead to a net increase in environmental sustainability. Since these should be specifically applicable to companies, special evaluation criteria are used and the improvement is quantified if possible. This is intended to enable an easy transfer to companies and lower the hurdle of application. In chapter 5 the expert interviews and their aggregated results about the status of sustainability in AI will be covered. The practical benchmark of selected classification algorithms will be found in chapter 6. A consolidation and discussion of the theoretical and practical results will be done in chapter 7 and finally a Conclusion about future work will be formed in chapter 8

2 THEORETICAL BACKGROUND

In these sections, artificial intelligence in enterprises and the definitions as well as metrics of sustainability are presented.

2.1 ARTIFICIAL INTELLIGENCE

Artificial Intelligence deals with the automation of intelligent behavior by machines and in particular machine learning. There are various definitions available. Therefore, the definition for Artificial Intelligence in companies, as it is used for this thesis, is presented below.

ARTIFICIAL INTELLIGENCE *In this context, Artificial Intelligence describes statistical modeling designed with primarily (but not exclusively) business data. The modeling is used descriptively or predictively in day-to-day business. The amount of data is usually very large and the methods used are from the field of machine learning and deep learning. The development or adaptation of the models is done in the company by programming or platform-supported specifically for the respective use case.*

To have a broad overview, of some machine learning algorithms and their concepts, some algorithms mentioned in this thesis are presented in the following section.

2.1.1 *Machine learning models*

The treated procedures are procedures of the *supervised learning*. Here a target value (regression) or a target class (classification) is given, which is to be reproduced by the model on the basis of data. In the case of classification, for example, data on customer behavior is used to predict whether a customer will buy a product or not.

The learning of a classification of the data can be realized by different approaches:

NAIVE BAYES The Naive Bayes Method works in essence with a probability model. Using the rules on conditional probabilities of Bayes, the class is selected which, given the data, has the largest conditional probability for the class. [Nas17] The different methods (Bernoulli, Multinomial, Gaussian) are according to the assumption about the probability distribution.

LOGISTIC REGRESSION Logistic regression is a discriminative classifier. [Nas17] In contrast to Naive Bayes, it is intended to find bounds that separate classes. In its basic form, it separates only binary variables (0 and 1), but can be combined to predict multiple classes. The data are fitted into a logistic function so that the output corresponds to a probability.

DECISION TREE In decision trees, the data is separated recursively. Starting with the root (all data), further nodes are formed by optimizing the purity of the classes through decision rules. In the leaves, the last node without outgoing edge, the class is formed by a majority principle.[Nas17]

K-NEAREST-NEIGHBOURS The KNN method is based on the proximity of a new data point to old known data points in the hypersphere of features. The class of the new data point is determined by the classes of the nearest points. The k indicates how many of the nearest points (neighbors) are included in the classification decision. [Muh15]

SUPPORT VECTOR MACHINES The technique of Support Vector Machines is based on the concept of solution planes. These separate the instances of the different classes in the highly diemensional space linearly from each other. [Muh15] The class separation line should have as wide a corridor as possible. For non-linear classification, the kernel trick is used, which extends the object space by additional dimensions.

BAGGING ENSEMBLE The idea of an ensemble is that a large number of models jointly decide which class to choose. In the method of bagging (bootstrap aggregating), the n base learners are trained on a sample of the data (drawn with laying back). [CCS12]

RANDOM FOREST A random forest is similar to the already presented bagging method with decision trees. One difference here is that only a random subset of features is available when deciding by which features to split. [CCS12]

BOOSTING ENSEMBLE In the process of boosting, multiple weak learnings are trained. In the case of Adaboost, decision trees with only one partition. In the iterative training of the base learners, adaptive weights are used to give more weight to data points that are difficult to classify. In the end, the weak learners are merged into a strong learner. [CCS12]

ARTIFICIAL NEURAL NETWORKS The field of deep learning was founded with artificial neural networks. Even if the architectures become more and more complex, the basic principle is simple. In the input layer, the features are acquired to one neuron each. These values are then forwarded to the neurons of a hidden layers. There, the input values are summed up at the corresponding neurons and subjected to an activation function. This results in the value of the neuron, which in turn can be forwarded itself. In the output layer there is one neuron per class. Here the highest value decides for the output of the class. [Muh15]

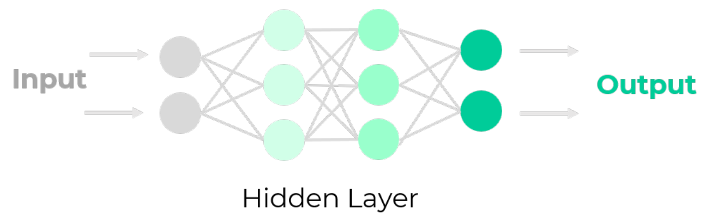

Figure 2.1: schematic structure of a neural network with two hidden layers

2.1.2 *Technology adoption in companies*

Since the technology adoption depends on various factors like management support, competitive pressure, resources and size [ACM18], artificial intelligence has different levels of maturity in enterprises, which is particularly important for the study of sustainability. In Alsheibani et. al. (2019), a first maturity model of general AI in corporations was developed based on four dimensions. The dimensions *AI functions* (tools and technologies), *data structure* (amount and structure of data), *people* (creator of AI systems) and *organizational* (influencing factors regarding business) serve as foundations for the model, which is displayed in table 2.1.

Table 2.1: Artificial Intelligence Maturity Model taken from [ACM19]

level	AI functions	Data structure	People	Organisation
Level 1 Initial	Very limited or no AI function exists and has no plans	Regular data structure; no data exists to train AI	Regular IT skills; Organisations lack the skills to evaluate, build and deploy AI solutions	No business case related to AI; existing structures are used informally
Level 2 Assessing	Discovery AI Technology	Integration of current usage of AI into data required to train AI	AI related training; Assessment of existing infrastructure with regard AI	Organisation initial AI strategy; for each AI application, have defined a value proportion
Level 3 Determined	AI project is at an advanced stage; determination of infrastructure needed to further implement AI	Custom AI data are introduced; data standardised are exist	Active management support; resources are provided, AI related employees training	Organisation has standard operating procedures that cover AI scenarios; change management is introduced
Level 4 Managed	AI process are defined throughout the organisations	Appropriate data science exists to make critical business decisions using AI	AI is being fully realised as employees' productivity	There is a well-defined value to support and full top management support
Level 5 Optimised	Full AI infrastructure adoption and standardisation	Proactive data analysis; Data is available in real time	Employees are engaged; centralised leadership	Role, responsibilities and accountability are clearly defined within each AI project; AI culture

The levels are later used to assess the maturity of AI in the companies surveyed in the expert interview and to further interpret the results.

2.2 SUSTAINABILITY IN IT

Sustainability in terms of energy efficiency is already an topic in IT since 2007[1] and in some cases also earlier. Labels like the American *Energy Star 4.0* and the German *Blauer Engel* should help consumers and businesses make sustainable decisions regarding hardware and also software[2].
While governmental organizations such as the German Federal Environmental Agency (*Umweltbundesamt*) and NGOs such as *Greenpeace* are addressing the issue, specialized networks such as *Association for Progressive Communications (APC)* are also emerging to push social and ethical issues alongside environmental ones.
The general buzzword for the ecological efforts is *Green IT* and it comes with various definitions. In [DJ15] over 16 definitions are gathered. For this thesis the definition by Elliot (2011) is selected:

> "Activities to minimize the negative impacts and maximize the positive impacts of human behavior on the environment through the design, production, application, operation, and disposal of IT and IT-enabled products and services throughout their life cycle." [Ell11]

This comprehensive definition covers the entire lifecycle of IT applications and their software and hardware components. In addition, ecologically positive use cases are approved as a contribution to greener IT by maximizing the positive impact.

STATUS OF GREEN IT IN COMPANIES If you want to determine how widespread the topic of green IT is in companies, a study by Capgemini [Ins21] from 2021 provides a good indication of this. Here, 1 000 companies worldwide were surveyed. Only 6% of the companies have a high level of maturity with regard to the sustainable design of their IT department. Overall, 57% of respondents do not know what the footprint of their corporate IT is. 34% do not know that the production of a cell phone or laptop generates more emissions than its entire lifetime. The awareness is therefore not yet established in the companies.
Only 29% of companies record their carbon footprint using suitable tools. Monitoring is often the first step in tracking a change.
Overall, there is a strong expectation for leading technology companies to establish more sustainable IT practices that companies can leverage. 52% believe that sustainability should be a dimension of the products and services of these technology companies.

Overall, with these thoughts about Green IT in mind, a variation for artificial intelligence has already been designed.

1 In 2007 the label for energy efficiency *Energy Star* became available for computers
2 Basic work on the formation of criteria for green software evaluation has been defined https://www.blauer-engel.de/de/publikationen/detail/entwicklung-und-anwendung-von-bewertungsgrundlagen-fuer-ressourceneffiziente

2.3 GREEN AI

The term "Green AI" was introduced in 2019 by Roy Schwartz et. al. [Sch+19] The position paper warns about the impact of the current development of Artificial Intelligence research. As already pointed out in the introduction of this thesis, the trend of computations required for state-of-the-art models was analyzed for this purpose, where for the period 2012-2018 a doubling every 3.4 months was determined.

Particularly in the AI research community, the trend is recognized to base the research of models predominantly on performance. Other factors are widely ignored. Researchers are accused of "buying" better results with larger models and larger data sets at the expense of computational costs and therefore lastly the environment. This dominant approach is summarized under the term **Red AI**. This trend is visible is the graph Here the ratio of the focus on accuracy, efficiency, both or other areas of 60 randomly selected top 2019 AI conference papers can be seen. The focus on accuracy is dominant in all conferences. It is worth noting that the NeurIPS papers often give convergence rate or regret bounds, which represent performance as a function of datasets and iterations, and so they also count to consider efficieny.

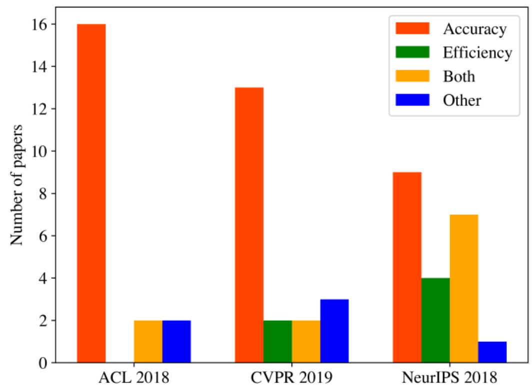

Figure 2.2: Focus of 60 randomly selected conference papers of top AI conferences determined by contents; taken from [Sch+19]

In contrast, **Green AI** is intended to characterize research on artificial intelligence that is more environmentally friendly and inclusive. Inclusiveness in the sense as in that students and researchers without extensive research budgets also have the opportunity to participate in the research. Here, the efficiency of the models is a priority. For this purpose, the calculation costs of model selection, model training, and model operation should be included in the research work.

Even though the term "Green AI" is primarily related to research, it has an impact on model development in companies. On the one hand, the sustainability of the application of AI fundamentally benefits from more efficient

models. On the other hand, it is desirable for model developers to be able to estimate the costs of the model by using provided information about the computational effort and the used data size in research. This enables the economic and sustainable comparability of different models.
Apart from green AI in research, there are a few programs out there like Microsoft's *AI for Earth*, which provide funds and tool access for organisations which aim to address the climate change.

Finally, in order to find a definition that brings together the aspects of Green IT and the business context of Green AI, the following definition is used in this work:

SUSTAINABILITY OF AI *In this context, sustainability is understood as the ecologically conscious actions oriented towards the long-term effects of Artificial Intelligence application on the environment. This includes environmentally and resource friendly design throughout the life cycle, as well as general resource conservation through the application itself. The ethical and social aspects of sustainability are not in the primary focus.*

The remaining question now is how to quantify the sustainability of artificial intelligence.

2.3.1 *Quanitfying Sustainability*

Quantifying sustainability is a complex undertaking. On the one hand, the reduction of ethical, social and ecological aspects to scales is challenging and ambiguous. On the other hand, comparability among the categories of sustainability is not objectively given.
Since this thesis primarily focuses on the ecological aspect in the application of artificial intelligence, the following section presents the challenges and approaches to solutions in the quantification of this aspect.

PRELIMINARY CONSIDERATIONS OF QUANTIFICATION A well-known problem of capturing an ecological footprint of systems is the definition of the system boundaries. With regard to artificial intelligence, it must be decided which components are to be included. In addition to model training and application of the model, data storage, preliminary work for modeling, monitoring and retraining of the final model, and hyperparameter optimization, which can be costly, can also be included. System boundaries may also include hardware and infrastructure acquisition and wear and tear. Here, details about usage and predicted lifetime must be available to get a reasonably accurate value, which is practically very laborious.

Since this work primarily focuses on the models and model operation, the ecological footprint is determined by the training time (and also later post-training time) and the use of the model. For this purpose, the following measures are commonly used in the literature:

- *runtime* - total time of model training or execution
- *CPU/GPU Hours* - real execution time on GPU or CPU[3]
- *floating point operations* - amount of floating point operation needed
- *energy consumption* - amount of energy needed for the computation
- *CO2 equivalent* - estimate the impact with energy consumption and energy mix

While all values allow comparability of models and their execution, some methods are better and worse per criterion. Table 2.2 summarizes the evaluation. It should be noted that even the floating point operation is hardware dependent due to the hardware optimized compilation of models. And further more in an experiment by [Hen+20] it was shown, that the floating point operations are in some cases uncorrelated with the energy consumption.
The metric CO2 equivalent should be highlighted because it offers a comparability to other areas like transportation or production. The other metric worth emphasizing is the CPU/GPU hours. It supplies a value, which is usually well measurable and cleansed out of the interference of other processes and remains interpretable by humans.

Table 2.2: Overview measurements for quantifying the footprint of an model based on [Tor+21]

Measure	**Measureability**	**Human Interpretability**	**Footprint Estimate**
Runtime	++	o	o
CPU/GPU Hours	+	o	o
Floating Point Operations	+	-	-
Energy Consumption	-	+	+
CO2 Equivalent	-	+	+

Especially if the environmental cost of AI is to be captured as a carbon footprint for comparability, there are many tools that easily integrate what is called *carbon accounting* for developers.

2.3.2 *Available libraries and tools for carbon accounting for AI*

The range of libraries and online tools for recording the footprint is large. An introduction to the most important ones is given here.

3 without other running processes interfere

Machine Learning Emissions Calculator [Lac+19b]

Since 2020, the web-based Machine Learning Emissions Calculator has come up from a research work. After entering the hardware type, the number of hours of calculation, the cloud operator and the region used, an emission value and its calculation are output transparently. Also integrated is how much emissions are already offset by the provider. Public data sources are used here to always obtain current values regarding the energy consumption of the hardware and the energy mix of the site. The primary focus here is to make the influence of different hardware and the calculation location visible. The calculation on the own infrastructure can also be specified.

Carbon Tracker [AKS20a]

For the development of deep learning models, Carbon Tracker is a good way to estimate the footprint via a library. Here, the energy consumption is projected over a given number of measured epochs. The output is directly in the output and can be further used with a integrated log parser.

EnergyVis [Sha+21]

Based on crabon tracker, EnergyVis can provide a dashboard for live tracking of emissions. The drawback, however, is that the current version only includes regions in the US. However, in further iterations, regions in Europe will also be added.

Code Carbon [AKS20b]

The tool Code Carbon can also be used to quantify an execution on a personal computer. The publicly available package is easy to use and has as output a CSV file with the current CO2 emissions. An API that tracks the emissions over time is currently available in alpha version. The emissions are also converted to kilometers driven to allow easy communication to the outside world.

2.3.3 *Current hurdles of Green AI in companies*

While no academic sources have yet been published on the status of Green AI in enterprises, major tech companies like Microsoft are looking into the issue. A blog post [Gup21] by a member of the Green Software Foundation and the founder of the Montreal AI Ethics Insitute provides an assessment of the evolution in terms of capturing the footprint of AI.
Here, hurdles such as providing a workflow-native tooling, introducing certifications, ensuring the comparability and correctness of the metrics, and forming standards are mentioned.

3

METHODOLOGY

The topic of sustainability of artificial intelligence in companies is evaluated with three aligned parts. First, a **literature review** lays the theoretical foundations regarding existing measures and concepts from research and the AI community. In the second part, the current status of the topic in companies is examined following a qualitative approach. For this purpose, **expert interviews** are conducted with individuals who are responsible for the design and development of internal AI applications in their company. The main focus will be on the awareness of the topic, hurdles and drivers as well as a personal assessment of the future development. The last part consists of a small contribution to greener AI. A **benchmark** of the energy consumption of popular machine learning algorithms for classification is produced.

The procedure and motivation for the choice of methodology can be found in this chapter. The peculiarities while conducting the method and the results are in the respective chapter of this thesis.

3.1 LITERATURE REVIEW

The sustainable design of AI systems is a topic that was started in research some years ago. Here, some measures for more energy efficiency in the development of AI are already known and published in papers. In order to evaluate which measures have arrived in the company and to make recommendations regarding them, a collection of measures must first take place.
The systematic literature search is based on [Wat+20] and adapted for this process. First, a research question is formed, which is to be answered in the context of the literature search. Here the taxonomy of the literature can be derived.

> *RQ1: What measures are known for sustainable AI and what is their quantitative effect?*

The company reference of this thesis was neglected here, as a later evaluation regarding applicability will take place as an own scientific contribution in this part. The process of the systematic literature is shown here to provide transparency and enable reproduceable results:

(1) searching for primary studies

(2) filter studies to inclusion criteria

(3) data extraction

(4) data analysis and presentation

In the first step pimary studies were searches. This started with the analysis of linked papers[1] of Roy Schwartz et. al "Green AI", which is known as the popular first attempt to create awareness of the topic of sustainable AI. Here also a key-word driven literature search was conducted in several specialized literature portals for academic search. The keywords were derived of the research question. The following key words and their German translations were used:

1. *measurements* actions, best practices
2. *sustainable* ecological, green, efficient
3. *AI* Artificial Intelligence, Machine Learning, ML, Deep Learning, DL, Statistical Learning, Natural Language Processing, Computer Vision

In a final step of the literature collection, ideas for measures were collected from the AI community on the World Wide Web. It would be a waste not to take up the ideas from the engaged AI community. To ensure the trustworthiness of the source, contributions from corporate websites (mainly in the field of consulting or big tech giants like Amazon, Google and Microsoft) and academic associations were integrated into the collection of primary literature.
The literature found was filtered according to whether measures for greener AI were mentioned or quantified in it. For example, literature related to sustainable use cases was discarded, as was literature that dealt with quantifying sustainability. Some papers mentioned a collection of measures as in [SGM19], [Pat+22], and [Xu+21].
In the data extraction step the measures themselves were singled out in a subsequent step and quantified with research. For this, the referenced works were read completely and the ecological improvement was elaborated. The circumstances of the studies are crucial in this context and are included.
In the data analysis, the measures are finally clustered and compiled in an overview. Here, an assessment is made for the use in the company, in which the criteria (A) time and resource requirements (B) model performance impact (C) model type agnositicity and (D) use case limitations are considered.

3.2 EXPERT INTERVIEW

The corporate part plays an important role in answering the central research question. Overall, the awareness of sustainability in the field of AI will be assessed, the current status in companies will be shown and measures for improvement will be derived. Since research so far has focused on the intersection of AI for sustainability in enterprises and Green AI is only an emerging topic, there is no source material on this. Therefore, an empirical research will be conducted by talking to companies to find out the listed issues. Expert interviews will be conducted for this purpose, which will be

1 Here www.connectedpapers.com was used

analyzed qualitatively.

Although expert interviews are often used in social sciences, [Jol15] and [Dun15] provides examples of how the method is applied in sustainability science and computer science.

3.2.1 *Qualitative Research*

In this section an introduction to qualitative research is given. Quantitative research is particularly popular in the field of data science, as statistical methods are used here to create transferable statements. However, this is not always possible, for example if the amount of data is too small, the time of research is rather early and the question is too complex. Qualitative research addresses these situations well. Nowadays, a number of recognized methodologies and supporting software tools have emerged to achieve qualitative results in this research method as well.
While quantitative research is more about measuring a subject, qualitative research is used to understand a subject. Especially when the sources are poor, a basis can be created on which quantitative work can be built. Also, in case of a discrepancy between the meaning that researchers bring in and the meaning that respondents give to a certain subject, qualitative research is used. Especially with the buzzword terms AI and sustainability, the respondent's understanding can be better investigated in this way. Especially the perspective on sustainability of AI and in this sense the awareness and relevance for the company can be mapped well.

In [Hel11, p. 21ff] the **basic principles of qualitative research** are explained. The following basic principles of qualitative research are important for the execution, text generation and interpretation:

1. The access to the meaning of the interviewee arises in a communication session
2. Respondents should be able to develop their sense of the subject
3. Openness and recognition of unfamiliar views are necessary
4. Reflection of one's own part in the situational understanding process

Especially the unfolding of meaning (2nd basic principle) ensures that an interview should be conducted very openly and narratively. The interviewees should have room to tell what is important for them [Hel11, p.23]. The third basic principle points out that no one's own thinking should be interpreted into others. Here it should be especially noted that the choice of the topic of the paper was done out of personal conviction about the importance and significance of sustainability for AI. In the interview itself, therefore, more attention must be paid to an objective appearance as a communication partner.

The expert interview is a special research form of qualitative research. In the context of the master thesis, an explorative research interview is conducted with the aim of exploring the issue. This type of research is recommended in [Kai14, p. 29f] when no firm theoretical assumptions or robust empirical data exist. Based on the knowledge gained, further interviews can be planned and hypotheses formulated. Due to the initial lack of knowledge, the interviews are not completely structured. A rough structuring can help here to ensure that specific information is drawn from the interview, which otherwise would not be (certainly) obtained. [Kai14, p.31].

At this point, it should be mentioned that qualitative surveys do not provide universally valid results. Statements can only be made within the framework of the sample size and the results are to be understood as an insight.

The framework of an expert interview means the selection of suitable experts, the design of the interview form and a suitable evaluation strategy. This points will be presented in the following sections.

3.2.2 *Preparation of the interview*

Before a guide for a semi-structured interview can be created, it is necessary to break down in more detail the research question that the questions are intended to cover. The following research questions will be answered by the interview:

RQ2 What are the reasons for companies to look deeper into sustainability?

RQ3 What is the relevance of sustainability in the data science department?

RQ4 What are hurdles to a more sustainable use of AI?

RQ5 What has to be done for a resource-saving design of AI in the future?

RQ6 How do expert see the topic developing in the future?

The interview takes place as an online meeting. The interview is whether conducted in German or English. The conduction via a semi-structured way helps to fulfil the second principle of qualitative research. Semi-structured means that there is the opportunity to add and remove questions depending on the course of communication. Here the interviewer is be able to respond to individual particularities and still achieve comparability among other interviews for the evaluation. The interview guide is a help for the interviewer to cover all important points.

To prepare for the question the theoretical knowledge is derived by the literature review regarding measures for sustainable AI in advance and doing the research for the chapter 2.

The created interview guide can be found in the appendix A.1. First, some introductory questions ask about the background of the company in the field of AI and the tasks of the expert. This serves to classify the expert for the evaluation according to the maturity model in table 2.1 and to finally evaluate the expert status. After the introduction, questions about the relevance of sustainability follow. Here, general factors for sustainability are first collected so that they can be evaluated for their relevance to the AI domain. Relevance in evaluating use cases is also discussed at this point.
The next section will then focus on the implemented tools and measures. First, the question of responsibility is clarified, whether companies see this more with the cloud or platform providers or with themselves internally. Here it will also be clarified whether companies also consider criteria such as sustainability in the selection of cloud providers and what responsibility the providers also have for a sustainable design of AI. Questions about implemented measures during model development and resource effort capture will follow. In terms of progress via standardization, it should also be clarified whether developers are individually responsible for the measures.
The final interview section is about the development of the topic and the necessary steps that still need to be taken.

The questions are formulated in an open-ended manner and are intended to stimulate narrative. An exception are questions regarding sustainability as a decision criteria for the selection of use cases and the responsibilities of the developers. Here, the goal is to make sure that these points are covered in the interview.
Since the relevance of sustainability is evaluated by the existence of measures and awareness, the interview also specifically asks about the capturing of the resource consumption of the AI applications as a first step towards more transparency. The results are available in chapter 5.

Selection of experts and realization

For the interview persons are selected who are responsible for the design and framework of AI applications. The person search is done through the social network for business contacts Linkedin. Keywords used for the search were "Head of Data Science", "Head of Data", "(Head of) Machine Learning". Then, the professional profile was checked to see if there was several years of professional experience working with AI. The expert status is justified by the work in commercially active companies, as an undistorted picture of the real situation in companies is to be provided here. For this reason, no experts for Green AI were approached on purpose. However, since the topic was already mentioned in the inquiry, a bias may arise here, since the people who find the topic interesting are more likely to report back.
After the success of the interview request, the interviewees were informed about the rough content in the form of keywords as part of the consent form.

However, there was no explicit preparation of the participants.

The interview is then recorded using the recording function of the CAMTASIA program and is exported as an audio file in MP3 format. The generated file will be used for manual transcription.

3.2.3 *Evaluation*

In this chapter, the methodology behind the data processing and analysis is justified and explained.

Transcription

In order for the interview to be processed further, it must first be put into text form. There are various methods for this. Since only content-related findings play a role here, the text should be transcribed word-for-word without paying attention to emotional signals such as pauses and gestures. Because of this, the transcription is designed according to Dresing/Prehl (2015). The specifics of the used scripting are:

- verbatim transcription with smoothing of double words
- half sentences marked with /
- pauses from three seconds are marked with (...)
- interview person is marked with *I* and interviewee is marked with *B*.
- time indication with minutes and seconds *(m:ss)* after each paragraph
- sections of meaning are formed and presented via headings

The last two items were added to the standards of [Dre16]. The time indication of milliseconds was not used, because this accuracy could not realistically be achieved. The headings of the sections are used as a guidance in the further analysis.
In addition to transcription, removal of the company name, the name of the interviewee, and clear references to the person were also performed for the purpose of anonymization.

Qualitative content analysis according to Mayring

The qualitative content analysis is conducted accoring to Mayring (2014). Therefore the software online tool QCAmap[2] is used. [May14]
In the direction of analysis, the object domain is essential. The topic of sustainability of artificial intelligence in companies is in the foreground and what is said about it by companies. The concretization of the research questions has already emerged in the design of the interview guide. The research

2 https://www.qcamap.org/

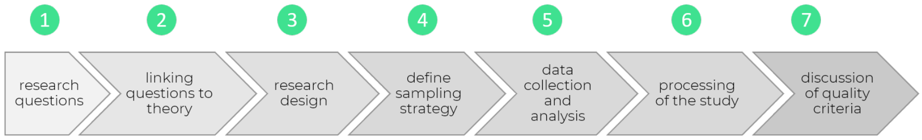

Figure 3.1: 7 Steps of qualitative content analysis according to Mayring

questions are taken up again directly and are entered into the QCAmap software tool.
The second steps would be to prepare for the theory behind the analysis. As for the creation of the interview guide a literature review about the big topic sustainability of AI had already been made.
The third step is to determine the research design. A exploratory content analysis will be performed, because the goal is the derive new categories inductively from the material. [May14, p.11f]. The procedure therefore has an iterative character.
In the fourth step the material is selected. For other types of analysis existing material like newspaper articles are chosen. As for this thesis the material is created, the materials will be the expert interviews which were conducted by the author in April to June 2022.
In the analysis phase the categories are then formed and coded in the text and the occurrence is examined and interpreted. This is made iteratively. After 25% of the interview the category system is reviewed and refined.
After that the results are presented regarding the research questions. Here the categories are described and an answer to the reasearch questions is formulated.
In the last step, the quality criteria are ensured. These consist of transparency, reproducibility and intersubjectivity. Transparency is to be ensured by a description of the person and the importance of AI for the company with the placement in the maturity model. Reproducibility comes from applying the categories to all interviews. If the data set is large enough, the category system should remains stable. The intersubjectivity is reached by publishing sample citations of the categories in the result chapter. There the categories can be understood by readers.

3.3 RANKING OF THE ENERGY CONSUMPTION OF CLASSIFICATION ALGORITHMS

An interesting measure to increase environmental sustainability in the field of AI is the use of resource-saving algorithms. In deep learning, a simple approach is to compare the number of parameters of two models to draw conclusions about an possible advantage in energy consumption.[3] When comparing between classical machine learning methods, the statements about energy consumption are difficult to make due to the diversity of the algorithms' operating principles. Here, a practical benchmark can help to get a

3 The more complex and different the architectures of the models are, the faster this tactic reaches its limits.

feel for the energy efficiency.
From the central research question *What is the awareness and status of sustainability of artificial intelligi oence in companies and how can it be improved?*, a contribution is made here to the second part, the improving sustainability. The formulated specific research questions is:

> *RQ7: Which machine learning method is most energy efficient in training and prediction?*

The structure of the benchmark is inspired by [FD+14]. There, different model types were trained on several data sets to show that there is no model that performs best on every data set. While there the performance is compared, here the energy consumption is in the foreground. The considerations regarding the data, the models including the library used and the measurement of the ecological footprint are explained in the next sections.

Data sets

The used datasets come from the UCI Machine Learning Repository[4]. Since the main focus is on the use of the model in a business context, datasets are considered that carry the label "business". First, classification models are to be tested, which is why the dataset should describe a classification problem. Here, the six largest data sets were selected, which a) had the target variable directly available and b) were readable via a csv format. Multiple data sets were chosen to reduce random effects of the data sets in favor of one model. For example, the neural network could converge very quickly on one data set and the energy consumption of the training would be correspondingly low in this special case. Theoretically, additional data sets can be added. In the evaluation it will be looked in particular whether this is necessary because e.g. the influence of the data sets on the energy consumption leads to very variable rankings.
The data sets were first transferred to an Excel table so that the secure transfer of the data types is given for the automatic benchmark. Since there is no concern about the performance of the models, the data is prepared in the same way for all models and all data sets. Also all available features of the data sets are used except ID columns. The following steps are performed as preparation:

- sort out ID columns
- fill missing values by most frequent ones
- convert text columns via one-hot-encoding
- standardize data set (mean 0, standard deviation 1)
- separating features and target value

4 archive.ics.uci.edu/ml/datasets.php

Since data preparation is the same for all models, it is not included in the measurement of energy consumption. In practice, different preprocessing steps are necessary depending on the model type, which is why it makes sense to include the entire data processing pipeline.

Model types

9 different algorithms are trained on the datasets and used for prediction. Inspired by [FD+14] different classes of algorithms are compared. Besides the pure model type, there is of course also a dependency on the programming language, the implementation in the library, the hardware used and the chosen hyperparameters. In order to eliminate here influencing factors, all model types used are from the python library `scikit-learn` in the version 1.1.1 and accordingly with their preconfiguration of the hyperparameters. These are available via the hyperlinks to the documentation. The models are listed in table 3.1.

Table 3.1: models for energy consumption benchmark.

model family	**model**	**remarks**
Logistic Regression	LogisticRegression	
Naive Bayes	GaussianNB	
Nearest-Neighbours	KNeighborsClassifier	
Tree Methods	DecisionTreeClassifier	
Random Forest	RandomForestClassifier	100 decision tree estimators
Boosting	AdaBoostClassifier	50 decision tree estimators
Bagging	BaggingClassifier	10 decision tree estimator
Support Vector Machines	SVC	rbf kernel
Neuronal Network	MLPClassifier	1 hidden layer with 100 neurons

Measurement of the energy consumption

To be able to measure the energy consumption of the models, the software tool `codecarbon` is used. In chapter 2.3.2 this and other possible tools for quantification of energy consumption and emissions are listed. This tool was chosen because it is very easy to apply within only a few lines of code and breaks down the power consumption to process level. Since the execution is done on a personal computer, this is desirable for the usability of the results. The hardware used is a CPU *Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz* with 8 physical cores and 32 GB RAM.
To minimize the influence of the number of data points and features on the power consumption, the datasets are normalized by dividing the power consumption by the number of rows $\times$ number of columns.

4 POSSIBLE ACTIONS FROM RESEARCH

The measures found following the procedure in chapter 3.1 are clustered into categories: (1) hardware and ressources, (2) model selection, (3) model training (4) model operation and (5) organizational. Each category starts with a tabular listing of the measures, their stated effectiveness from research, and possible limitations in corporate use. In short explanatory texts, the measures are explained as well as the circumstances of the measurement of the resulting sustainability effect. Also the considerations regarding a business use will find a place here. Which will be assessed with the criteria (A) time and resource requirements (B) model performance impact (C) model type agnositicity and (D) use case limitations.

It should be noted that the factors to assess the improvement or effectiveness of the measure depend very much on the starting point. For example, sometimes simple methods are compared with state-of-the-art methods e.g. the execution on a Central Processing Unit (CPU) with an advanced Tensor Processing Unit (TPU). As research progresses, old specialized methods are usually compared to newly developed methods (e.g. a current widely-used Graphical Processing Unit (GPU) is compared to a newly developed TPU). The improvement from the already better method to an even better method can therefore be smaller, since much has already been optimized here. Therefore, in the explanation of measures, the corresponding backgrounds are pointed out, which should be considered besides the table for full understanding.

4.1 HARDWARE AND RESOURCES

If AI is to be made more resource efficient, the considerations start with the hardware and infrastructure, regardless of the later models. Table 4.1 provides an overview of the measures identified and their impact on sustainability. Overall, the independence from the chosen algorithms is a great advantage of this category. However, the applicability depends on the flexibility of the company (e.g. use of a public cloud).

Specialized hardware

Especially in the area of deep learning, when a lot of data and large models are in use, the selection of the processing unit can make a big difference. According to a recent Google paper [Pat+22], using optimized processors like modern GPU or TPU instead of CPU,which is a general purpose chip and not optimized for model training, leads to an increase of the energy efficiency by the factor $2 - 5\times$.
A previous publication from Georgetown University's Center for Security and Emerging Technologies of 2020 summarizes 18 studies from 2011-2019

Table 4.1: Collected measures for hardware and infrastructure selection

measure	improvement	limitation	references
CPU $\rightarrow$ specialized chip	energy efficiency $3 - 368\times$	(costs), availability	[Pat+22; KM20]
GPU $\rightarrow$ advanced GPU	less energy consumption $1.2 - 10\times$	(costs), availability	[SV22; Pat+22]
choose public cloud	higher PUE $1.4 - 2\times$	data protection	[Pat+22]
energy-aware scheduling in data centers	higher PUE $1.1 - 3\times$ $2.5\% - 36.6\%$ less emissions	use case requirements, long jobs	[McD+22; Wie+21]
cloud provider choice	up to minimum emissions	business contracts	[SGM19]
cloud location choice	$5 - 12\times$ less emissions	data residency	[Pat+22; Hen+20; SGM19; Tor+21; JS19]
power capping the hardware	energy usage decrease $10 - 45\%$	higher computing time	[McD+22; Fre+22]

comparing CPU, GPU, TPU and Field-Programmable Gate Arrays (FPGA) and Application specific Integrated Circuits (ASIC) with each other in model training and interference for different deep learning architectures. Modern chips, meaning advanced GPU, TPU, FPGA and ASIC, with energy-saving transistors, better parallelization logic and special design for AI calculations could reach energy efficieny factors ranging from $3 - 368\times$ compared to a CPU. Additionally, the energy efficiency win comes with an speed advantage ranging from $2 - 500\times$. [KM20, p. 38ff] However, it should be noted that the main application here is state-of-the-art AI research with hugh models, which may not be necessary for businesses.
There are also differences when comparing newer and older chips of the same type. In a Neural Machine Translation deployment, two types of GPU[1] were compared. Depending on which deep learning model type was used, the energy consumption could be reduced by a factor of $1.2 - 10\times$. [SV22] For a comparison of the other chip types with each other, reference is made here to the summary of [KM20] and the studies cited there.

LIMITATIONS IN THE APPLICATION IN COMPANIES The applicability in the company benefits from the independence of performance and usually also of use cases (e.g. not possible for embedded systems applications with given chip). Since the chips are primarily designed for deep learning, there is no direct advantage for all models. Specific libaries like `RAPIDS`[2] allow the use of a GPU for classical methods as well. The time savings also advocate the use of specialized hardware. However, modern chips initially cause an investment expense and they are not always available on the market due to a high demand because of Bitcoin mining and others. When using instances in

1 4 NVidia GeForce 1080Ti with 11GB of vRAM (user-class GPU) vs. 3 NVidia Tesla P100 (for workstations) with 16GB of vRAM
2 docs.rapids.ai/api/cuml/stable/

the public cloud with those processing units the costs are higher than CPU instances and the availability is not given for every location. However, the cost aspect can be countered with [KM20], which says the cost-effectiveness improves greatly when using specialized hardware.

Choose public cloud over on-premise

Due to the higher utilization of servers in public cloud data centers and the usually higher efficiency, energy savings could be up to $1.4 - 1.5\times$ [Pat+22]. The efficiency of data centers is measured by the Power usage effectiveness (PUE) which is globally 1.59 [McD+22]. Meaning approximately 40% of the energy is not used for the actual computer equipment.

$$PUE = \frac{\text{IT energy} + \text{facility energy}}{\text{IT energy}}$$

Google reported a PUE of 1.1[3] in their data centers for comparison.
Further, the total amount of servers for all customers is lower because of the economics of scale which results in an additional sustainability gains. And following up the previously mentioned measure, specialized hardware is immediately available.

LIMITATIONS IN THE APPLICATION IN COMPANIES The use of a public cloud (at least for machine learning training) is associated with reasonable effort in the opinion of the author. The cloud providers advertise cost savings through economics at scale and pay-as-you-go payment plans. A critical point, however, could be internal or governmental regulations regarding data protection and data residency, which do not allow going into the cloud.

Intelligent scheduling in data centers

The PUE of a data center is actually not a stable value and varies due to the influence of factors such as server load (heat generation) and outside temperature (need for cooling). In figure 4.1 this variation is shown for the institutional data center of the MIT Lincoln Laboratory. Temperatures in summer (A) and the temperatures in hot afternoons in summer (B) are affecting the PUE.
A naive approach would be to move the scheduling to a time when the PUE reaches a low value. During summer the effect from moving a short-running job from day to night is estimated with 10% lower PUE and moving long-running jobs from summer to winter is estimated with a 33% reduction. [McD+22] Smarter methods can exploit further fluctuations in PUE evolution and dynamically incorporate that into the scheduling of machine learning processes. However, when scheduling at night, it is important to note that

3 www.google.com/about/datacenters/efficiency/

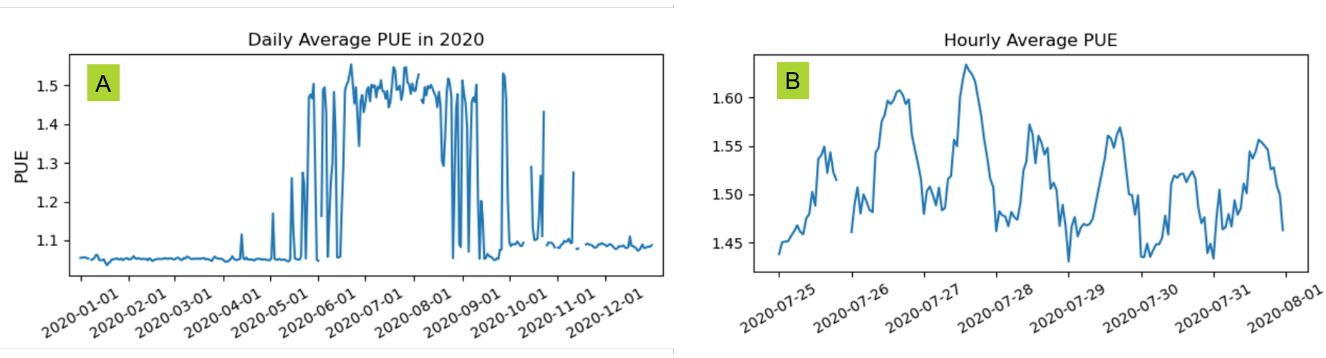

Figure 4.1: Measurement of the PUE of a data center daily average in a year (A) and hourly average in the summer time (B). Taken from [McD+22]

the energy mix at night is sometimes less green due to the absence of solar energy and other factors (see Figure 2 of [Pat+22]). In [Wie+21] the factor time and energy mix are brought together for Germany, France, Great Britain and California. Here the weekends are especially recommendable for the lowest overall carbon intensity. Possible shifting strategies were also evaluated. Depending on which strategy (interruptable machine learning, +-2/8 hours shift) is chosen the carbon savings could be 2.5% – 36.6%.

LIMITATIONS IN THE APPLICATION IN COMPANIES In the simplest form, the adapted scheduling of the models, using low temperature at night and winter time, does not require elaborate systems. Otherwise the effort of determining the current PUE and integrating it into the scheduling would be incurred. However, the requirements from everyday business cannot always provide the necessary flexibility. For some models, real-time requirements have to be realized or competitive advantages cannot be only advanced in winter. In addition, very complex model training may take several days and will be executed during the hot afternoon temperatures anyway. But as shown in [Wie+21] there is high number on workflows which can be shifted with simple requirements as 'finished until the end of the workday'. However, especially retraining or making batchwise predicitions can usually be moved well into the night.

4.1.0.1 *Cloud Provider Choice*

The choice of cloud services provider has an impact on which energy mix is used. For example, in 2018 Amazon Web Services used only 17% renewable energy overall, while Google was already using 56%. [SGM19] However, energy mix numbers are constantly changing, and some smaller cloud providers offer 100% renewable energy deployment and compensate the rest with emissions trading. In addition, it must be weighed whether the current status is important or the provider's sustainability goals.

LIMITATIONS IN THE APPLICATION IN COMPANIES The decision in favor of a cloud provider is usually made centrally in the company in order to secure special conditions and to build up the necessary cloud platform knowledge. Depending on the flexibility it may be possible to use multiple cloud providers. Furthermore greener cloud service providers may come with more costs and may not reach the PUE of larger providers.

Cloud Location Choice

The actual energy mix is primarily dependent on the location of the executing data center. In [Pat+22] the factor $5-10\times$ is given, which is also mentioned in past works of the authors. For the year 2017 the paper [Hen+20] discovered even a difference of the factor 30 between the cleanest available cloud location in Quebec and Estonia. The influence of the cloud location is also examined in [Tor+21]. According to this, the AWS Frankfurt location would cause $12.2\times$ more CO_2 equivalent than the AWS Stockholm location. In addition to general considerations about which locations are basically ecologically preferable, the decision can also be made dynamically by incorporating carbon intensity data in the decision process of a Kubernetes scheduler. A paper on this can be found in [JS19].

LIMITATIONS IN THE APPLICATION IN COMPANIES The selection of the cloud location is a measure that is theoretically easy to integrate, provided that no internal regulations or special restrictions on the geographical location of the data processing speak against it. The latency for the data transfer also does not have to play a role for the use case. Even for real-time models, at least the training can be moved to a sustainable location. Overall, it is to consider, that even green energy has some emissions e.g. energy from photovoltaics generated a total footprint of 3,055 kt CO_2 equivalent in Germany during the year 2018. [LMS19, p.33ff] A footprint remains, but it is very minimal. In [Wie+21] it is also shown, that the carbon intensity (gCO_2 per kWh) can very strongly (Germany) or weakly (France) over the year. This is to be considered.

Power capping the hardware

GPUs like V100 offer power capping possibilities, which enable energy savings. If the power is capped at the found optimal 200W (instead of 250W in normal mode), energy savings of at least 10% occur for all models without significantly increasing the training time. At 100W, greater energy savings are possible (25-45%), but the training time is affected (35-40% less speed).[Fre+22]
In [McD+22], power capping of 150W was also used for natural language models. On average, the reduction in total energy was 13.7% and the increase in training time was 6.8%. For the language model BERT, the energy gains from power capping are much greater in the interference, which is more relevant for commercial use. Here, power capping at 150W can reduce energy consumption by 24.2% and interference time increases by 22.7%.

LIMITATIONS IN THE APPLICATION IN COMPANIES Especially power capping during interference can reduce the footprint of the operated AI for companies. In computer vision and natural language processing, operation on a GPU is common, so power capping can be set up without much effort via the GPU settings.

4.2 MODEL SELECTION

Models are usually selected based on criteria such as explainability, required accuracy and amount of available data, and other process-critical requirements such as a real-time requirement. However, some guidelines can also be established with regard to sustainability.

Table 4.2: Collected measures for model selection

measure	improvement	limitation	references
usage of simpler models	less training and interfernce time	accuracy limits, feasibility	[FD+14]
consider scalability of models	different energy consumption overhead	limited choice	[Fre+22]
usage of optimize models	faster computation $2.5 - 44\times$	model types, technology radar, implementation effort	[Pat+22; So+21; Tam+20; ALV14; HB20]
compact architecture design	less energy consumption of model, shorter runtime	accurracy loss, effort implementation, complexity	[Xu+21]
pretrained models	less training time & data amount	applications limited	[SS21; Ber+16; Rad+21; LNH09]
Selection of libraries	less training time $25 - 32\%$ less interference time 65%	employees knowledge, implementation effort	[Chi+21; SH19]

Usage of simpler models

The use of simpler models is mentioned here for completeness. Especially in business when a structured database is used, classic machine learning models should first be used before deep learning is applied. Complex models like neural networks do not necessarily have to have the better performance. In a study of 121 classification data sets with 17 classes of model types, for example, random forest models prevailed overall over neural networks.[FD+14] For automatic mode selection see the measures for automated machine learning in 4.3.

LIMITATIONS IN THE APPLICATION IN COMPANIES If structured data is processed, the best practice is already to begin with simple models. More complex models also tend to overfit the training data. Simpler models can yield better performance in such cases. However, for complex contexts, more powerful models are often required to realize the use case at all.

Consider scalability of models

In [Fre+22], the scaling of deep learning models was tested for distributed training. The resulting energy consumption overhead was measured and varied greatly by the type of model used (computer vision vs. geometric deep learning vs. natural language processing). When scaling from 2 to 16 GPUs the energy requirement of training DimeNet[4] is higher by a factor of 3.6. For other models, including other geometric deep learning models, the factor is between $1.1 - 1.2$. The scalability of the models should therefore also be of interest in model selection.

limitations in the application in companies Here, two aspects stand in the way of the application. On the one hand, the information about the scalability of the models must be known, and on the other hand, the choice between several models that are otherwise comparable must be given. To obtain the scalability information, comparisons must be made. Additionally, limiting the number of GPUs can also be a measure.

Usage of optimized models

Optimized model architectures result from subsequent research work after the base model is published. They are produced by applying efficient architecture changes, performing a neural architecture search or applying common optimization techniques due to application limitations. This not only bring a performance gain, but often also a time gain. Especially in this section, the listing of model improvements is inexhaustible. The selected examples are therefore only intended to provide an estimate for the measure.
As an example for efficient machine learning architectures, sparse models help to reduce the overall computation time by $5 - 10\times$. This is enabled by sparse activation, that only uses a small fraction of the total architecture for a task. As an example, the sparse version Generalist Language Model (GLaM) (2021) with more parameters than the predecessor GPT-3 (2020) was trained $2.8\times$ faster and also achieved a higher accuracy. In interference only 8% of the model parameters are active which results in a runtime improvement. [Pat+22]
The Primer model (2021) is a successor of the original Transformer model of 2017 and found by a neural architecture search. It is $4.2\times$ faster. [So+21]
A common reason for model optimization is also required latency and planned deployment on resource-constrained platforms such as those used in the Internet-of-Things domain. Here an accuracy loss is more acceptable. As an example, EdgeBERT (2021) enables energy savings between $2.5 - 7.5\times$ per sentence in interference. This is achieved through entropy-based early-exit, adaptive attention span, network pruning, and floating point quantization. The accepted accuracy loss is adjustable for $1\% - 5\%$. [Tam+20]
In [HB20] are the efficiency factors for famous models noted, when they are

4 a unoptimized geometric deep learning model used in chemistry and materials science

replaced by an optimized model. Here a reduction by the factor of 44× could be reached by replacing AlexNet with EfficientNet.
Even if deep learning models are mainly listed here, there are also further developments of classical machine learning methods such as tree-based methods. See [ALV14] for this.

LIMITATIONS IN THE APPLICATION IN COMPANIES More optimized models usually lead to efficiency and/or performance gains. In companies, however, the adoption of the optimized models faces several challenges. First, the optimized models must be known, which is a matter of internal technology radar. Furthermore, the models must be applicable in some form through manual implementation (additional effort) or programs and libraries used (library version dependencies!). Should an old model be replaced, more effort goes into testing and a parallel provisioning. If possible, however, the first "raw" version of the state-of-the-art models can be skipped and the next iteration step can be waited for.

Compact architecture design

The techniques already used in optimized models such as EdgeBERT can of course also be applied when building a deep learning models from scratch. Due to the multitude of possibilities, only an overview is given in figure 4.2. In [Xu+21, p.9-17] all measures are explained accordingly.

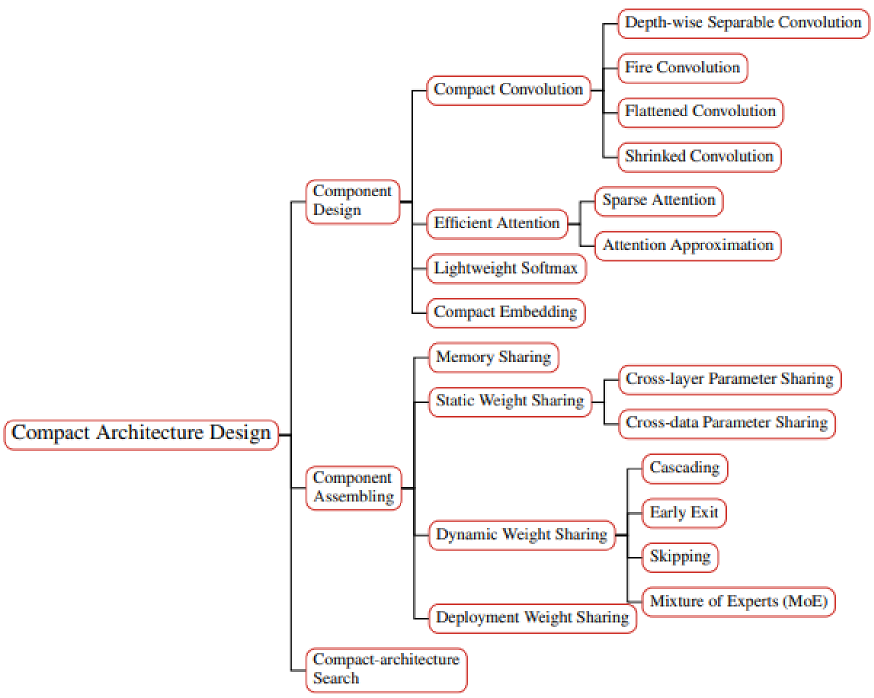

Figure 4.2: methods for compact architecture design. Taken from [Xu+21]

LIMITATIONS IN THE APPLICATION IN COMPANIES The presented measures are only suitable for deep learning where the design of an architecture plays a role. The implementation of the measures naturally requires an effort and the technical know-how of the model creators. The additional complexity introduced also requires in-depth knowledge of deep learning for further enhancement of the model. However, the implementations are mostly available with sample code on GitHub. Some measures like the *early exit* are associated with a potential lower accuracy [Tam+20].

Pretrained models

For complex projects in the field of natural language processing and computer vision, pre-trained models are a very good way to reduce the training time and the required amount of data and still achieve good results. State-of-the-art architectures are trained on very large data sets and extract the general knowledge there and convert them into their parameter values. Later, they are adapted with further training for fine-tuning on specific business data e.g. language models such as BERT and RoBERTA only need to learn specific terminology. This procedure is also known as transfer learning. An overview of different methods can be found in [Wan+19].
The sustainability factor includes savings in training time as well as using less data. The impact is immense. In [MS20], a pretrained RoBERTA model is used for the SNLI dataset. Only about 2% of the data are necessary to get the same performance as a fully trained model. It should be noted here, though, that this depends heavily on the new context and its similarity to the orginal training data.
Extreme concepts in transfer learning are one-shot-learning [Ber+16]and zero-shot-learning [LNH09]. For one-shot-learning, the reduction is not based on efficiency but from use case requirements. As an example, an automatic passport control should be able to detect a person only by using a single image.
Zero-shot learning is about generalizing a model to classify unseen classes without explicit training with labeled data. For example, in [Rad+21] natural language supervision is used to use the textual descriptions of the images (no labels) to describe the new classes. Training data and descriptions must therefore also be available here.
The various methods all have in common that they save a specific costly training from scratch. It should be noted for the sake of using simpler models, that the extracted basic knowledge not only comes in huge models. In [SS21] a model three times smaller than GPT-3 achieves the same performance after a few-shot learning.

LIMITATIONS IN THE APPLICATION IN COMPANIES In fact, using pretrained models leads to a time gain and also the amount of processed data points decreases. Since state-of-the-art models are used, performance is usually very good. From these advantages, the model-as-a-service business line is evolving. One-shot and zero-shot learning enable new use cases to be implemented efficiently. The only disadvantages are that the pretrained models have to be available in the plattform or library used and the application has to be suitable to use it.

Choice of libraries[5]

Also impacting execution time and energy consumption is which library is used. While general statements are to be avoided here, there are however studies that benchmark the same model architecture in `Tensorflow` and `PyTorch`. Here, the execution time in `Pytorch` was 25% shorter in training and 65% shorter in interference compared to `Tensorflow 2.0`. However, the accuracy of the `Tensorflow` model was better by 1.13%. [Chi+21]
This can be confirmed with another study. In [SH19], the execution time of `Pytorch` in training is 35% shorter for the same accuracy (exact value not given).
Two things are noted about this finding. First, the execution time does not allow direct conclusions about the energy consumption.[Abd+15] An extended investigation here would be of interest to assess the actual sustainability of the measure. Second, the difference varies depending on the implementation of the network architecture. Although in general the core of `Tensorflow` is written in C++, while `PyTorch` uses C and Lua in its core. In a study on the energy efficiency of programming languages, C is always ahead of C++ in terms of energy consumption and execution time. [Per+17] A consistent difference would thus be explainable.

LIMITATIONS IN THE APPLICATION IN COMPANIES The choice of the library has to do with several factors. In [Par+17] many of the deep learning frameworks are compared. Here capability, available platforms, interfaces and support of pretrained models are mentioned as factors. In the business context, the knowledge of available developer resources is added to enable the development and maintenance of the deployed models without facing any problems. A switch of the main programming language is associated with educational effort. To what extent the performance is influenced must still be found out with further studies.

5 The measure of program library selection fits into several categories presented here. Since the decision affects which models are available, the measure is placed here.

4.3 MODEL TRAINING AND AUTOMATED MACHINE LEARNING

An efficient and resource-saving training process of a machine learning model eliminates unnecessary resource consumption. Hyperparameter tuning, which is the process of finding optimal parameters to fit the data, is particularly significant due to the execution and comparison of a large number of model variants. Similarly, the evaluation of multiple models, available features and data preparation pipelines in Automated Machine Learning (AutoML) is particularly resource intensive. For this, the measures are mainly taken from the paper [Tor+21], which explicitly presents measures for greener AutoML.

Table 4.3: Collected measures for model training and automated machine learning

measure	improvement	limitation	reference
prefer other search strategies than grid-search	reduced runtime 1.6×	potential accuracy loss	[Lac+19a; LL19]
warmstarting	faster to fixed accuracy zero-shot: less runtime by 1972×	accuracy loss, information required	[MWH18; Van18; Sin+21]
avoid time-outs with time-out prediciton	reduce wasted CPU time 2×	implementation effort	[Tor+21; Moh+21]
multi-fidelity performance measurement	less cpu time for fixed sized calculations $2.5-53\times$ faster time to result level $10-100\times$	performance loss, available implementations, rebound effect	[Pet00; FKH18; Kle+17]
energy consumption as objective	t.b.d.	no implementation yet, research to be done	[Tor+21]
exploit heterogeneous hardware resources	potentially lower energy consumption (t.b.d)	information base needed, implementation effort	[Tor+21]

Prefer other search strategies than grid-search

If a hyperparameter search covers a small parameter space, grid search, i.e. the systematic search of all combinations, can be a suitable choice. For large parameter spaces, the approach is not recommended due to resource intensity. In [LL19] a grid search, a random search and a genetic algorithm (for deep learning) were investigated. All three methods achieved the same accuracy, while random search was 1.6× faster than grid search.

LIMITATIONS IN THE APPLICATION IN COMPANIES The faster retrieval of the candidate is in the interest of the company due to the associated time savings. In the popular Machine Learning (ML) framework `sklearn` the random search is already implemented and genetic algorithms can be realized via code examples on the Internet with the help of the Python package `scipy`. Here the effort is a bit higher.

Warmstarting

Warmstarting means that knowledge gained from previous executions is incorporated into the current execution. The optimization process does not start from scratch. This helps to find a good candidate earlier. Here metalearning techniques [Van18] are used. Statistical characteristics of the data set and the performance of models on similar data sets are the basis for this. A simple idea of the library `ML-Plan` [MWH18] is the hierarchical planning of the execution depending on how well the algorithms perform in general. In `auto-sklearn` a static portfolio is used for this.
In the extreme case of zero-shot AutoML [LNH09] the pre-information leads to the suggestion of only one candidate and its acceptance by the system without further evaluation. For this, textual descriptions of the dataset and computed meta-features are used. In the paper a benchmark between common libraries and the approach was performed. The libary `autosklearn` needs $295.94s$ in average to find a good candidate for the selected datasets. The zero-shot approach only needs $0.15s$ while the average accuracy decreases from 0.88 (autosklearn) to 0.87. Warmstarting and the extreme case zero-shot-AutoML thus have a great effect.

LIMITATIONS IN THE APPLICATION IN COMPANIES Simple warmstarting techniques are already integrated in the known libaries for AutoML. Advanced methods such as the zero-shot method presented are only available as code and have only been tested on simple data sets that are not representative of business data. Since a lot is currently happening in research, new features in the known libraries should be kept an eye out for.

Avoid timeout

In [Moh+21] it was discovered that 20-60% of the CPU time is lost in automated machine learning because the algorithm reaches a time out during the evaluation. By predicting the runtime of an algorithm within a so-called safeguard component, a time out should be prevented before execution. The lost time can be reduced by 49% with this approach. This is particularly recommended for resource-intensive problems such as predicitve maintenance. Similar methods for runtime prediciton of algortihmes and piplines can be found in [Tor+21].

LIMITATIONS IN THE APPLICATION IN COMPANIES The runtime prognosis in the safeguard componente can be a meaningful addition of AutoML projects. A corresponding implementation in known AutoML libraries was not found. Therefore the measure would come with additional effort for the saving of computation time.

Multi-Fidelity performance measurement

The idea of multi fidelity performance measurement is to compute the relative performance of a potential candidate in automated machine learn-

ing by using easily computable functions instead of costly cross-validation. [Tor+21] There are a lot of available techniques and optimizers for AutoML. In [Pet00] effective subsampling is used. The paper evaluates the technique with 35 datasets of different length and 12 algorithms to choose from. In total the CPU time to find the best algorithm could be reduced by the factor $10 - 53\times$. The rate of selecting the best of 12 algorithms depends on the subsample size: The rate of using a fixed size train and test dataset with 1,000 data points[6] each is $0.55 - 0.69$. Even if all the data is used (divided in a train and a test dataset) rate goes up to $0.86 - 1.00$. This 2-fold-cross-validation saves resources compared to the usage of a 10-fold-cross-validation while having no effect on the order of algorithms for data sets with more than 10,000 data points.
Of course, there are more advanced methods to find the best candidate in a cost-effective manner like (adapted) Bayesian optimization and bandit-based methods or the combination of both (BOHB)[FKH18], which are used for AutoML as well as for hyperparameter optimization of a single model type. The time saved in evaluation is hereby great. If the results of BOHB and a random search are compared, BOHB can reach the result faster by the factor $55\times$. Also, the method "Fast Bayesian Optimization of machine learning Hyperparameters on Large Datasets" (FABOLAS) claims to be $10 - 100\times$ faster than 2017 state-of-the-art Bayesian optimization methods and the bandit strategy Hyperband.

LIMITATIONS IN THE APPLICATION IN COMPANIES In particular, lowering the k-fold-cross-validation is an easy-to-use measure, especially for very large data sets. A follow-up study on more recent algorithms could further prove its effectiveness. In the AutoML framework `auto-sklearn 0.14.7` the splitting into training and test data (67:33) is already stored as default value.
The key takeaway from this measure is that there are any number of complex methods and the factors speak very much for themselves. However, the rebound effect must be taken into account, since the speed advantage does not necessarily lead to less energy consumption and thus CO2 emissions. With the rebound effect the time is just used to search further checks more options for the optimal algorithm.

Energy consumption in objective function

One idea is also to incorporate energy consumption directly into the automated machine learning search algorithm. Here, the expected improvement of the pipeline and the expected energy consumption of the execution can be opposed. This approach would be feasible with an adapted optimization (Bayesian or Hyperband). The estimation of the energy consumption of a pipeline is a problem that is currently being researched.

6 The minimum size of a dataset was 2,000 data points.

LIMITATIONS IN THE APPLICATION IN COMPANIES Due to the fact that no implementation from research is available yet and no estimation of possible performance and sustainability trade-offs is given, the measure is not recommended for use in companies for the time being.

Exploit heterogeneous hardware ressources

The basic idea, mentioned by the authors of [Tor+21], is that the heterogeneous candidates run on the most suitable processor according to their respective hardware-dependent energy consumption. For example, a GPU is much better suited for neural networks than a CPU. The corresponding types should be available for selection in the execution cluster.

LIMITATIONS IN THE APPLICATION IN COMPANIES The method has only been proposed by the authors and has not yet been evaluated. Accordingly, an information base must first be created in research or in the company, which examines the energy consumption of the models on different hardware. The implementation of the efficient allocation of the candiates to the hardware must also take place in order to be able to draw any conclusions at all about a potential benefit. Presumably, the further drive will come from research, since the incentive for companies to achieve unquantifiable cost savings through lower energy consumption is simply too small, and large clusters with specialized hardware are not available to every company to this extent.

Conclusion

In conclusion, one recommendation can be made regarding automated machine learning for everyday business: Attention should be paid to always using the latest versions of the libraries, since corresponding findings from research are implemented there fairly promptly. Knowledge of the various adjustable parameters and their effect on the computing time should be introduced by training the developers regularly.

4.4 MODEL OPERATIONS

Since the operation of models is the main concern in companies, measures in this area have a great deal of potential impact. The operation section describes everything that happens after model selection and the first successful training. Special focus is here on the optimization of retraining of the models and the compression for shorter interference time after deployment.

Table 4.4: Collected measures for model operations

measure	improvement	limitation	references
retraining after concept drift	less retrainings	implementation effort	[BAK22]
distiallation	reduced runtime $1.8 - 10\times$	small additional effort, only ensembles and deep learning	[JHW22; HVD15]
quantization	lower energy consumption $4 - 6\times$	accuracy loss, only deep learning	[SV22; Jac+17]
other compression tactis	faster runtime $1 - 4.4\times$	accuracy loss, libaries	[Tha+20]

Retraining after concept drift

In a machine learning system operated long term, the up-to-dateness of a model must be maintained. Re-training is therefore necessary to integrate fresh data into the model. Instead of retraining at fixed intervals, the necessity can be determined by detecting so-called concept drifts. There are data-based and performance-oriented detection options for this. [BAK22] Overall the number of retrainings could be reduced.

LIMITATIONS IN THE APPLICATION IN COMPANIES Of course, the retraining after concept drifts is dependent on monitoring the data or the model. Since in many companies both are already monitored, this can be linked to. Depending on the selected lower limits for the model performance, it is possible to control how much performance loss is acceptable. The measure is otherwise applicable for all model types and different strategies can be applied for gradual, cyclical and abrupt changes of the trends.

Distillation

Knowledge distillation compresses the knowledge of a large model into a small one, or the knowledge of an ensemble with many predictors into a single one. Here, the output of the teacher model is used to instruct the student model. In research, an ensemble with 10 identical models could be distilled into a single model with no change in performance. [HVD15]
In another study [JHW22] a state-of-the-art distillation technique is used to train a small student model on the knowledge of a bigger neural machine translation model and here reducing the runtime by factor $1.8 - 1.9\times$.

LIMITATIONS IN THE APPLICATION IN COMPANIES By this second-level learning step the model size can be reduced strongly without causing considerable restrictions of the performance. In some cases the performance even increases. The application is suitable for ensembles or (strongly regualarized) neural networks. Many code examples are available on the Internet for different framework like `keras`.

Quantization

Quantization is a low-effort compression techniques, which approximates the parameters of a deep learning model. These are usually stored as 32-bit floating point numbers and are approximated with lower precisions (e.g 8-bit integers). In [SV22] the results for a quantizied transformer model (float32 $\rightarrow$ int8) could achieve an power consumption reduction by factor 6. Libraries for specific usage like `CTranslate2`[7] and general deep learning libaries such as `pytorch` already support this measure. It should be noted, that the model performance is decreasing by applying this technique.

LIMITATIONS IN THE APPLICATION IN COMPANIES The technique is easy to use and many deep learning libraries already support it. By selecting the precision (float16, float8, int8) the performance loss is at least adjustable. This can be further reduced by quantization aware training [Jac+17]. There a quantization is already simulated during the training and flows into the adjustment of the parameters. This weakens the performance loss.

Other Compression techniques

Compression techniques like the presented quanitzation and distillation are applied after the successful training of the models and intent to reduce the model size or lead to a faster execution. There are more tactics available in research: Pruning, low rank matrix factorization, tensor decomposition, weight sharing and dynamic skipping of RNN state updates as examples. In the paper Thakker et. al. (2020) a new compression tactic Hybrid Matrix Factorization HBF is introduced and compared with other techniques. See figure 4.3 as an example of one of four evaluated application scenarios. The range of runtime improvement with the compression techniques is $1 - 4.4\times$ depending on the compression rate. The proposed tactic HBF is the best compromise regarding speedup and accuracy loss.

LIMITATIONS IN THE APPLICATION IN COMPANIES Balancing compression against loss of accuracy is the biggest challenge in applying this measure. Practical comparisons will probably have to be made here to make the choice. The presented hybrid matrix factorization has not yet been implemented in any known library and the code is also not available, which makes the method unattractive for companies. Furthermore, the compression methods are not applicable to all model types.

7 github.com/OpenNMT/CTranslate2

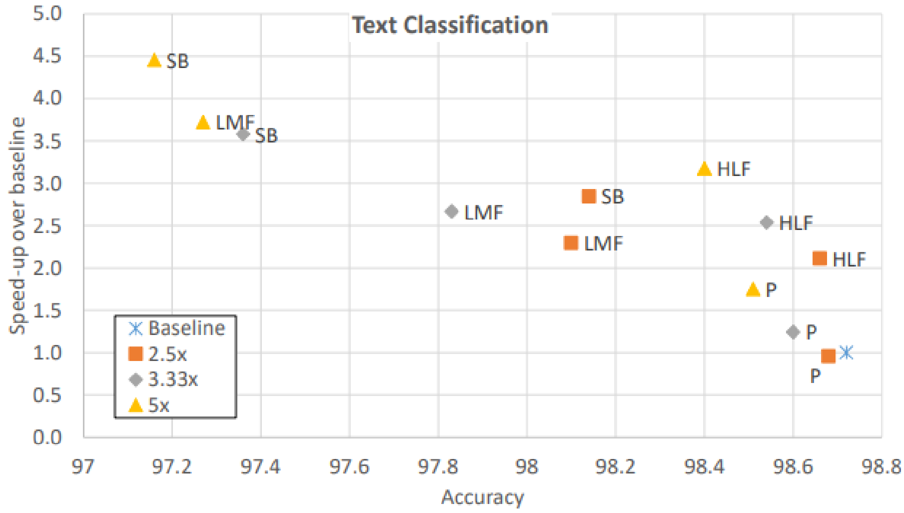

Figure 4.3: Comparison of the compressing tactics hybrid matrix factorization (HBF), pruning (P), low rank matrix factorization (LMF) and train a smaller baseline model (SB). The accuracy and speed up depending on the compression rate is displayed. Taken from [Tha+20]

4.5 ORGANIZATIONAL

Finally, there are some organizational measures help to make AI systems sustainable. Since quantification of the measures is not directly possible, this section is more for awareness and discusses the applicability directly.

Table 4.5: Collected measures on an organizational level

measure	limitation	references
evaluation of use-cases	complex, monetary interest prevails	[Lig+21]
define needed accuracy	dependent on application, just estimations and less existing data	[Dev+19]
track emissions	complexity	[Lac+19a]
shared models in community	business secrets	[Che+22]

Evaluation of use-cases

When evaluating potential use cases, the environmental, social and ethical consequences of the application of AI can be assessed. In [Lig+21], the life cycle assessment methodology is applied to the data science domain. The footprint of AI is systematically compared to the environmental gain. In the company itself, the approach is not easy to implement due to its great complexity. For example, it is difficult to estimate the footprint of the AI in advance, as well as the possible energy gain. Even in sustainability-promoting application ideas from the academic research only 9.5% publish quantified

data on the actual energy gain.[Lig+21] The search for comparative values is therefore challenging.
However, the point "environmental impact" can be included in the discussion list as a first step when evaluating new use cases and rough estimations could be made. The trade-off against financial interests also needs to be discussed.

Define needed accuracy

Often a baseline is set at the beginning of the project, which is considered to be the minimum requirement for the model. Further target values can be defined in the course of the project. On the one hand, a necessary minimum *improvement* can be defined here, which, if not met, leads to a non-deployment of the model. On the other hand, it can be decided in the course of the project how much an increase in performance is worth. For example, in [Dev+19] the number of training batches was doubled for a 1% performance increase when introducing the language model BERT. A siginficant increase in training time or model size should therefore always be balanced. Domain knowledge should clarify the necessity of a certain performance increase in terms of the caused costs transparently for all stakeholders and developers.

Track emissions

One of the first steps is to consistently record and measure the CO_2 equivalent so that informed decisions can be made. The number of existing methods in form of libaries and web-interfaces such as [Lac+19a] make it easy for companies to approximate emissions. Though, automated collection still needs to be developed and integrated into existing systems. In addition to precautionary captures for informed decision-making, interpretable values have the power to influence behavior.

Shared models in community

An often mentioned idea is to establish an AI community where models are shared. A very specialized, optimized model for e.g. customer churn can reduce the repeated model training effort in individual companies. Here, however, the concerns for giving up business benefits are very high. However, since the open-source idea with free commercial use and joint further development is very established in IT, initiatives can be started here. In the enterprise, obtaining a model from a provider specialized in the use case can also be a solution.

4.6 DISCUSSION

The measures found with the help of the literature search could also be quantified for the most part via the associated publications. Many measures were discussed by several sources and the individual values of the studies are mentioned in the description of the measure. The corresponding range of improvement in the tables should therefore be seen as a rough guide value. For some measures, such as the use of optimized models or compression methods, the listing can be extended much further due to a large number of studies. The mentioned works are intended to illustrate the potential of the method by presenting a rough range of improvement. In some cases, it is still possible to fall below or exceed this range. The improvement was listed in the metric used in the papers. Creating a consistent score on, for example, the percentage of avoidable emissions would be a use that would simplify the selection of measures. Especially since runtime and energy consumption are not necessarily correlated.

The applicability of the methods is often a question of additional effort. In particular, if the measures are not yet integrated into commonly used libraries, the company must already be strongly committed to sustainability in order for the effort to be made. Research in particular should therefore not just stop at developing methods, but should also integrate them into open-source libraries.
Actions related to a public cloud or the choice of programming language must be communicated to the corporate world at an early stage with their effectiveness. Especially now, when AI is still being established in companies, it is easier to make it sustainable than when legacy systems and other burdens of the past make restructuring difficult.
Some measures are only applicable with a tradeoff of accuracy. An internal discussion on the tradeoff is needed here. Fortunately, performance loss is adjustable for compression techniques.
Many measures also produce a speed advantage, which can be an additional incentive for companies.

The method of literature research could lead here to a large extent of the measures. Many of the publications on collections of measures are from the current year, which indicates an increasing importance of the topic of efficiency and sustainability. With regard to the literature search, there is of course a risk that works that have just been published are not yet recorded in the databases. Since these are not referenced by other papers directly after publication, the direction from the basic works such as [Sch+19] to referencing papers is especially crucial.
Especially since the focus was placed on the use in companies, the inclusion of blog contributions from organizations dedicated to the topic can be seen as a useful addition.

5 RELEVANCE OF SUSTAINABILITY IN COMPANIES

In this chapter the results of the expert interview can be found. The justification and explanation of the research method and the structure of the interview can be found in chapter 3.2.

In order to answer the research questions, a corresponding questionnaire was designed. Among other things, the questions of Capgemini's Green IT Study of 2021 were incorporated here, which is presented in 2.2. The questionnaire was not changed in the course of the interviews due to the satisfactory information density, only the finer questions were refined. Specific questions about the measures were often omitted because companies already said in advance that the topic did not play a role for them. The whole interview guide can be found in the appendix A.1.

Selection of experts

In total over 100 persons were contacted this way. As a result 12 experts were aquired via Linkedin. 10 of the experts are active as leaders in the field of data analytics/data science/machine learning in their companies. Two respondents have several years of experience as developers of machine learning solutions.

The companies' industries are widely spread: metal industry, transportation/logistics, mobility platform provider, real estate sector, chemical industry, cloud provider, print media, technical IT consulting, banking, data-based biotechnology, consulting and software house for banks, automation and digitalization. Consulting companies are often drivers of new topics in companies, which is why including them also leads to a good representation of the real status.

The maturity level of AI in the company was assessed by the creator of the work using the maturity model presented in table 2.1. For this purpose, statements were taken from the interviews and matched to the different levels. In the interview itself, the introductory questions about the importance of Data Science and the personal scope of duties were intended for this purpose. The classification was not always clear. In some cases, just started initiatives were mentioned, which in themselves pave the way to a higher level. In case of doubt, the lower level was always taken in order to reflect the actual state as best as possible. In the digital appendix, the classification can be found on the basis of the supporting documents.
The number of companies that reach a certain maturity level is shown in the

graphic 5.1. Level 2 is predominantly reached, which still represents a low prevalence of AI and few productive applications.

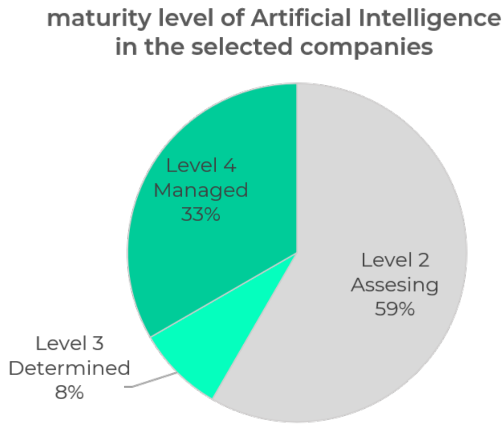

Figure 5.1: amount of interviewed companies in a specific maturity level

Remarks on the process

The interviews lasted on average about 32min and were conducted online without exception. The recording of the interview (B9) was interrupted at the end for technical reasons, but only the closing was lost. The interview was therefore added as completed. Another interview (B12) had to be shortened to the extent of 15min due to the time constraints of the interviewee. However, since the gain of knowledge through active measures regarding sustainability was of importance here, the interview is included. Due to the speed of the process, it was not always possible to ask follow-up questions to the extent of the other interviews. Therefore, the absence of categories is not necessarily significant here. Reference is made to this accordingly in the evaluation.

5.1 RESULTS OF THE GAINED DATA

For each research question, the expert interviews inductively derived categories that provide answers to the research question. The results are presented according to the following scheme:
The *categories* formed for the qualitative content analysis are first presented and backed up with examples from the texts. The *number of occurrences in the documents* is then evaluated. It should be mentioned once again that due to the small scope and the qualitative survey, conclusions can only be drawn in the context of the interviewees. The percentages refer only to the proportion of respondents.

5.1.1 *Reasons for sustainability in general*

To better discuss the drivers of corporate sustainability as it relates to Data Science, companies were asked what generally lead to increased interest in it. Participants were encouraged to list all the factors they were aware of. Overall, the following drivers emerge which are further broken down in figure 5.2. Example citations for each category are given in table 5.5

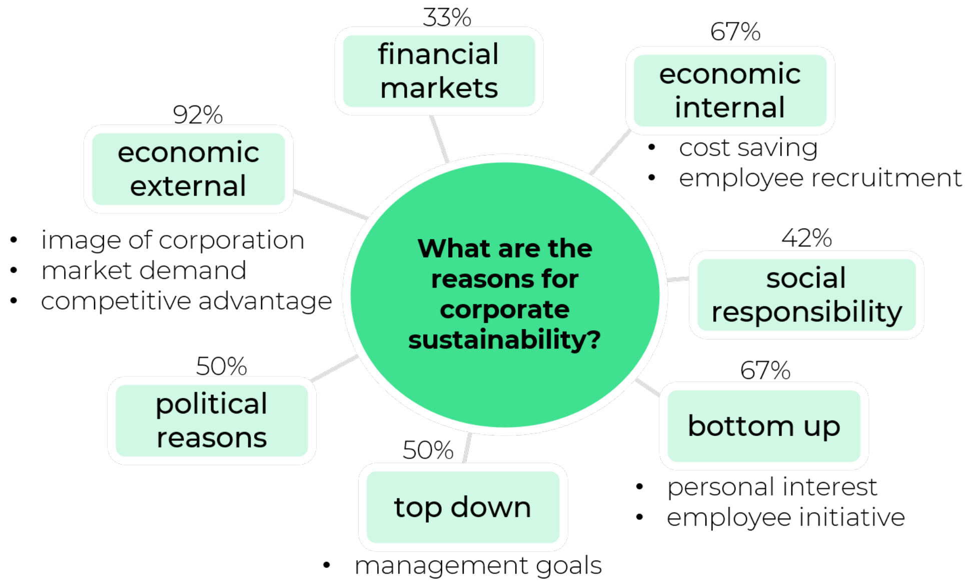

Figure 5.2: categorized reasons for companies to look deeper in sustainability

Bottom Up

Overall, the reason is cited by two-thirds of respondents. Half of the interviewees report about initiatives which originate from the personal interest in sustainability of the employees. Three of the interviewees are themselves well versed in the subject and contribute this as leaders. It should also be noted that there are opposing opinions: *'sustainability pops up here and there. But there aren't any major initiatives that I'm aware of.'* (B7, translated)

Top down

The management level is also doing a lot to promote sustainability. Here, too, out of personal conviction or through strategic target agreements. This was the case for half of the interviewees.

Political reasons

In half of the cases, political regulations in the mandatory sense and funding in the motivational sense were identified as drivers for sustainability. Literally, CO_2 taxes were named by B5 here. A shortage of resources due to Russia's current offensive war against Ukraine is also a factor for companies.

Financial markets

The pressure from financial markets to act more sustainably was mentioned by two company representative. In particular, the refinancing of debt incurred as a result of the Covid19 pandemic makes the issue economically interesting, especially for affected businesses. Another two respondents who work in the banking industry confirmed the development in corporate financing to consider sustainability plans for better rates.

Social Responsibility

42% believes that the social responsibility to do something about climate change as an existential threat to humanity also applies to companies.

Economic external

The economic trend of sustainability is beyond question for almost all experts.[1]. The corporate image benefits from measures and competitive advantages are created by products. Especially in the B2C market, the customer demand for sustainability is a driver that would otherwise exclude the company from the market.

Economic internal

Two-thirds of the respondents also see the economic benefits of sustainability internally within the company. On the one hand, it makes it easier to recruit and retain employees, and on the other hand, it saves resources and thus reduces costs.

Table 5.1: Example lines for the categorization of reasons

Category	Example	Cited
bottom up	So, in general, I think this is close to the heart of many employees in our company.	B4, translated
top down	We have made it one of our strategic goals to work sustainably ourselves and to enable our customers to produce more sustainably	B12, translated
political reasons	CO_2 taxes	B5
financial markets	But what I see is that especially the institutional investors are very keen on this. So for this in the meantime a must.	B4, translated
social responsibility	And yes I think especially in IT consulting, if you are responsible for that, in the end, that companies are now using more and more IT, which we then contribute to the growth of consumption in the end.	B8, translated
economic external	If you don't react, then you'll probably be left behind. I mean, it's just like adopting a new technology, right.	B9
economic internal	I think that's always nice for the companies themselves, of course. They can then save costs in the course of this	B6, translated

1 The non-mention is in an interview which had to be shortened due to scheduling reasons

5.1.2 *Relevance of the topic*

One of the basic objectives of the expert interview is to determine the current status of sustainability in the company. For this purpose, the experts were asked for their assessment with a separate question. In particular, they were asked whether they capture resource consumption in any way. The categories formed can be seen in figure 5.3.

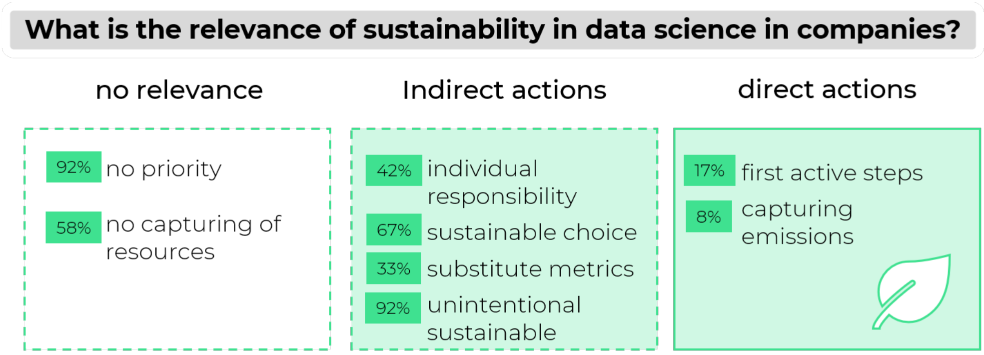

Figure 5.3: categorized relevance of sustainability of AI in companies

Since the state of research on the sustainability of AI is only just emerging and AI as a technology has not yet become established in companies, it was to be expected that the topic would be very new for the companies. Therefore, instead of differentiating into individual maturity levels, the division into no relevance, indirect measures, direct measures is made. Likewise, no fully established, automated solutions were expected, but ideas and approaches that are already present in companies.

No relevance

The survey clearly shows that the relevance of green AI for companies is not yet established. Indirect and direct measures have already arrived in companies in isolated cases, but the companies surveyed pointed out that there is generally no priority (92%) and that the recording of which resources were used in the course of AI does not take place (58%). Only in IT consulting has the topic arrived and own solutions are being developed for customers.
Reasons were often given that have a higher priority in the development of an AI:

- get the use case to work
- fast way to minimum viable product
- only performance as main focus
- high monetary return on the use case

Especially the companies with their own production or companies with few productive models indicate that even from the opposite direction AI is not yet relevant for sustainability. There are other "low hanging fruits" in the company to improve its ecological sustainability.

Indirect actions

All companies say that they at least contributed indirectly to sustainability. The implementation of sustainable use cases also plays a role here. A resource-saving design was recognized in the basic approach of starting with simpler models, which is a common practice for structured data. Also the use of a public cloud was recognized for the better utilization of servers. Optimizing steps of models were done for other reasons like scalability or runtime optimization due to requirements of the use case.
Five experts stated that all measures, however, are in the hands of the individual model developer and are not controlled centrally.
The selection of AI solutions (in the case of buy instead of make) and in particular the selection of the cloud provider took place, according to two thirds, at least with the inclusion of sustainability criteria. Whereby these acted more as nice-to-have factors instead of changing a purchase decision.
After all, one third of the experts state that they include the cost factor or the calculation time. However, this is more for estimation purposes and is evaluated only sporadically. In the banking environment, this information is recorded in the course of model validation, but not stored centrally, which makes it difficult to evaluate or track in terms of sustainability. Interestingly, the idea of recording costs is classified as monitoring sustainability multiple times, since these are closely related and can be easily recorded in the cloud.

Direct actions

The first active steps were also mentioned in two interviews. The development of an own product for sustainability in AI in companies is the approach of IT consulting. One company developed a nano production app where users can create their own time series and see the footprint of the chosen model. In this case, an initial recording of emissions in the company is also integrated, with day-to-day business evaluating the observation via a carbon tracking tool from the cloud provider.

Table 5.2: Example lines for the categorization of relevance

Category	Example	Cited
no priority in daily business	So if the if a longer computation produces a better result, I would imagine that most companies would prefer the better result rather than the shorter computation time.	B7, translated
	We have development guidelines, but they have not yet been optimized for sustainability.	B12, translated
no capturing of resource consumption	To be more sustainable, we would not set up tracking. Not yet, then.	B10, translated
unintentional sustainable	but we always go for the easiest and simplest solution that can solve the problem. So we don't start with like okay, we have a huge model that can solve any problem. We always start simple and then put on top of it.	B9
	It is in the actual interest that the machines do not calculate forever, because you simply want to finish faster. But not in the sense of okay, you have to become more sustainable.	B11, translated
sustainable choice of products and vendors	Which of their locations have the best clean energy and we looked at that again and also chose based on that.	B10, translated
individual responsibility of the developer	This is always with the people who build the model, of course	B1, translated
capturing substitute metric	but only from the time perspective, not from energy perspective. So from time perspective, we have already some kind of a view of what it's already taking or consuming time.	B5
first active steps	We do this economically, so that we can see how we can somehow also make this a product, in the broadest sense, so that we can say that sustainable consulting is also a product.	B8, translated
capturing emissions	And specifically with one of the big cloud providers, there's now dashboard for users of how much of their last month of their last quarter CO2 was emitted, their usage.	B12, translated

5.1.3 *Hurdles for more sustainability in AI*

The difficulties and challenges in the sustainable design of AI emerged indirectly in the course of the discussion. They are the problem areas for the further steps that were asked of the experts. The lack of priority given to sustainability in the development of AI systems has already been considered in terms of relevance and is therefore left aside here. The resulting further hurdles are shown in the figure 5.4.

Lack of knowledge

The biggest hurdle with a mention in three quarters of the interviews is the lack of knowledge in the field of sustainable AI. Knowledge about measures and about the sustainability of models is either not available in the enterprise or not available in research either. It is not visible what the footprint of AI

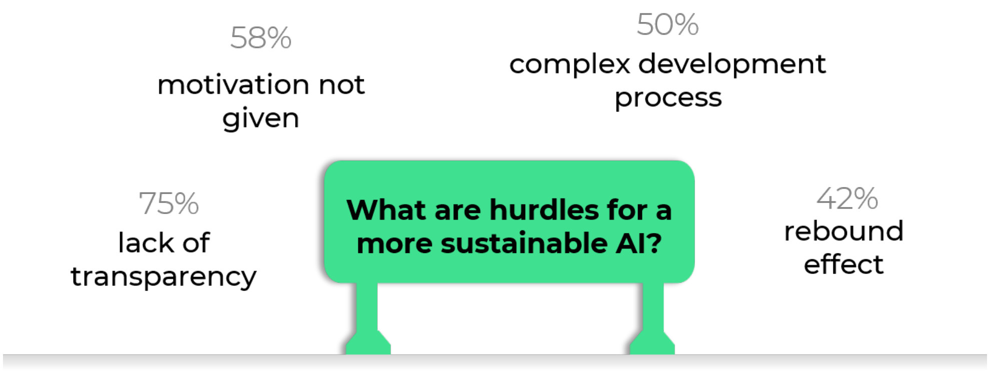

Figure 5.4: categorized hurdles of sustainable AI

is. Even if measures are known they cannot be quantified. This also leads to companies not knowing where to start and identify the best use case. In particular, the lack of quantification makes communication with management and external communication difficult.

Motivation not yet given

Often, the incentive to take care of sustainability is not given. There is no doubt that attempts to make artificial intelligence more sustainable in companies still require time and effort. 58% of the experts surveyed do not yet see the topic as worthwhile. The predominant reason given is that it is not economically beneficial (especally in B2B). Even if the reduction in energy consumption results in lower costs, these are not in proportion with the developer resource required. The relevance of AI for the overall sustainability of the company is seen as too low to get into doing even if the sustainability in the company is to be improved.
Another factor here is that, as already mentioned, the companies' priorities lie elsewhere and the developers are working to full capacity.
Following on from the study on drivers of sustainability, the lack of personal interest and commitment in the company should also be mentioned here. Interviewee B12 often spoke of the fact that they were able to establish so many measures because the management and team themselves are enthusiastic about the topic. The latent factor of the right person in the role is therefore added here, even if it could not be formally derived from the interviewees' point of view.

Complex development process

One hurdle is itself in the nature of data science applications accoridng to half of the participants. They are associated with trial and error. There is no one model that works equally well on all problems. In Deep Learning, there is a clear tendency for larger models to perform better with more data. One third of the experts surveyed found the unpredictability and data hunger of models to be disadvantageous for development in the interests of sustainability. As long as no standards or best practices have been established, the topic will be slowed down, according to 41% of the respondents. Developers are currently having a hard time, as they have to deal with the topic of sustainability in addition to their own tasks.

Rebound effect

A gain in efficiency does not necessarily lead to lower resource consumption, but often generates an increase in demand for the resources. The effect is known as the rebound effect and, according to five experts, jeopardizes true sustainability once measures are applied. Many of the runtime optimizations that resulted in lower energy consumption enable the model to be used at a finer granularity, e.g., one country is scaled to 30 countries by the efficiency gain.

Table 5.3: Example lines for the categorization of hurdles

Category	**Example**	**Cited**
lack of knowledge	I believe that if we want to get there, we have to make the issue of CO_2 emissions much more transparent overall. So I think the first step will be to play the whole thing on the purely monetary factor.	B3, translated
motivation not given	But we sort of sell it to companies as. There is no goal or incentive to do anything.	B10, translated
	And maybe in IT you're still (...) flying under the radar a bit. That's my current perception.	B2, translated
nature of data science	How long did the training take? But don't worry about it beforehand. It is also difficult to estimate such things beforehand.	B1, translated
missing procedures	But this sustainability field, I think is still a bit young. It also needs to be established.	B9
	And we're not yet at the point where we're like, let's say, classical programming, where there are standardized design patterns that you use to solve certain tasks.	B3, translated
rebound effect	Yes, yes, in any case, of course, there is a small rebound effect in what you save. In the end, you use that to do it more often.	B7, translated

5.1.4 *Necessary steps for more sustainability*

The experts were asked in a separate question what steps need to be taken to promote the sustainability of AI in companies. The results can be seen in Figure 5.5. The ideas for solutions emerged in a spontaneous assessment.

Knowledge generation

Because the topic is still new in research and almost not present in companies, the topic of knowledge generation is important for 11 out of 12 respondents. Only one participant (B11) sees the problem as too distant and sees the transparency of emissions as the only necessary step. Particular hope is

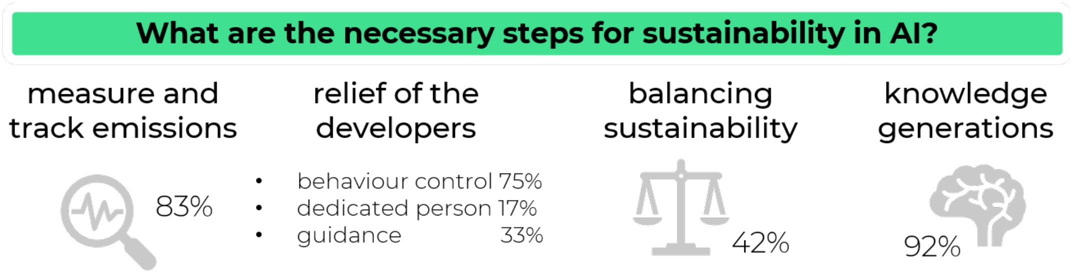

Figure 5.5: next steps for more sustainability in AI

placed in the assessment of which model is suitable for a given problem. Benchmarks are also often mentioned, which are either shared publicly or created within the company. At this point, cooperation with research institutions is also not excluded. Especially for platforms that offer models, the memory and resource footprint for each of them should be specified.
Overall, sustainability of AI will still lead to unique selling points in isolated cases, but a publication of knowledge for the common is a long-term step. Cross-company publications can help learn from failures and eliminate inefficiencies. The open-source area is also here of importance.
Moving the topic into education was also seen as a necessary step.

Measure and track emissions

For 10 out of 12, measuring sustainability is a major first step. On the one hand, it is seen as a necessary step to derive measures (in this sense, knowledge generation) but also as a consistent awareness tool. Measurement serves as a continuous reminder that systems carry a footprint. The transferability of the CO_2 equivalent to other areas encourages individual responsibility and simplifies communication with management. It is also seen here that the net sustainability of a use case can be calculated. That is, the emissions that are emitted before emissions can be saved.

Relief of the developers

For sustainability to be successfully established in the field of data science, the developers must be relieved of responsibility. The study of relevance showed that the developer is often individually responsible for the measures. 4 of the 12 experts therefore think it is necessary to provide better guidance for developers. This could be through the provision of best practices, development guidelines or support via training courses. It should be specified in the consultation that sustainability is also to become a fixed component of the consultation.
Only 2 of the 12 companies surveyed see a need for a dedicated person to discover sustainability potentials (also) in the field of AI.
Three-quarters suggest an incentive and 'punishment' system. It is currently too easy to use very resource-intensive models. High resource consumption must be penalized in some form or certain contingencies introduced. The

good, sustainable behavior should be simple, convenient and in the best case automated. More concretely, Kubernetes flags for sustainable execution or badges for an energy level ranking for models would help here. Automation via pre-checks before committing the code or automatic optimization are also imaginable. The step into the cloud is also seen as a facilitation that offers ease and comfort.

Balancing sustainability

Sustainability, cost and performance have to be weighed against each other according to 42% of the respondents. Although an attribution of sustainability in terms of cost is constructed to a certain degree, the trade-off and communication is simplified. The question must be answered, because otherwise it will be asked again and again.

Table 5.4: Example lines for the categorization of next steps

Category	Example	Cited
knowledge generation	We will have to research this much more deeply, this entire area, in order to have this expertise that will then also allow me to say, okay, I now want to design my implementation for sustainability. I also have the relevant background knowledge to make this possible at all. I think we need more research in this area.	B3, translated
	Transparency would definitely help on the cloud level, like giving some kind of ranking of the algorithm and implementation with their corresponding memory footprint, resource footprint, CO2 footprint, energy consumption.	B5
	So I think that education also plays a very important role in this area and that the topic of sustainability and such aspects must also be taught.	B7, translated
measure and track emissions	as continious reminder	B6, translated
	And I think that's probably the starting point a conscious decision to say "we want to measure how sustainable we are and how our actions impact that."	B4, translated
guidance of employees	very concrete best practices would help us a lot	B6, translated
dedicated person	I think that's the first significant responsibility, that we really have a central area that's going to bundle that together.	B2, translated
Behaviour control	Git has a bunch, so when you commit a code in the repository, there's a bunch of pre-checks that go through. Which then recognize redundancies quasi in the end effect simply	B8, translated
balancing sustainability	And where you can use Data Science, of course, to develop an attribution model.	B4, translated

5.1.5 *Forecast development*

With a dedicated question, the experts were asked for their opinion on how the topic of sustainability of AI and with AI will develop in the future. Possible scenarios were freely formed by the experts. An overview of the views is displayed in figure 5.6. The categorization is formed with a positive and a negative sentiment about the two topics. The named aspects are taken up in the following explanation.

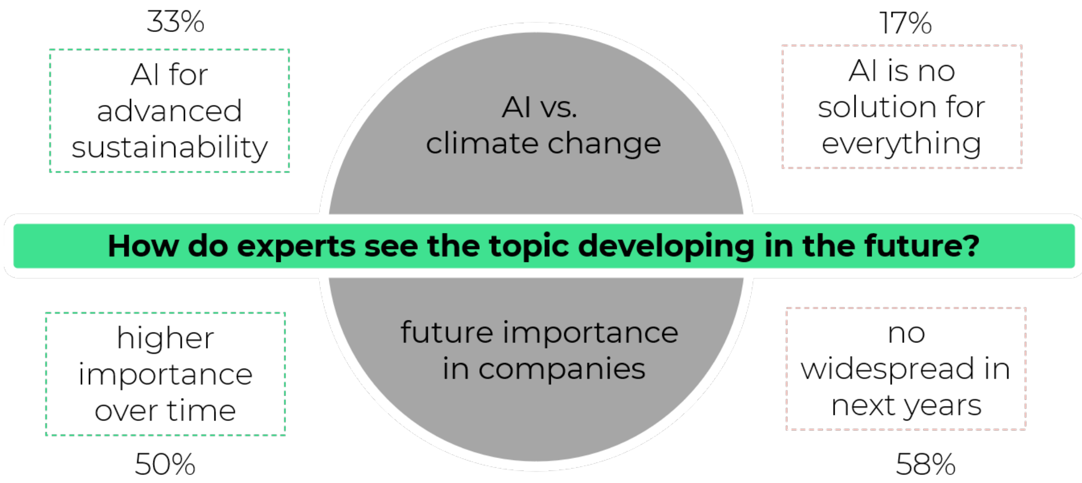

Figure 5.6: Experts' opinions on the development of environmental sustainability by and with AI

Artificial intelligence vs. climate change

Opinions differ on whether AI can be one of the big solutions to global climate change.

AI FOR ADVANCED SUSTAINABILITY In the conversation with the experts, AI is often identified as a great tool against climate change. Especially the complexity of the use cases, the amount of data and real-time requirements of the problems makes AI a necessity. This opinion is shared by a third of the participants.

AI IS NO SOLUTION FOR EVERYTHING The other opinion held by two experts is that the use of AI on all problems is due to the current hype. The importance on the climate change related is equal to other tools.

Future importance in companies

Whether sustainability is becoming more important in relation to artificial intelligence is also disputed in the survey. In two cases, the experts were undecided. The spread of AI is bringing the topic to the forefront, but from the current status in the company, its significance is not foreseeable.

SUSTAINABILITY OF AI GAINS IMPORTANCE Half of the respondents mention that the sustainability of AI will play a role. On the one hand, AI is gaining in importance as an approach to economic sustainability. On the other hand, the scope of AI systems in companies will reach a critical mass, according to the experts, which makes the topic inevitable.

NO WIDESPREAD IN NEXT YEARS 58% of the respondents do not yet see the sustainability of AI emerging as an issue in companies in the next few years. Either, in their opinion, digitalization as a basic requirement will not be sufficiently advanced or they do not see the demand in the B2B area clearly. Potentially, the sustainability of AI is described as a permanent niche topic.

Table 5.5: Example lines for the categorization of the topic development

Category	Example	Cited
AI for advanced sustainability	And this and that - how do I manage all this optimally in real time? You have to do all that with an AI. No human being will do that. That's where I see the huge benefit of AI. Big scale, then.	B1,translated
	(...) we can't manage that in our heads. We have to use computers to do it. And computers are not fast enough to provide analytical solutions. In this respect, we always have to make approximate systems like AI	B8, translated
AI is no solution for everything	I would like to see more natural intelligence than artificial intelligence, to be honest. (...) For me, it's more a change of consciousness of natural intelligence than artificial intelligence. AI is only somehow an aid in implementing that.	B11, translated
higher importance over time	Now, I also assume that [AI] will become much, much more widespread in our company as a result of this [data] literacy program. And then you also reach the point where this critical mass slowly emerges (...) where costs are subsumed. (...) And then this topic will also come up	B2, translated
no widespread in next years	They only do it when the customers want them to. (...) But we sell it to companies. There's no goal there or no incentive to do anything.	B10, translated
	I don't know if any other branches, but in B2B, we are talking still about digitalization and we are still on the baby steps of the digitalization.	B5

5.2 DISCUSSION

The interviews lead to an answer of the research questions. The scope of the interviews was sufficient for a first insight. It would be interesting to look at very large technology companies such as Google, Amazon or Microsoft. However, it was not possible to obtain a contact person for the interviews there. When considering the results, it is important to remember that the individuals volunteered for the interview. A certain bias of interest is therefore given. Under certain circumstances, the image conveyed of the sustainability of AI is rather too optimistic. In the same way, not all AI experts are represented on the social network Linkedin. The subjectivity of Mayring's evaluation is also a point of criticism here. Intersubjectivity can be increased considerably if different people take over the evaluation and come up with the same categories.

A further interview can provide new insights into the topic.For a more in-depth questioning of the persons, prompting questions could be asked on the one hand, such as "In your opinion, what could be opimized at the time of execution in the direction of ecological sustainability?". Here, however, there is a risk that participants will feel overwhelmed and that the atmosphere of the interview will lead to less open conversations. In [AENEN20] the awareness of students on the topic of sustainability was checked in which concrete measures were ranked, according to their contribution to sustainability. A similar design of a survey could help in the next step to evaluate the level of knowledge in depth.

The reasons for increased interest in sustainability are plausible. For the Data Science area, especially the categories *bottom up*, *top down*, *economic internal* and the *social responsibility* become relevant. Due to the small share of AI, which is often highlighted by the experts, financial markets and external economic reasons are rather secondary. Likewise, there are no known political regulations in the field. Personal interest is of great importance. The interview showed that if the leader has a personal interest in the topic of sustainability, measures in this direction are also more likely to be established and the topic receives more attention[2]. Both observations each led for the scope of the interviews conducted. Especially B12 stated that the progress is due to her and her team's motivation. This is also crucial in ongoing studies such as [CS16]. Here, the first step towards corporate sustainability is stated to be that employees (and management) should have developed a common ecological vision and everyone should be actively involved in it.
Overall, an interesting aspect to promote sustainability of AI can be that the cost advantage is highlighted and that certifications etc. are available as a marker of sustainable AI for companies for self-promotion purposes.

2 Here the causality directions needs to be discussed. Maybe the leader was was hired because he or she fit well into the company's vision with sustainability.

The relevance of sustainability in companies was not surprising for the fact that the topic has only been emerging in research since around 2019 and companies are only just beginning to really establish AI. This is also shown by the determination of the maturity level of the companies. The majority are still at the assessing level. There, the priorities are first of all to get AI applications up and running and to establish standards.
In order to bring sustainability in companies further forward, measures can be subdivided according to maturity level in order to address companies with a lower maturity level in particular. This can be linked to the solution of typical problems of the stage.
All companies were able to name indirect measures. The fact that the experts classified many measures as promoting sustainability that were actually implemented for other reasons shows at least that there is not a problem in understanding environmental sustainability. In this sense, there is no need to create extra understanding, as is the case with other sustainability topics such as recycling.
However, the measurement of costs and training time should be replaced by better metrics, precisely because the necessary steps see more transparency as necessary. For this, it can help to convince companies of the benefits of a more accurate evaluation and to show them simple solutions like the carbon tracker tools mentioned above. Especially the introduced dashboard for checking the carbon footprint of cloud services from Amazon Web Services can easily enable a replacement of monitoring over cloud costs - at least for this provider.
Even if the first active steps are beginnings, the communication of these steps can also help here, so that other companies find an inspiration.

The hurdles mentioned are mirrored by the statements about the relevance of sustainability. Transparency is a particularly important issue here. In a further survey, the hurdles can be broken down even further. Subcategories from the hurdles formed can be used for this.
The complex development process in particular will be simplified in the next few years, as AI should be broadly applicable in the company. A special significance is therefore seen in the sustainable design of AutoML solutions. It was surprising here that, despite the complex development process, only two people in the solution approaches saw a dedicated person as support on the topic of sustainability as a necessary step that would relieve the burden on the developer. A lack of priority of the topic may be a possible justification here.
The rebound effect should be taken seriously as it threatens efforts to increase sustainability. Especially since many identified measures lead to a shortening of the runtime. Organizational measures such as limiting calculation capacities or setting a maximum growth are effective here. However, caution against this must also be communicated at the same time as the topic of sustainability.

The next steps for more sustainability address the above hurdles well. Transparency is improved by measuring and tracking emissions and exploring further measures. Here, benchmarks from research and in companies can help. A contribution is also made by this thesis, that important, applicable measures were pointed out. This transfer from research to practice must take place more regularly.
Also, the developers must be relieved. According to the experts, developers are otherwise far too often individually responsible for sustainable measures. An investigation into management support in terms of time and resources could also be interesting here. Overall, it must be convenient to make sustainable decisions. This can be achieved through standards and governance structures in companies. In a cloud, for example, only sustainable locations can be selectable, or the execution of certain models can be allowed at only a certain time window. The said pre-checks also prove that ideas already exist in the companies, but the execution still has to take place. Especially these ideas confirm the method choice of a qualitative interview.
A special attribution of sustainability in cost form can be interesting for the whole company. However, since a whole capture of sustainability via metrics has to be established first, simpler approaches should be found here. In the opinion of the author of this thesis, transparency is also needed here. In the other direction, sustainability costs can be made more transparent. 'How much effort does a performance increase of 1% need?" for example.
In a blog post by Micorosoft on the topic of sustainable AI, similar necessary steps were discovered.[3] Metrics, standardization and automation are also seen as necessary by the experts.

3 https://devblogs.microsoft.com/sustainable-software/the-current-state-of-affairs-and-a-roadmap-for-effective-carbon-accounting-tooling-in-ai/

6 RANKING OF CLASSIFICATION ALGORITHMS

In this chapter are the results of the described benchmark of classification algorithms. This is intended to create awareness among developers about the energy consumption of the chosen models. The code of the execution is available at `https://github.com/LuWiesa/ranking_classification_models`.

6.1 RESULTS OF THE BENCHMARK

The benchmark of the nine models with six data sets was created within 44 min 14 s. The Pearson correlation between runtime and energy consumption is 0.99 in training and prediction.

6.1.1 *Training of the models*

When training the models, the data sets were prepared according to the preparation described in chapter 3 and the models were fitted on the data. Here, convergence warnings occurred 5 out of 6 times for the multilayer perceptron (neural network) and twice for the logistic regression. In practice, the data would have to be adjusted here or the neural network would have to be designed to fit the complexity of the problem accordingly.
The results are shown in figure 6.1.

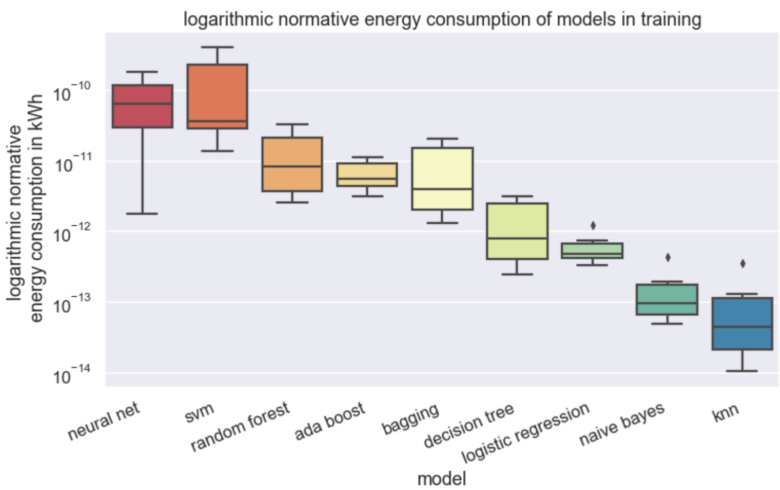

Figure 6.1: logarithmic normative energy consumption of models in training

The logarithmic scale of the energy consumption standardized on the data set should be noted. Between the lowest value of the K-Nearest-Neighbors method and the Support Vector Machine lies a factor greater than 10 000.

If the median values[1] are compared, a factor of 1390 is obtained. An interpretation and explanation for this can be found in the following discussion. Furthermore, it is visible that the values of the energy consumption fluctuate widely despite normalization to the data set. The neural network leads the energy consumption in the training. After the support vector machines come the ensemble methods, which use several models. The results are shown in figure 6.2.

6.1.2 *Prediction of the models*

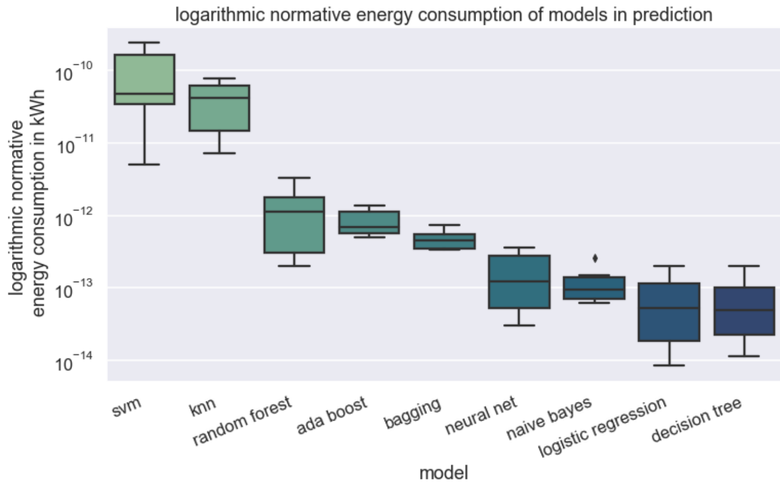

Figure 6.2: logarithmic normative energy consumption of models in prediction

In the prediction, the normalized energy consumption of the Support Vector Machines and the K-Nearest-Neighbors method is larger by a factor of 42.6 and 37.1 from their median values than the median of the following Random Forest method. There are some differences in the ensemble methods. The lowest values are reached by the Logistic Regression and the Decision Tree. These lie almost on the same level. The neural network lies with a factor of 2.4 higher energy consumption in the median over the decision tree.

6.2 DISCUSSION

In the benchmark, an impression of the energy consumption of the models was determined with a total execution time of less than 45 minutes. Only a few lines of extra code were necessary for this, which is why the recording of energy consumption is an easy measure for developers to integrate and benchmarks should also be introduced in companies. The execution on the own laptop was well enabled by the option to track the energy consumption of the associated process.

1 The median is used here so that the interpretation can be understood on the basis of the available charts. Graphs with the mean values can be found accordingly in the appendix B

Significant differences between the algorithms are apparent, and possible explanations are provided below. First, it should be noted that the sample size for a boxplot chart is very tight, but the plot still shows the values better than simply giving a range of results, which is why it was chosen anyway.

TRAINING The maximum number of iterations for the neural networks is 200. Thus, in total, in 5 out of 6 cases, all data were pushed through the network 200 times and the errors were propagated back again. The high value in the training is thus understandable. Support Vector Machines react strongly to a large number of features. Since the datasets use a total of 23 to 386 features as a result of the one hot encoding, the variability here is not surprising. For the ensemble methods, the number of base estimators in the vanilla implemantation is also different. Random Forest, Adaboost, and the bagging have 100, 50, and 10, respectively. The order reflects this well. Since only truncated trees or decision stumps are used here, the factor is not directly comparable to the energy consumption of the unconstrained decision trees. It is also not surprising that the KNN method is the fastest. The data values are only stored in the model, since later in the forecast the distance of a new data point must be calculated with all known points.

PREDICTION In contrast to the training, the KNN method now involves a very high energy consumption. The ensemble methods retain their order. This seems to be influenced by the number of estimators. The decision trees and the logistic regression are once formed very resource efficient. For companies, the effect of good explainability can be of additional interest. Especially in companies the prediction is involved to 90% of the total consumed energy of a model. Especially the ranking of the prediction of the models should therefore be kept in mind during development work.

FURTHER RESEARCH The investigation can be deepened in many ways. Some ideas are listed below

- execution on different hardware with parallelization options
- testing of different implementations and libraries
- include more model types
- repeat benchmark for regression methods
- try out different hyperparameter configuration an their influence on the energy consumption
- include more datasets and develop best practices for model-compatible data preparation

In particular, explaining the variability of models on different datasets is a worthy goal toward predicting the energy consumption of models on particular datasets.

7 DISCUSSION OF THEORETICAL AND PRACTICAL RESULTS

In this work, the research question *What is the awareness and status of sustainability of artificial intelligence in companies and how can it be improved?* was solved in three parts, which are merged here.

Through the 12 interviews with companies, important insight was gained on the current state of sustainability of AI. Similar to the 2021 study on Green IT (chapter 2.2), the topic is not yet established. Again, 92% of the companies explicit did not see the topic as a priority and only two out of 12 companies are implementing first steps. While in the interviews conducted, only one company is actually capturing emissions in a use case, the Green IT study would suggest that more, 29%, are. Thereby the tool landscape for the first steps is given and even first dashboards and API are available for a productive operation. In the practical benchmark, the applicability could be experienced first hand. The biggest hurdle here is seen in the need to spread the tools further. The dashboard for monitoring the carbon footprint of Amazon Web Services' cloud services also needs to be made more widely available.

Especially the fact that the developers are mostly responsible for the measures themselves shows that more needs to be invested in training and spreading the best practices. During the introduction of the work, initiatives were already mentioned that deal with the topic and also the measures from the research already show that the niche topic of a few years is slowly gaining momentum. This is also visible if you look at the year of the studies of the found measures, since a large part was published in 2021 or 2022. Especially measure collection in specific areas like AutoML, Neural Machine Translation and NLP exist as support for the developers.

Other issues such as balancing sustainability and monetary factors or model performance are not resolved. Here discussions probably have to be solved individually for the companies. However, the concern that sustainability always entails a lot of monetary effort and a loss of performance is unfounded. Measures such as power capping the GPU or using random search are easily applied and have no effect on performance according to the studies found. Random Search can even provide faster training than Grid Search.

The companies often mentioned that they use particularly simple models either for reasons of comprehensibility of the AI systems or because they are still in the early stages of AI. In this case, the benchmark conducted offers a good opportunity to enrich the selection of models with the dimension of energy consumption.

It would have been interesting if the expert interviews could have been more closely intertwined with the measures found. As in [AENEN20], the interviewee could have been asked to classify the measures according to their effect. However, this was not yet possible because the basic work of providing an overview of the status of Green AI was not given.

8 CONCLUSION AND FUTURE WORK

Overall, it was possible to answer the central research question *What is the awareness and status of sustainability of artificial intelligence in companies and how can it be improved?* based on the current state of knowledge of the topic. The current status has been well determined, at least for the selected, interviewed companies. Hurdles and next steps were further determined in the interview itself. The listed measures and the ranking of the classification algorithms are options for companies to further improve this sustainability. Thus, all parts were answered.

As the topic of Green AI is just starting to enter companies, there is a lot of room for further in-depth investigation. In addition to quantitative evaluation of the status of sustainability in AI, which could take the qualitative knowledge as a starting point, other company-focused research ideas are conceivable. For example, analogous to the maturity model of AI, a maturity model for sustainable AI can also be developed.

Standardizing and communicating the automatic implementation of measures and carbon accounting of the AI department can be tackled as a joint project between research and business. Overall, many of the efforts should be shared as they contribute to the larger society-wide goal of tackling climate change. Many beneficial methods need to be quickly implemented in common libraries so that they can be used worldwide. The open source idea, so buried in IT, has special potential to achieve the common goal here as well.

In the discussions and the studies on the measures there are many ideas like a carbon intensity forecast for a renewable aware scheduling, or the design of pre-checks for sustainable code in the AI domain. In the discussion section of the results, other ideas of the author of this paper are also listed. In this sense, it is important to become active in order to achieve further progress in the field of Green AI in companies. In the author's opinion, the work for this is seen more in awareness-raising measures. The footprint of AI must become visible. Consulting companies have to carry the topic into companies. The topic must be integrated into the education of future AI experts. Employees have great leverage when it comes to the topic of sustainability, as evidenced by the experts' statements. Personal awareness and interest are the biggest drivers, at least in the period before standardization.

Part II

APPENDIX

A EXPERT INTERVIEW MATERIAL

A.1 INTERVIEW GUIDE

As for most of the interviews are held in german, the questionaire is in german with the translated questions for the english interviews.

Einstieg (5min)

1.1 Welche Bedeutung hat Data Science für ihr Unternehmen? *Which relevance has Data Science for your corporation?*

1.2 Inwiefern tragen Sie Verantwortung für die Gestaltung von internen Data Science Anwendungen? *How are you responsible for the design and general conditions of internal data science applications?*

Teilbereich A – Relevanz Nachhaltigkeit (20min)

2.1 Welche internen und externen Faktoren sehen Sie aktuell, welche zu einer gesteigerten Bedeutung von Nachhaltigkeit in ihrem Unternehmen beitragen? *Which internal and external factors do you currently see, which contribute to increased interest in sustainability in the company?*

2.2 Was ist die Relevanz von Nachhaltigkeit im Bereich Data Science in ihrem Unternehmen? *What is the relevance of sustainability in the field of data science in your company?*

2.3 Werden die ökologischen Auswirkungen der KI-Modelle vor Projektstart in ihrem Unternehmen abgeschätzt? *Is sustainability directly or indirectly represented as a decision criterion with regard to the selection and generation of new use cases?*

Teilbereich B – Implementierte Instrumente (12min)

3.1 Welche Verantwortungen für eine nachhaltigere Gestaltung von KI-Modellen und deren Anwendung liegen im Unternehmen und welche bei den Cloudbetreibern)? *Which responsibilities do they see in the company and which in platform and cloud providers for sustainable AI?*

3.2 Wie wird der Ressourcenaufwand hinsichtlich Datenhaltung und Berechnungsaufwand der Algorithmen in der Entwicklung und im Betrieb einer

Data Science Anwendung erfasst? *Are you currently tracking the used ressources regarding development and operation of AI models?*

3.3 Wie wird für eine ressourcenschonende KI-Modellierung während des Entwicklungsprozesses gesorgt? *Are there any actions or guidelines for a ressource-preserving development process of AI solutions?*

3.4 Sind die Entwickler:innen individuell für diese Maßnahmen zuständig? *Are the developers themselves responsible for the sustainability of their work?* ***or*** *Do you think that the developers should be responsible for the sustainability of their work?*

The transcribed interviews are not attached for document length reasons. Requests in this regard to Luise Wiesalla via Linkedin.

Teilbereich C – Ausblick (8min)

4.1 Wo sehen Sie die nächsten Schritte um einer umwelt- und ressourcenschonenden Gestaltung von Künstlicher Intelligenz den Weg zu ebenen? *Where do you see necessary steps for making AI more sustainable?*

4.1 Wie sehen Sie persönlich die Bedeutung von Nachhaltigkeit in der Künstlichen Intelligenz entwickeln? *How do you personally see the topic in future?*

A.2 MATURITY OF AI IN THE INTERVIEWED COMPANIES

Table A.1: coded passages reasons for categorize the interviewee in the maturity model

	Einstufung nach Maturity Model	Beurteilung aus Beleg	Belege
B1	Lvl 4 - Managed	Unterstützung Management, große Bedeutung (Lvl 4)	als Chief Digital Officer. Das sagt schon, wenn es so eine Rolle gibt, dann heißt es , dass das Unternehmen der Digitalisierung eine große Rolle beimisst. (
		Rollen definiert	Data Scientists machen eigentlich die meiste [Arbeit] (.) Data Engineerins natürlich auch / und verwandte Rollen machen eigentlich die meiste Arbeit
B2	Lvl 4/5 - Managed to Optimised	Datenhintergrund sehr umfangreich	Wir haben über viele Jahre - ich sage mal - wir nennen es immer so zwei Assets an Datentöpfen angesammelt
		Lange Zeit für Data Science	Und das ist tatsächlich 2014 gestartet worden und hat wirklich hier auch das Bewusstsein rund um Data Science weiter geschärft
		Schaffung der entsprechenden Infrastructur (Lvl 5) und Standardisierung	Chief Data and Analytics Architect, das heißt, ich werde mich jetzt ein bisschen wieder mehr darum kümmern in dieses Thema der Daten, Daten, die vorbereitet sein müssen, die eine gute Governance Struktur brauchen, die gemanagt sind, mich darauf zu fokussieren, diese Asset Daten einfach noch besser und professioneller bereitzustellen.
		Mitarbeiter werden eingebunden, AI cluture	Und erst letzte Woche für uns auch einen neuen Meilenstein erreicht hat, indem wir ein sogenanntes Data Literacy Programm aufsetzen, um Data Science in die Bereiche zu tragen
B3	Lvl 4/5 - Managed to Optimised	Top managment support (Lvl 4)	vermisse zum Beispiel einen Chief Digital Office im Senior Management (...) . Wobei wir allerdings den Vorteil haben, dass wir einen CEO haben, der einen guten Data Background hat.
		Zukauf von Daten	dass wir uns weitere Daten dazukaufen,
		Zentraler Führung (Lvl 5)	Data Bereich bei uns mit insgesamt um die 60 Personen doch deutlich groß ist.
		Reaktion in Echtzeit (Lvl 5)	Also so wie das Unternehmen aufgestellt ist und auch wie die Marktsituation sich ergibt, ist es für uns unerlässlich, dass wir mit sehr hoher Geschwindigkeit, im Endeffekt in quasi Echtzeit, auf das Marktgeschehen reagieren können
B4	Lvl 2 - Assessing	Noch nicht weit fortgeschritten (nicht Level 3)	Wir sind im Moment,würde ich sagen, in dem Prototypen Status
		KI auf dem Schirm und gewillt	Data Science würde ich sagen, ist ein strategisches Potenzial.
B5	Lvl 2 - Assessing	KI in Einsatz	The role of Data Science, AI , Digitalization is more or less to automatize manual work
		Indivudelle Beratung, noch nicht standardisiert	I am consulting the business owners of external partners and suitable solutions.

		Noch nicht durchdringt (kein Lvl 3)	Immer wieder kommen Machine Learning Algorithmen dann auch zur Geltung. Jetzt abgesehen von dem Beispiel, das ich vorhin genannt habe, aber proportional ist das aber ein Stück weit weniger
B6	Lvl 2 - Assessing		
		Einsatz von produktiven Modellen (Lvl2)	Und in diesem Milieu zumindest haben wir dann auch tatsächlich dann auch angefangen, Machine Learning Algorithmen zu implementieren
		Laufende Modelle, Verknüpfung von Daten (Lvl 2-3)	Den Verkaufserfolg bewerten und das Ganze mit redaktionellen Daten verknüpfen, um hinterher Auswertungen zu machen.
B7	Lvl 2 - Asessing		
		Änderung der bestehenden Infrastruktur (Lvl2)	Und in dem Zusammenhang haben wir oder habe ich mein Team technologisch neu aufgestellt mit Cloud Technologien
		Management Support (Lvl 3)	Ja einer der Hauptpfeiler der generell IT-Beratung
		Fortgeschrittene IT projecte, Produktiv laufen werden keine Modelle bei einer Beratung (Lvl 3)	Das setzen wir dann im internen in Demos um, die wir dann dem Kunden präsentieren können
B8	Lvl 2/3 - Assessing to Determined		
		Aufbereitung der Daten für KI (Lvl 2)	Wir kriegen im Endeffekt einen Datensatz oder Datensätze oder Zugang zu einer Datenbank, schauen, dass wir die Daten irgendwie aufbereiten, dass die das die nutzbar sind oder vereinheitlichen
		Lot of developers	But what we did was to collect all other analytics teams together and create this bigger organization. And now we have an organization of 300 people
		fortgeschrittene Anwendungsfälle, Transaktionsdaten sind geschäftskritisch	like recommendation engines, personalization, transactional data
		Auslagerung Prozesse definiert	Except the operations, we don't do operations, we only do the cool stuff.
		Wert von KI is bekannt	We look at the business value of it
		Standardisierung	we actually have to document it because we also need to go through model validation phase
B9	Lvl 4/5 - Managed or Optimised		
		Management Unterstürtzung	I work very closely with the chief analytics office
		Standardisierung, Rollen und Verantwortlichkeiten sind definiert, KI Kultur (Lvl 5)	we also create best practices for the whole organization. So we actually train both like business teams and also the rest of the analytics organization in terms of best practices of ML, operations, software engineering and how to write better code and this kind of thing
		Definierter Mehrwert durch KI und Support durch Gründer (Lvl 4 auf Organisationsebene)	Ja, es hat sehr große Bedeutung, weil die Hypothese ist, dass es funktioniert im Bio Anwendungsfall. Das heißt am Ende ist bei uns Data Science dann auch sehr spezifisch das Verkaufsargument gerade.
B10	Lvl 2 - Assessing		
		Erst der Aufbau der Struktur, die für Level 3 benötigt wird	Wir sind och ein ganz frisches Start-up, wir haben noch kein Produkt. Das heißt, alles was wir jetzt machen, ist quasi Entwicklung.
		Entdeckung der Technologie	Auf der anderen Seite ist das bei Banken und Versicherungen nur zum Teil angekommen

B11	Lvl 2 - Assessing		
		Entdeckung der Technologie	Und ja, da sind Banken und Versicherungen was was wir beraten noch total weit vorne weg und von daher ist das überhaupt erst mal ein Prozess dahin zu kommen was machen wir überhaupt mit KI und was machen wir mit Data Science?
		KI bezogene Einstellungen	Die Bedeutung nimmt zu. Wir stellen immer mehr Data Scientist ein, und das wird in allen möglichen Unternehmensbereichen eingesetzt.
		KI bezogenes training, fortgeschrittenes Stadium der Produktivsetzung erreicht	Schulen viel. Trainieren viel. Genau. Und implementieren Einzellösungen. (nach Rückfrage automatisiert)
B12	Lvl 3 - Determined	Standard Arbeitsanweisungen für KI Entwicklung	Es gibt bei uns Development Guidelines
		Veröffentlichung eines Tools (Schritt in Richtung Level 4)	Und wir haben selber eine sogenannte Nano Production App entwickelt, die lokal läuft, mit der jeder User, jede Userin Time Series Forecasts erstellen kann.

A.3 RESULTS OF QUALITATIVE CONTENT ANALYSIS

A.3.1 *Reasons for sustainability in companies*

Table A.2: document statistics reasons sustainability

Category Name	Absolute Count	% of SUM	N of Documents	% of Documents
Economic external	*33*	*36*	*11*	*91*
Image of Corporation	11	12	6	50
Market Demand	18	19	6	50
Competitive Advantage	4	4	4	33
Political Reasons	*9*	*9*	*6*	*50*
Political Reasons	9	9	6	50
Bottom Up	*14*	*15*	*8*	*66*
Personal interest	6	6	3	25
Employee Initative	8	8	6	50
Top Down	*8*	*8*	*6*	*50*
Management Initative	8	8	6	50
Social Responsibility	*6*	*6*	*5*	*41*
Social Responsibility	6	6	5	41
Financial Markets	*5*	*5*	*4*	*33*
Financial Market	5	5	4	33
Economic internal	*16*	*17*	*8*	*66*
Economic Efficiency	10	10	4	33
Employee Recruitment	6	6	4	33

Table A.3: coded passages reasons for sustainability

Interviewee	Category Title	Marked Text
B09	Competitive Advantage	If you don't react, then you'll probably be left behind. I mean, it's just like adopting a new technology, right.
B11	Competitive Advantage	W ir haben ein ein Produkt, das wir den Kunden anbieten, um deren Kunden wiederum nachhaltige Produkte zu empfehlen.
B3	Competitive Advantage	Bei den klassischen Ride Hailing ist es im Moment für uns eher so was wie ein Alleinstellungsmerkmal. Ist noch nichts, was jetzt gefordert wird, was dann so etwas wie ein Must have wäre, sondern eher etwas, wo man sagen kann, das setzt uns vom Wettbewerb ab.
B8	Competitive Advantage	irgendwo in Produkte packen und so, aber das ist mehr für die Firma gesprochen
B1	Economic Efficiency	Shortages / und Metalle sind sehr energieintensiv
B1	Economic Efficiency	Aber auch natürlich Themen wie Kreislaufwirtschaft, Ressourcenknappheit und so was hängen damit zusammen. Energie natürlich gerade jetzt im Brennpunkt wieder mal
B1	Economic Efficiency	Und für die Wirtschaft ist einer der großen Aspekte natürlich nicht nur etwas Gutes zu tun, sondern gleichzeitig auch Kosten zu sparen und resilient zu werden.
B1	Economic Efficiency	Wie kann ich Ressourcen sparen, um weniger abhängig zu werden von externen (..) Lieferketten und Materialien? Wie kann ich Kosten sparen bei steigenden Energiepreisen, zum Beispiel steigenden Rohstoffpreisen? Effizienz ist sowieso immer ein großes Thema und das steigert eben die Resilienz. Und es wirkt sich natürlich positiv auf den Profit eines Unternehmens aus. Und das ist natürlich höchstes Interesse in der Wirtschaft, ganz klar.
B1	Economic Efficiency	Und dann natürlich, wie ich eben gesagt habe, das ganze Kostenthema, was eher indirekt ist.
B1	Economic Efficiency	Aber wie gesagt, in der Wirtschaft zählt erst mal mehr Kosten. Und klar, das ist immer so ein Win-Win. Spart man Kosten, spart man meistens auch damit Ressourcen.
B12	Economic Efficiency	aus der Ressourcenknappheit.
B5	Economic Efficiency	circular, circular economy and green economy,
B5	Economic Efficiency	ban of importing material from Russia
B6	Economic Efficiency	Ich glaube, das ist immer nett für die Unternehmen an sich natürlich. Die können dann Kosten sparen im Zuge dessen
B10	Employee Initative	uns in der Firma jetzt so der Idealismus ist,

B12	Employee Initative	Wir haben gerade vor zwei Wochen einen Drei Tages Innovation Space gemacht, wie unser persönlicher Software Carbon Footprint aussieht und wo wir den reduzieren können.
B12	Employee Initative	Das liegt in meinem persönlichen Umfeld. Jetzt an dem Team, den Leuten.I
B3	Employee Initative	Also den besten Hebel haben sicherlich die Cloudbetreiber und wenn ich mir angucke, die beiden Großen, also Amazon und Google, haben ja auch durchaus mit ihren Mitarbeitern relativ viele aus der jungen Generation, die das Thema auch sehr, sehr hochhalten und die dort intern eine Menge Druck aufbauen werden
B4	Employee Initative	Also, ganz allgemein ist das, glaube ich, für viele Mitarbeiter in unserem Unternehmen Herzensthema.
B4	Employee Initative	Die Mitarbeiter selber sind dann im Moment oft begeistert mit Möglichkeiten.
B5	Employee Initative	The employees are taking initiative
B6	Employee Initative	Und das war etwas, das quasi auch wirklich von uns Mitarbeitenden kam, jetzt weniger von der Oberen, äh Exekutives. Und basierend darauf kam eigentlich viel die Motivation da, sich auch dran zu halten und auch mehr in Richtung Nachhaltigkeit zu denken und es voranzubringen.
B11	Employee Recruitment	Und zum anderen ist es, glaube ich, wichtig, um die Mitarbeiter zu halten und zu gewinnen. Das ist wahrscheinlich in der Zukunft der größte Punkt, dass die neuen Mitarbeiter sehr darauf achten werden, dass sich das Unternehmen nachhaltig aufstellt. Und da haben wir auch schon diverse Maßnahmen unternommen.
B2	Employee Recruitment	Und ich nehme wahr - ich bin auch selbst eingebunden in das ganze Thema der Ausbildung - hier bei der Betreuung junger Kollegen, dass für die natürlich so was auch eine Herzensangelegenheit ist.
B2	Employee Recruitment	Und ich glaube auch, auch da vor dem vor der Brust, dem demografischen Wandel, Fachkräftemangel, glaube ich, muss man als Unternehmen einfach auch den jungen Menschen Nachhaltigkeit als Wert vermitteln, damit man sie überhaupt gewinnt als Fachkraft für ein Unternehmen wie einen Flughafen zu arbeiten.
B4	Employee Recruitment	wenn die Mitarbeiter das Gefühl hätten, dass wir einen sozial oder für die Umwelt schädlichen Dienst, schädliche Arbeit tun oder unser Geld damit verdienen, dass wir Leute ausbeuten oder Miet-Mogul sind oder so. Ich glaube, dann würden wir ganz schnell viele Mitarbeiter verlieren.
B8	Employee Recruitment	Ich meine, ich glaube, viele Leute fühlen sich derzeit eher nachhaltig agierenden Unternehmen verpflichtet oder nicht verpflichtet, angezogen
B8	Employee Recruitment	Und wie du sagst, junge Leute anzuziehen. Das ist schon eine große Motivation, sonst eben tatsächlich politisch mit CO2 abgaben und so weiter haben, so fein kann das wahrscheinlich irgendwann gerechnet wird.
B09	Financial Market	some sustainability index by the ECB - European Central Bank.
B09	Financial Market	all this started actually giving out this green loans. For example, in 2019, [anonymized - other company] asked for a loan and then [anonymized] actually said something, some goals and said if you, I don't know, increase your sustainability index this much or if you keep it at this level, we will give you a better interest rate offer.
B11	Financial Market	und hat dann im Prinzip für für die Banken einmal direkte Auswirkungen in der Kreditwürdigkeit selber und in der Außendarstellung der Banken.
B2	Financial Market	Aber ich denke, dass die Refinanzierungsstrukturen sich auch verändern werden. Und ich glaube, dass man da in Zukunft auch stärker auf dieses Thema der Nachhaltigkeit blicken wird. Zu welchen Konditionen kriege ich da meine Refinanzierung. Wird sicherlich davon abhängen, wie nachhaltig ich wirtschafte oder welche Nachhaltigkeitsziele ich mir setze und wie die Entwicklung dahin aussieht. Von daher glaube ich, dass wir den Druck auch von der Finanzbranche bekommen werden, da aktiv zu sein.
B4	Financial Market	Aber was ich sehe, ist, dass vor allem die institutionellen Investoren sehr großen Wert darauf legen. Also für dieses inzwischen ein Muss.
B1	Image of Corporation	Nehmen wir mal an, der CEO, also der Chief Executive Officer der Firma oder das Leadership hat einen ganz starken, sozusagen grünen Bezug irgendwie. Dann bringen die das natürlich mit ein. Das gehört dann aber zum Branding der Firma dazu.
B1	Image of Corporation	von der Unternehmensführung durch eine starke Überzeugung her getrieben wird.
B10	Image of Corporation	einfach ein gutes Marketing.
B2	Image of Corporation	Also ich meine, als Flughafen stehst du immer im Generalverdacht, nicht besonders nachhaltig zu sein.
B2	Image of Corporation	Wir haben diese Demonstrationen bei uns in den Terminals. Also wir sehen das ganz direkt und unmittelbar.
B2	Image of Corporation	Klar, aber das halt im Einklang mit diesen Nachhaltigkeitszielen auch zu vereinbaren und das wird dadurch, dass wir auch so sichtbar sind, haben wir das vor der Brust und wir erleben das ganz direkt, wie gesagt, von außen oder von innen.
B3	Image of Corporation	Und am Ende ist halt auch das, was wir als Unternehmen uns auf die Fahne geschrieben haben, dass wir in die Richtung wollen.
B3	Image of Corporation	Bei den klassischen Ride Hailing ist es im Moment für uns eher so was wie ein Alleinstellungsmerkmal. Ist noch nichts, was jetzt gefordert wird, was dann so etwas wie ein Must have wäre, sondern eher etwas, wo man sagen kann, das setzt uns vom Wettbewerb ab
B3	Image of Corporation	Das sind schon Themen, die nachgefragt werden. Allerdings sind wir noch nicht so weit, dass Anbieter, die das nicht haben, abgelehnt werden. Es ist eher etwas, was den positiven verstärken Effekt hat.
B5	Image of Corporation	circular, circular economy and green economy

B6	Image of Corporation	Allerdings werben wir auch darum, dass wir Ökostrom benutzen, um die Server zu betreiben.
B09	Management Initative	You always see it like every senior manager has sustainability as part of their goals and this also has an impact on us. We actually already worked on some sustainability projects also in our team
B10	Management Initative	bei uns in der Firma jetzt so der Idealismus ist
B10	Management Initative	Grundideen war, dass man damit kann man vielleicht sogar helfen, nachhaltiger zu werden, indem man plastikfressende Proteine findet.
B12	Management Initative	unserer strategischen Ziele auf die Fahne geschrieben, dass wir nachhaltig selbst arbeiten, dass wir unseren Kunden ermöglichen, nachhaltiger zu produzieren.
B12	Management Initative	l in der IT eine eigene Stabsstelle für Sustainability@IT,
B2	Management Initative	Wir haben ein Umweltmanagement bei uns und jetzt sind wir gerade dabei ein wirkliches Nachhaltigkeitsmanagement aufzusetzen und aufzubauen.
B5	Management Initative	but also the management
B6	Management Initative	Ja, tatsächlich so ein Gemisch von allem
B09	Market Demand	sustainability index by the ECB - European Central Bank.
B09	Market Demand	Now somehow the world is changing. Right. And then if you don't act on these kind of things, then customers also get away from you
B1	Market Demand	Ich meine, nur in die Medien gucken, da siehst du was los ist.
B1	Market Demand	Mit sehr viel Pressezeit
B1	Market Demand	Das andere [der Treiber] ist, die Kunden wollen es. Das heißt, wenn es einen starken Bedarf nach umweltfreundlicheren Produkten gibt, wenn die Kunden selbst draufgucken würden
B1	Market Demand	Aber diese Leuchttürme - diese Vorbilder - kommen eigentlich nur, wenn es die Kunden wirklich wollen
B1	Market Demand	Aber in dem Falle ist das meistens erst mal getrieben dadurch, dass es entweder verlangt wird, dass die Kunden zum Beispiel auf den Carbon Footprint draufgucken - dass es eben sustainable sein muss.
B11	Market Demand	und hat dann im Prinzip für für die Banken einmal direkte Auswirkungen in der Kreditwürdigkeit selber und in der Außendarstellung der Banken.
B11	Market Demand	der Kunde bringt den den Druck und auf den reagieren wir auch schon.
B11	Market Demand	Und da ist es einmal die die Kunden werden es verlangen, die die Banken und Versicherungen selber, dass man sich nachhaltig aufstellt.
B3	Market Demand	Also gerade bei uns in dem Markt, in dem wir uns befinden, ist das Thema Nachhaltigkeit ein sehr wichtiges geworden, weil gerade mit dem Eintritt diverser Mitbewerber gab es ja doch relativ viel Diskussion darüber, inwieweit gerade dieses ganze Thema "Ride Hailing" oder Transportation in Städten eigentlich zu einem negativen Effekt führen könnte, dadurch, dass man dann vielleicht mehr Fahrzeuge auf der Straße hat.
B3	Market Demand	Und wenn man sich dann noch anguckt, wie sich unsere, unsere Zielgruppen zusammensetzen, da haben wir auch sehr, sehr viele Personen aus den jüngeren Zielgruppen, so Generation Y and Z
B3	Market Demand	Halt einmal, weil wir die Zusammenarbeit mit den Städten damit optimieren können und für die Städte natürlich auch ein interessanter Partner werden. Und auf der anderen Seite aber halt auch unsere Kunden das natürlich sehr gerne mögen.
B3	Market Demand	Bei Themen wie wie bei diesen Rollern, da geht es schon sehr stark in die Richtung, dass das verlangt wird.
B3	Market Demand	Nachhaltigkeit ist im Moment wirklich ein Thema, was natürlich auch zu einem Alleinstellungsmerkmal in dem Unternehmen führt.
B4	Market Demand	Wir haben eigentlich einen guten Beitrag in Nachhaltigkeit, haben eigentlich einzelne Initiativen. Wir haben zum Beispiel jetzt einen ESG Newsletter den wir intern an unsere Investoren vertreiben, wo wir eben Initiativen aufzeigen, aber auch Entwicklung in dem gesamten Real Estate Markt, die es in dem Bereich gibt, teilen.
B4	Market Demand	Also wir haben Investoren, die ganz konkret fordern, dass die Portfolios, die wir für sie aufbauen und betreiben, quasi das es einen klaren ESG-Plan gibt.
B7	Market Demand	sich die Einstellung der Konsumenten verändert zu dem Thema, dass Nachhaltigkeit einfach in den Köpfen stärker verankert ist, das stärker wahrgenommen wird, dass alle, insbesondere auch Unternehmen, eine Verantwortung haben.
B1	Personal interest	Nachhaltigkeit ist ja eigentlich ein sehr alter Hut, ohne das negativ zu meinen. Also Sustainability/ Sustainability Development Goals zum Beispiel sind ja echt alt. Brundtland /
B1	Personal interest	Ich zitiere mal gerne "Limits to Growth".
B1	Personal interest	Aber ich kann mir natürlich viele Projekte vorstellen und ich kenne auch viele Projekte aus dem Data Science, aus Beispielen, also Use Cases, wo das natürlich absolut super überall eine Rolle spielt.
B1	Personal interest	eben eine besondere Überzeugung existiert im Unternehmen.
B12	Personal interest	h war sozusagen Product Owner und mein Ziel war es nicht nur für uns selber herauszufinden, wo haben wir Energie, Einsparmöglichkeiten, sondern auch Wo ist unser Fußabdruck denn groß, eher klein?
B8	Personal interest	Also ganz persönlich muss ich sagen, finde ich das einfach auch.
B09	Political Reasons	Actually there's a lot of interest in sustainability, I guess because of some external regulations.

B1	Political Reasons	Ich habe meinen Doktor in [Ort anonymisiert] gemacht hat, das ist das glaube ich noch mal eine Stufe stärker. Da wird es noch mal mehr getrieben. Durch Auflagen und durch ein Bewusstsein.
B2	Political Reasons	Politik
B2	Political Reasons	Ich glaube, es geht auch viel bei uns da drum, - vielleicht auch noch mal als Motivation Nachhaltigkeit - viel auch der Förderungen mit abzugreifen, die in Zukunft dann sozusagen aufkommen werden.
B5	Political Reasons	CO_2 taxes.
B5	Political Reasons	on the national level.
B6	Political Reasons	Ja, tatsächlich so ein Gemisch von allem
B8	Political Reasons	auch politisch ist es natürlich immer mehr gefragt
B8	Political Reasons	Sonst eben tatsächlich politisch mit CO_2 abgaben
B1	Social Responsibility	da gab es eine große Studie von Wissenschaftlern des MTI, die ein System Dynamics Modell der Welt gebaut haben und vorhergesagt haben, dass wir in ein Sustainability/Ressourcen/Nachhaltigkeitsprobleme reinkommen werden, wenn wir den Business as usual einfach weiter folgen.
B1	Social Responsibility	Ich habe meinen Doktor in [Ort anonymisiert] gemacht hat, das ist das glaube ich noch mal eine Stufe stärker. Da wird es noch mal mehr getrieben. Durch Auflagen und durch ein Bewusstsein.
B10	Social Responsibility	Na ja, für alle Menschen vielleicht ein wichtiges Thema.
B12	Social Responsibility	aus dem Klimawandel gegeben
B3	Social Responsibility	Auf der andern Seite muss ich auch sagen, dass die alleine es nicht nicht werden hinbekommen können. Und ich sehe auch auf Unternehmensseite bei Unternehmen wie uns eine Verantwortung, da sinnvoll mit den vorhandenen Ressourcen umzugehen.
B8	Social Responsibility	Und ja glaube ich gerade in der IT Beratung, wenn man ja dafür verantwortlich ist, im Endeffekt, dass die Unternehmen jetzt mehr und mehr IT nutzen, was wir dann zu dem Wachstum des Verbrauchs im Endeffekt beiträgt.

A.3.2 *Relevance of sustainability in the data science department*

Table A.4: document statistics relevance sustainability

	Category Name	Absolute Count	% of SUM	N of Documents	% of Documents
no relevance		*36*	*34*	*12*	*100*
	no priority in daily business	27	26	11	91
	no capturing of ressource consumption	9	8	7	58
Indirect actions		*58*	*56*	*12*	*100*
	individual responsibility/actions of developer	8	7	5	41
	sustainable choice of products and vendors	15	14	8	66
	capturing substitute metric	8	7	4	33
	unintentional sustainable	27	26	11	91
Direct actions		*9*	*8*	*2*	*16*
	first active steps	6	5	2	16
	capturing emissions	3	2	1	8

Table A.5: coded passages relevance of sustainability

Interviewee	Category Title	Marked Text
B12	capturing emissions	Und wir haben selber eine sogenannte Nano Production App entwickelt, die lokal läuft, mit der jeder User, jede Userin Time Series Forecasts erstellen kann. Und da haben wir ein Snippet integriert Code Carbon, was jeden Algorithmus ausgibt, wie viel CO2 emittiert wurde mit dem gerade abgeschlossenen Lauf.
B12	capturing emissions	. Und speziell bei einem der großen Cloud Anbieter gibt es inzwischen Dashboard für die Anwender, wie viel ihres letzten Monats ihres letzten Quartals CO2 emittiert wurde, ihrer Nutzung.
B12	capturing emissions	Ja, also wie gesagt, es ist noch nicht im Alltag im Monitoring drin, aber zumindest nutzen wir die Möglichkeit, dass wir uns da ein Bild verschaffen und dass wir da ein Bewusstsein herstellen.
B3	capturing substitute metric	Und dabei meine ich, würde ich den Verbrauch an Rechenleistung - dieser Art Impact - haben wir uns bisher nicht angeguckt. Da gibt es halt mehr das Thema, dass wir uns angucken, was ist der Kostenfaktor da.
B3	capturing substitute metric	wo Rechner im Hintergrund arbeiten, den charmanten Vorteil, dass das natürlich mit dem verbrauchten Geld und den verbrauchten Ressourcen sehr eng aneinander gekoppelt ist. Insofern hat man da einen gewissen Automatismus.
B3	capturing substitute metric	Also was ich konkret sehe ist aber im Endeffekt die Rechnung, die ich für die Ressourcen in der Cloud jeden Monat bekomme. Die kann ich auch im Endeffekt auf einzelne Modelle runterbrechen. Damit kriege ich schon einen ganz guten Überblick darüber, was einzelne Modelle an Kosten und auch an Ressourcenverbrauch erzeugen.
B5	capturing substitute metric	We track the amount of time that it required, but not in the/ but only from the time perspective, not from energy perspective. So from time perspective, we have already some kind of a view of what it's already taking or consuming time.
B09	capturing substitute metric	We do it project by project but I don't have them in a central place.
B09	capturing substitute metric	How much time we spend on training, how much memory it uses, and how much storage we need for this and all these kinds of things. Yeah. We don't use this information so much like later.
B09	capturing substitute metric	But we don't store this information later.
B11	capturing substitute metric	Kosten nicht explizit, weil wir die zum Teil gar nicht kennen, aber die Zeit schon. Also es wird einfach gemessen. Wie lange läuft das Modell? Wie lange braucht das dafür? Allerdings nicht aus Nachhaltigkeitsgesichtspunkten, sondern einfach um eine Abschätzung zu bekommen.
B8	first active steps	. Ich denke, das war schon immer irgendwie ein Thema. Man versucht es jetzt auf jeden Fall noch mehr zu intensivieren
B8	first active steps	wir machen das ökonomisch, dass wir schauen, können wie wir irgendwie das auch als ein Produkt machen, quasi im weitesten Sinne, dass man, dass man sagt, nachhaltige Beratung auch als Produkt
B8	first active steps	a, auch die, die Codebasis, die ganze Modellierung und die Softwareentwicklung, dass man auch einen gewissen stärkeren Fokus nicht mehr nur mit dem Gedanken Ja, es soll schnell laufen, sondern auch es soll schon auch Ressourcen optimiert laufen

B12	first active steps	Ich war sozusagen Product Owner und mein Ziel war es nicht nur für uns selber herauszufinden, wo haben wir Energie, Einsparmöglichkeiten, sondern auch Wo ist unser Fußabdruck denn groß, eher klein?
B12	first active steps	Und wir haben selber eine sogenannte Nano Production App entwickelt, die lokal läuft, mit der jeder User, jede Userin Time Series Forecasts erstellen kann. Und da haben wir ein Snippet integriert Code Carbon, was jeden Algorithmus ausgibt, wie viel CO2 emittiert wurde mit dem gerade abgeschlossenen Lauf.
B12	first active steps	Wir haben uns vor kurzem angeschaut, wie verschiedene Programmiersprachen im Verhältnis zueinander performen und welchen Fußabdruck hinterlassen. Und als ganz einfaches Beispiel schneidet da Java wesentlich besser ab als Python.
B1	individual responsibility/actions of developer	Das ist immer bei den Leuten, die das Modell natürlich bauen.
B1	individual responsibility/actions of developer	Genau das ist die Verantwortung. Eben von einem Data Scientist (..) der guckt darauf / oder die [Data Scientistin]. Du weißt immer der oder die / Es kommt einfach darauf an. Die müssen ja auch so denken, dass sie sagen okay, es lohnt sich nicht, ein riesen komplexes Modell zu bauen und damit so eine kleine, minimale Verbesserung zu erreichen. Das ist schon automatisch in deiner Ausbildung drin, dass du willst das auch so eine Art Pareto Prinzip - 80 20/
B1	individual responsibility/actions of developer	Vom Entwickler. Das ist einfach nur Best Practice.
B2	individual responsibility/actions of developer	Und vor dem Hintergrund versuchst du dann schon da, wo es für dich pragmatisch Sinn macht, das zu optimieren.
B2	individual responsibility/actions of developer	Ich glaube heute, die Optimierung kommt viel vom Entwickler. Heute sehe ich das erstmal aktuell noch sehr stark vom Entwickler kommend.
B3	individual responsibility/actions of developer	Also man guckt natürlich, dass man nichts mit einbaut, was jetzt sinnlos Ressourcen verschwendet.
B7	individual responsibility/actions of developer	Gibt es auf jeden Fall keine klaren Konzepte für. Eine ganz klar Individual Entscheidung.
B12	individual responsibility/actions of developer	Auch da sind wir gerade am Anfang und im Moment sind es noch die Individuen, die davon wissen oder nicht
B1	no capturing of ressource consumption	Nein, eigentlich nicht.
B1	no capturing of ressource consumption	Derzeit wird nur [geschaut] "Wie akkurat ist mein Modell?" Es wird erst dann darauf geguckt, wenn es einen signifikanten Kostenanteil verursacht.
B1	no capturing of ressource consumption	Wenig. Wenig, wenig bis gar nicht.
B2	no capturing of ressource consumption	Nein, wenn dann nur in so einer hoch aggregierten Form. Nach dem Motto, was unser Data Center vielleicht als als Strom im Jahr verbraucht und dann vielleicht mit einem Schlüssel weiter verrechnet, aber aber gar nicht runtergebrochen, glaube ich auf Applikationen oder Domänen
B4	no capturing of ressource consumption	Da sind wir noch nicht.
B6	no capturing of ressource consumption	Aber das ist etwas, das ist uns tatsächlich auch im Hinterkopf geblieben, dass wir das gerne machen wollten. Einfach auch, um Dinge zu optimieren.
B7	no capturing of ressource consumption	Nein, wird es nicht. Aber jetzt in der Cloud haben wir natürlich wirklich die Möglichkeit, ganz anderes Monitoring noch aufzubauen, deutlich noch anwendung individuelles Monitoring aufzubauen. Das ist auch auf jeden Fall der Plan. Aber natürlich auch da weniger, sage ich mal mit der Nachhaltigkeitsperspektive als mehr mit der Kostenperspektive.
B8	no capturing of ressource consumption	Wenn das aber intern oder wie hier auf , also wenn ich auf meinem Laptop jetzt zwei, drei Tage irgendwas mache, dann wird das soweit ich weiß, nicht erfasst
B10	no capturing of ressource consumption	Um nachhaltiger zu sein, würden wir kein Tracking aufsetzen. Also noch nicht.
B1	no priority in daily business	Also direkt im Data Science Bereich ist es noch nicht so relevant.
B1	no priority in daily business	Aber wenn du eigentlich Data Science oder KI im Unternehmensumfeld betrachtest, dann ist der Fokus, wie ich eben gesagt habe / (&) der [Fokus] ist auf Prozess, auf Produkt und auf Geschäftsmodell.
B1	no priority in daily business	Natürlich willst du ein Produkt entwickeln, ein Auto zum Beispiel, das weniger Treibstoff verbraucht. Im Endeffekt es muss es aber verkauft werden. Das heißt ganz klar, das spielt immer eine Nebenrolle derzeit, nur - leider.
B1	no priority in daily business	Danach wählst du ja auch deinen Data Science Use Case aus. Da wo du mit geringstmöglichen Aufwand den höchstmöglichen Return hinkriegst. Aber eben (...) monetär hauptsächlich
B2	no priority in daily business	Wir haben - man muss halt auch sagen - wir haben so viel Druck aktuell auch an einer anderen Stelle, dass wir einfach diesem Fachkräftemangel entgegen laufen
B2	no priority in daily business	Und vor dem Hintergrund ist jetzt erst mal/ Zuerst müssen wir wirklich erst mal wieder unsere Qualität sichern. Das heißt, der Passagier, der will eben nicht zwei Stunden in der Halle warten auf seinen Koffer vorher. Es ist schwierig, über Nachhaltigkeit in allen Aspekten dann zu philosophieren und das zu berücksichtigen
B2	no priority in daily business	Und dann glaube ich wird das, wenn wir das gelöst haben, ist es wie so eine Bedürfnispyramide und wir sind am untersten Level und dann kommen wir der Bedürfnispyramide nach oben und dann können wir uns auch noch ein bisschen stärker mehr auf Nachhaltigkeit [fokusieren].

B3	no priority in daily business	lso im Moment ist es halt meistens so, dass wir gerade dabei beim Modellieren erst mal nach Möglichkeiten suchen, überhaupt die, die Anforderungen abdecken zu können. Das ist mit den Daten, die man so zur Verfügung hat und auch vor dem Hintergrund der Datenschutzgrundverordnung immer schon ein herausforderndes Unterfangen.
B4	no priority in daily business	Und ich glaube, für mich als Projektleiter ist immer am Anfang des Projekts, die auf Pragmatismus zu trimmen und vergisst die Optimierung
B5	no priority in daily business	If we have a backlog of tasks where we should minimize the energy footprint, the resources that computers consume within the manufacturing branches is probably the least
B5	no priority in daily business	But as I said for us right now, it's with the lowest priority in the future. I don't see within our field that is coming and I don't see it honestly, I don't see that would come in any kind of priority.
B5	no priority in daily business	Because we have a lot of tasks that is not relevant to sustainability in terms of computation. It's more of other topics.
B5	no priority in daily business	There are bigger fishs out there.
B6	no priority in daily business	Tatsächlich nicht. Also es ist auch für mich ein extrem neues Thema, ein absolut notwendiges und tolles Thema
B6	no priority in daily business	So retrospektiv, wenn man wenn ich darüber nachdenke, wäre das definitiv ein Aspekt gewesen, den wir hätten mit einbeziehen sollen in der Findungsphase.
B6	no priority in daily business	Ich glaube wir hatten einfach nicht nicht in den Zeit und nicht nicht die Priorität.
B7	no priority in daily business	Also überhaupt nicht. Das wäre mir überhaupt nicht bekannt.
B7	no priority in daily business	. Ja, da hast du natürlich schon auch große Datenmengen, weil natürlich der ganze Content aus Texten, Bildern, Videos etc. besteht. Da weiß ich nicht, ob es da vielleicht das irgendwie schon eine Rolle spielt. Da wir wie gesagt leider noch relativ weit weg davon sind und ich will da näher rankommen.
B7	no priority in daily business	Und vielleicht spielen dann ja irgendwann solche Dinge auch auch mehr eine Rolle. Aber Stand heute? Hat sich glaube ich über Nachhaltigkeit, Stromverbrauch oder ähnliches bei uns keiner Gedanken gemacht
B7	no priority in daily business	Also dass es kein Thema sein wird, das besonders nachhaltig zu gestalten. Ich glaube, ich kann mir vorstellen, dass das, dass viele Unternehmen eigentlich in der selben Situation wie wir sind, dass es eher für sie erst mal gerade die Herausforderung ist, überhaupt die Kompetenzen aufzubauen, überhaupt die richtigen Use Cases zu finden, um überhaupt, sage ich mal, an dem Nutzen von Deep Learning zu partizipieren.
B7	no priority in daily business	Also ich kann mir vorstellen, dass Unternehmen weniger solche KPIs im Auge hätten als eher Performance KPIs, also Performance im Sinne von Wie akkurat sind meine Modelle?
B7	no priority in daily business	Also wenn die wenn eine längere Berechnung ein besseres Ergebnis produziert, könnte ich mir vorstellen, dass das die meisten Unternehmen eher das bessere Ergebnis bevorzugen würden als die kürzere Rechenzeit.
B09	no priority in daily business	We consider the outcome as money. We look at the business value of it, but so far we haven't really considered the sustainability of it.
B09	no priority in daily business	For example, if a solution is going to be really expensive to produce and if it's going to be used only for a short amount of time, then it doesn't make sense to invest in this because these resources that we put in this are actually very expensive, especially the human resource on this.
B10	no priority in daily business	Hm. Nee.
B11	no priority in daily business	Und ja, da sind Banken und Versicherungen was was wir beraten noch total weit vorne weg und von daher ist das überhaupt erst mal ein Prozess dahin zu kommen was machen wir überhaupt mit KI und was machen wir mit Data Science? Und der nächste Schritt ist dann eigentlich erst mal Was müssen wir denn da alles beachten dabei?
B12	no priority in daily business	Es gibt bei uns Development Guidelines, die sind aber noch nicht Nachhaltigkeit optimiert.
B2	sustainable choice of products and vendors	dann die eigenen Data Scientist aktivieren oder dass sie eigentlich über den Markt versuchen zu gucken, ob diese Probleme schon jemand für einen gelöst hat. Also man macht nicht immer alles selbst als Flughafen, oft geht man halt raus und guckt erst mal hat man für diese Fragestellungen schon andere gute, clevere Lösungen
B2	sustainable choice of products and vendors	Ja, im Moment ist halt auch die Frage, inwieweit gibt man diese Verantwortung auch ein Stück weit ab an den Betreiber eines Rechenzentrums und verlangt von ihm Nachhaltigkeit?
B3	sustainable choice of products and vendors	Es gibt ja Cloud Betreiber, die ihr Rechenzentrum in der Form nachhaltig machen, dass sie die Abwärme als Heizwärme oder wieder zur Energieerzeugung benutzen.
B3	sustainable choice of products and vendors	Das ist nicht, was wir heute verwenden, weil da eher der Fokus darauf steht, einen der großen Betreiber zu verwenden, die sehr standardisiert ihre Leistungen anbieten, sodass es dann deutlich einfacher wird, mit den vorhandenen Personen, die man im Team hat, die auch zu nutzen.

B3	sustainable choice of products and vendors	o Amazon und Google, haben ja auch durchaus mit ihren Mitarbeitern relativ viele aus der jungen Generation, die das Thema auch sehr, sehr hochhalten und die dort intern eine Menge Druck aufbauen werden. Und die Firmen haben natürlich auch nach außen hin einen großen Bedarf, sich da positiv darzustellen. Insofern habe ich die Hoffnung, dass dort eine Menge passiert
B4	sustainable choice of products and vendors	Wir haben das diskutiert im Rahmen von der Auswahl unserer Cloudanbieter, wo die sich natürlich auch unterschiedlich aufstellen. Also Google macht dann seine, glaube ich, sehr viel stärkeres Versprechen als Amazon
B4	sustainable choice of products and vendors	Das ist, glaube ich, am Ende trotzdem nicht das Ausschlag oder noch nicht noch nicht wahrscheinlich das ausschlaggebende Kriterium. Zweck der Arbeit ist es, ein offenes Interesse da zu sein, bei einer Technologieauswahl AUCH darauf zu schauen.
B7	sustainable choice of products and vendors	Und das insgesamt natürlich die Cloud Provider einen deutlich größeren Hebel haben. Weil das ist einfach deren Kerngeschäft und sie verkaufen das allen Unternehmen und wenn sie es auf ihrer Seite schaffen, besonders nachhaltig, die Dinge besonders nachhaltig zu tun. Dann haben die natürlich einen ganz anderen Hebel damit
B09	sustainable choice of products and vendors	We actually already worked on some sustainability projects also in our team. And it is mostly about finding the next best client and who is actually getting more and more sustainable. And I also actually looked into it a little. And the funny thing is it's not only about how sustainable companies are but also about the roadmap, how sustainable they want to be.
B10	sustainable choice of products and vendors	wir haben schon mal drüber geredet und so allgemeines Wissen glaube ich, dass es schon sehr sehr viel Energie braucht und die großen Transformer Modelle und so zu trainieren. Energie, Zeit, auch Geld. Wir haben Cloud System, wo wir schon geschaut haben, dass es zumindest lokal ist.
B10	sustainable choice of products and vendors	Welche ihrer Locations die am besten saubere Energie haben und das haben wir uns noch mal angeschaut und auch danach ausgesucht.
B10	sustainable choice of products and vendors	Also es muss natürlich/ am Ende des Tages müssen wir verdienen und damit muss es dann auch zusammenpassen. Aber es gibt auch einige Firmen, wo wir schon gesagt haben, mit dem wollen wir eigentlich nicht arbeiten, einfach weil der Fokus in die richtige Richtung gehen soll
B11	sustainable choice of products and vendors	Das wird wenn dann wahrscheinlich allgemein gemacht von Kunden über ihre gesamte Software, die sie betreiben. Das heißt, die, die werden dann vielleicht die Server bei einem Unternehmen einkaufen, was eher nachhaltig aufgestellt ist, also auch im CO_2 Ausstoß.
B11	sustainable choice of products and vendors	Ja und. Vielleicht wenn es könnte bei der bei der Server Auswahl oder der Provider Auswahl für Server. Da könnte es eine Rolle spielen was die für ein CO_2 Abdruck mitliefern.
B12	sustainable choice of products and vendors	er natürlich, wenn wir Cloud Anbieter auswählen, dann sind wir zum Teil darauf angewiesen, was der Anbieter für Ressourcen einsetzt. Nichtsdestotrotz glaube ich, dass wir da schon in geringem Maße Einfluss nehmen können, indem wir uns dann wieder entscheiden, der mehr auf nachhaltige Energieversorgung oder Produktion achtet als ein anderer.
B1	unintentional sustainable	Ja indirekt schon: wir haben Data Science Projekte im Rahmen der Material Entwicklung. Und dort ist Nachhaltigkeit auch an neue Innovationen und neue Materialien geknüpft
B1	unintentional sustainable	Aber ich kann mir natürlich viele Projekte vorstellen und ich kenne auch viele Projekte aus dem Data Science, aus Beispielen, also Use Cases, wo das natürlich absolut super überall eine Rolle spielt.
B1	unintentional sustainable	Das ist schon automatisch in deiner Ausbildung drin, dass du willst das auch so eine Art Pareto Prinzip - 80 20/ Du willst eigentlich mit dem 80 % Modell / Oder mit dem mit dem Modell (..)/ Das eben reicht. 80 % gut zum Beispiel /(...) Da sparst du dadurch eine Menge Trainingszeit und kriegst trotzdem einen großen Return damit hin
B1	unintentional sustainable	Ist der Aufwand, den man treibt, ist der in Relation zum Nutzen. Das wird sowieso immer geguckt
B2	unintentional sustainable	Und jetzt öffnet sich ja auch eigentlich erstmalig so ein Deckel, der uns immer zurückgehalten hat. Davor war das Thema Agieren im Data Science Umfeld sehr, sehr stark gekappt worden, weil wir wirklich im eigenen Rechenzentren nur Leistung zur Verfügung hatten.
B2	unintentional sustainable	Diese Thematik, die Idee, die haben wir jetzt mit der Cloud. Eben diese Flexibilität zur Verfügung. Vorher mussten wir das eigentlich ganz viel Hardware einkaufen, um diese Spitzenlast abzudecken, aber wir haben es jetzt noch nie aus der Nachhaltigkeitsperspektive so bewertet im Detail.
B2	unintentional sustainable	Daher will man den Zyklus für sich selbst schon optimieren und verkürzen. Und dann ist man auch geneigt, etwas zu optimieren oder den Algorithmus besser zu machen.
B3	unintentional sustainable	Wo wir aber durchaus sind, ist, dass wir so was wie die Grüne Flotte, also elektronisch betriebene Fahrzeuge, in den Vordergrund schieben können
B3	unintentional sustainable	Oder dass wir mit den Modellen dafür sorgen können, dass die zurückgelegte Strecke von Fahrzeugen reduziert wird.

B3	unintentional sustainable	Und insofern das Produkt oder der Service, den wir anbieten, der ist schon sehr in Richtung Nachhaltigkeit getrimmt. Worauf ich hinaus wollte, ist halt das Thema Wie kann ich das intern im Data Science bei der Umsetzung mit unterstützen? Und das können wir halt in dem Output, den wir liefern machen.
B3	unintentional sustainable	sondern es gibt den Best Practice Austausch mit dem Fokus auf Effizienz der Modelle, was am Ende auch zu Ressourcenschonung führen kann, aber halt nicht zwingend muss.
B3	unintentional sustainable	Allerdings zugegebenermaßen vor allen Dingen im Moment mit Fokus auf Kosteneinsparung. Aber meines Wissens nach geht das durchaus gut zusammen mit dem Thema Ressourcenschonung.
B3	unintentional sustainable	Solche Dinge nutzen wir aber vor dem Hintergrund der Kostenersparnis
B3	unintentional sustainable	Ich sagte ja zu Anfang schon uns ist zum Beispiel auch wichtig, dass wir near-realtime unterwegs sind und dabei nicht so viele Ressourcen verbrauchen. Also nicht so viel Kosten verbrauchen. Da kommen die Dinge automatisch zusammen.
B4	unintentional sustainable	Also wo ich jetzt drauf achten würde, wobei es natürliches Interesse ist, darauf zu optimieren.
B4	unintentional sustainable	Für mich als als Projektleiter im Maschine Learning Bereich ist das höchst relevant, wie schnell ein Modell rechnet und trainiert, weil ich teilweise auch einfach Kapazität und Zeit Grenzen habe, an denen ich arbeiten muss. Also sein Modell muss in einer gewissen Latenzen Ergebnis liefern.
B4	unintentional sustainable	ay, jetzt müssen wir eine Lösung finden, wie wir die 30 Länder in der Zeit laufen lassen, in der wir jetzt aktuell fünf laufen lassen.
B5	unintentional sustainable	So as I said, the amount of time we try to minimize it as much as possible, not invoke very complex algorithms at the beginning.
B6	unintentional sustainable	der Kernpunkte, wofür wir uns wie entscheiden, ist Wie schnell rechnet dieser Algorithmus? Weil wir denken, je schneller es es rechnet, desto eher können wir, kommen wir meistens ans Ziel.
B7	unintentional sustainable	Das heißt, wir haben jetzt im Moment bei Stufe eins sozusagen müssen wir 60 Geschäftspartner regulieren. Nennen wir das Bahnhofsbuchhandlung, gehen wir dann über 350 und bei Einzelhändlern reden wir dann über bis zu 30 40.000. Und da muss man natürlich sozusagen schon sich Gedanken machen, dass man das Ganze so konzipiert, dass es überhaupt möglich ist, hinterher 30.000 Einzelhändler zu regulieren, aber eher vor dem vor dem Hintergrund der Machbarkeit.
B8	unintentional sustainable	Also soweit ich das bisher gesehen habe, nie im Namen der Nachhaltigkeit, sondern wenn man immer so was macht, was man ja schon lange im Endeffekt macht im Namen der Erklärbarkeit eines Modells, ist es das einfache Modell, es ist besser erklärbar, was aber natürlich gleichzeitig auch bedeutet, dass es eben kein sehr komplexes großes Modell und damit natürlich auch nachhaltiger.
B8	unintentional sustainable	Wie gesagt, also vieles eben unbewusst irgendwie in die Richtung, dass es auch gleichzeitig nachhaltiger ist. Aber nicht im Namen der Nachhaltigkeit.
B09	unintentional sustainable	but we always go for the easiest and simplest solution that can solve the problem. So we don't start with like okay, we have a huge model that can solve any problem. We always start simple and then put on top of it. I think in that way we are very sustainable.
B09	unintentional sustainable	tarting with really simple solutions and then building on top of it and only if nothing works out, then we really go for complicated solutions. I think this is also coming from the fact that we are a bank.
B10	unintentional sustainable	Gerade die Tendenzen gehen ja auch in die Richtung, dass man viel wiederverwendet.
B10	unintentional sustainable	Du willst ja auch nichts Unnützes machen, oder? Wir werden immer ein ML-Modell trainieren, wenn ich denke, dass es sinnvoll ist.
B11	unintentional sustainable	Es ist im eigentlichen Interesse, dass die nicht nicht ewig rechnen die Maschinen, weil man einfach schneller fertig werden will. Aber nicht in dem Sinne okay, man muss, muss nachhaltiger werden.

A.3.3 *Hurdles to a more sustainable AI*

Table A.6: document statistics hurdles sustainability

	Category Name	Absolute Count	% of SUM	N of Documents	% of Documents
lack of transparency		*25*	*37*	*9*	*75*
	lack of understanding	25	37	9	75
no real sustainability		*7*	*10*	*5*	*41*
	rebound effect	7	10	5	41
complex development process		*14*	*20*	*6*	*50*
	missing processes	8	11	5	41
	nature of data science	6	8	4	33
Economic reasons		*21*	*31*	*7*	*58*
	motivation for effort not yet given	21	31	7	58

Table A.7: coded passages hurdles sustainability

Interviewee	Category Title	Marked Text
B7	rebound effect	Das heißt, wir haben jetzt im Moment bei Stufe eins sozusagen müssen wir 60 Geschäftspartner regulieren. Nennen wir das Bahnhofsbuchhandlung, gehen wir dann über 350 und bei Einzelhändlern reden wir dann über bis zu 30 40.000. Und da muss man natürlich sozusagen schon sich Gedanken machen, dass man das Ganze so konzipiert, dass es überhaupt möglich ist, hinterher 30.000 Einzelhändler zu regulieren, aber eher vor dem vor dem Hintergrund der Machbarkeit.
B7	rebound effect	.Ja, ja, auf alle Fälle ist es natürlich so, es steckt ein kleiner Rebound Effekt drin, was man einspart. Das nutzt man dann ja letztendlich dafür, um es öfter auszuführen.
B4	rebound effect	Okay, jetzt müssen wir eine Lösung finden, wie wir die 30 Länder in der Zeit laufen lassen, in der wir jetzt aktuell fünf laufen lassen.
B2	rebound effect	an muss halt immer nur sagen, so Effekte wie durch den Weg in die Cloud spare ich CO2 ein. Auf der anderen Seite ist es das ist wie beim Verbrennungsmotor, wir haben den Verbrennungsmotor über die Jahrhunderte, sozusagen Jahrzehnte optimiert. Aber auf der anderen Seite viel mehr Autos auf der Straße. Also dieser komplementäre Effekt ist immer da.
B10	rebound effect	Egal was am Ende ist, man ist effektiver in dem, was man macht. Ja heißt ja nicht, dass man weniger macht. Dann geht man einfach noch in andere Sachen auslagern. So läuft es ja meistens.
B10	rebound effect	Ja, also ich glaube, Forschungsabteilungen sind auf jeden Fall immer zeit- und geldfüllenend die Kassen. Deswegen ist es irgendwie schwierig.
B09	rebound effect	So to really unlock this whole potential of AI, we actually need more resources. And one way to do it is also to optimize resources. I think they need to go hand in hand. Otherwise probably the development will be slower.
B3	nature of data science	Aber wenn ich hingehe und was Neues modeliere, dann ist es schon sehr schwierig, im Vorhinein zu sagen okay, ich baue auf eine bestimmte Methode, auf ein bestimmtes Modell, was weniger Ressourcen verschwendet.
B10	nature of data science	Dann sagen wir, wir haben festgestellt, dass mehr Daten und mehr Zahlen auf jeden Fall helfen.
B10	nature of data science	lso Sie würden immer noch größere Modelle bauen gerne. Weil es auch immer noch mehr Daten gibt. Und weil sich das beides so hochschaukelt.
B1	nature of data science	wie lange hat das Training gedauert. Aber machst du dir vorher nicht so drüber Gedanken. Es ist auch schwer sowas vorher abzuschätzen.

B1	nature of data science	Na ja, so ist ja grundsätzlich so ein Vorgehen, aber wenn du von vornherein sagen können Sie hier genau das ist das Modell, diese Größe, diese Struktur, das bräuchte ich, da wäre jeder Data Scientist super happy, aber das gibt es noch nicht. Da wird gerade dann geforscht, wie man das vielleicht vorgeben kann, dass man das nicht mit Trial and Error rausfinden muss.
B09	nature of data science	When we actually create some kind of application that we need to run on premise, we need to actually spend a lot more resources on this than a cloud provider
B8	motivation for effort not yet given	Im Endeffekt trägt der Kunde dann die finale Entscheidungen
B8	motivation for effort not yet given	Derzeit habe ich manchmal das Gefühl, es es noch so, so ein Luxus, den sich Unternehmen dann irgendwie leisten. Da wollen sie natürlich das natürlich nur bedingt, dass sie da geben wir extra Geld für aus.
B8	motivation for effort not yet given	Und erfahrungsgemäß in der Beratung ist es natürlich, dass die Unternehmen haben natürlich auch nicht kein Interesse, die Beratung unendlich lange laufen zu lassen.
B7	motivation for effort not yet given	Also wenn die wenn eine längere Berechnung ein besseres Ergebnis produziert, könnte ich mir vorstellen, dass das die meisten Unternehmen eher das bessere Ergebnis bevorzugen würden als die kürzere Rechenzeit.
B3	motivation for effort not yet given	Und dabei meine ich, würde ich den Verbrauch an Rechenleistung - dieser Art Impact - haben wir uns bisher nicht angeguckt. Da gibt es halt mehr das Thema, dass wir uns angucken, was ist der Kostenfaktor da.
B3	motivation for effort not yet given	Also es gibt natürlich einen Best Practice Austausch, aber ehrlich gesagt nicht mit dem Fokus auf Nachhaltigkeit, sondern es gibt den Best Practice Austausch mit dem Fokus auf Effizienz der Modelle,
B3	motivation for effort not yet given	Aber man muss halt auf der anderen Seite auch sagen, wir haben jetzt nicht genügend Teams oder nicht genügend Personal an der Stelle, sodass wir zum Beispiel zwei verschiedene Ansätze für dasselbe Modell parallel entwickeln lassen können, dass man dann hingehen könnte und sagen kann okay, welches von beiden ist am Ende ressourcenschonender oder sogar nachhaltiger?
B3	motivation for effort not yet given	Denn wir werden nicht zu einem Punkt kommen, wo meine Mitarbeiter hingehen und für einen Lösungsvorschlag 10, 20, 30 Modelle durchprobieren, um am Ende zu gucken ,okay, welches davon ist das Nachhaltigste?
B2	motivation for effort not yet given	Ich glaube die "low hanging fruits" was Nachhaltigkeitsmanagement angeht [sind] an anderer Stelle im Konzern sind einfach viel viel größer.
B2	motivation for effort not yet given	Und da ist man vielleicht aus der IT noch etwas/ Da fliegt man noch ein bisschen unter dem Radar. Aktuell hab ich so die Wahrnehmung.
B2	motivation for effort not yet given	Ich finde, dass funktioniert nur dann glaube ich, wenn du tatsächlich schaffst, Nachhaltigkeit auch mit einem Kostenfaktor zu preisen und dann wirklich auch gegeneinander zu so ein Stück weit auszuspielen.
B2	motivation for effort not yet given	Also du musst so eine gewisse Masse erreichen, weil du mit dieser Masse dann auch die Awareness schaffst. Und das ist etwas, das beginnt jetzt bei uns. Bis jetzt ist es so, dass wir noch im überschaubaren Rahmen sind.
B2	motivation for effort not yet given	uf der anderen Seite ist es das ist wie beim Verbrennungsmotor, wir haben den Verbrennungsmotor über die Jahrhunderte, sozusagen Jahrzehnte optimiert. Aber auf der anderen Seite viel mehr Autos auf der Straße. Also dieser komplementäre Effekt ist immer da. Gibt es auch einen Fachbegriff? // Rebound Effekt // wie man den klar macht. Weil klar mit der Cloud habe ich eine viel höhere Effizienz. Gleichzeitig ist es für mich viel leichter an Ressourcen zu kommen und ich mache viel mehr und schnellere Projekte.
B12	motivation for effort not yet given	ass wir bis jetzt für mein Team gesprochen noch nicht den Schritt gemacht haben und gesagt haben Der Nachhaltigkeit zuliebe machen wir uns einfach viel mehr Aufwan
B12	motivation for effort not yet given	Es kostet einfach noch nichts, außer dass es vielleicht lange dauert.
B10	motivation for effort not yet given	Aber wir verkaufen es an Firmen quasi als. Es ist kein Ziel da oder kein Incentive was zu tun.

B1	motivation for effort not yet given	der [Fokus] ist auf Prozess, auf Produkt und auf Geschäftsmodell. Du siehst, da taucht der Begriff Sustainability erst mal gar nicht auf, der ist nur immer indirekt mit da drin.
B1	motivation for effort not yet given	Aber ich denke mal, die Auflagen werden da schon auch immer höher.
B1	motivation for effort not yet given	Aber diese Leuchttürme - diese Vorbilder - kommen eigentlich nur, wenn es die Kunden wirklich wollen oder wenn es irgendwie von oben, von der Unternehmensführung durch eine starke Überzeugung her getrieben wird. Ansonsten würde es nicht kommen. Ganz klar.
B1	motivation for effort not yet given	Aber wie gesagt, in der Wirtschaft zählt erst mal mehr Kosten.
B1	motivation for effort not yet given	Wie akkurat ist mein Modell?" Es wird erst dann darauf geguckt, wenn es einen signifikanten Kostenanteil verursacht.
B3	missing processes	Und wir sind noch nicht so weit, dass wir wie bei, sagen wir mal, klassischer Programmierung, wo es standardisierte Entwurfs Muster gibt, mit denen man bestimmte Aufgaben löst.
B2	missing processes	Wenn das kompliziert wird oder es Aufwand generiert, dann wird man schnell in alte Muster verfallen
B2	missing processes	Wenn man jetzt die ganzen Themen noch aufbürdet, dann wird es immer schwieriger für ihn. Und ich sage mir, darum sollte man sollte es versuchen und das glaube ich, wird dann auch machbar sein, das eigentlich weg zu kapseln, so gut es geht.
B12	missing processes	Es gibt bei uns Development Guidelines, die sind aber noch nicht Nachhaltigkeit optimiert.
B1	missing processes	Eine Art Life Cycle Analyse für das ganze Training. Ich denke mal, da muss ein Standard geschaffen werden.
B1	missing processes	Aber es ist noch nicht so quantifiziert (...) Und das liegt einfach daran, dass es keinen richtigen STANDARD dafür gibt.
B1	missing processes	okay, wenn ich meinen Energieverbrauch weiter und weiter runter schraube, wird irgendwann meine Accuracy mit der Targetprediction wird immer schlechter. "Oh, was mache ich denn dann? Wo ist denn da der Trade off?" Keine Ahnung (lacht). Deswegen muss man, wie gesagt wieder den Standard schaffen. Dann musst du sagen, wie viel Verlust in der Vorhersagegenauigkeit entspricht eigentlich später - in der Implementierung - eingesparte Energie zum Beispiel.
B09	missing processes	But this sustainability field, I think is still a bit young. It also needs to be established.
B8	lack of understanding	Wiel der Rest kommt, Green Coding und so weiter, wenn es mal im Bewusstsein der Leute ist.
B6	lack of understanding	Ich glaube, ich bin nicht so stark in nachhaltiger Optimierung bewandert, dass ich wüsste, welche Knöpfe ich drücke. Ich jetzt wie triggere um jetzt noch eine größere Nachhaltigkeit zu initiieren.
B6	lack of understanding	ielleicht ist es gar nicht so bewusst, wie viel das doch an Strom frisst, wie viel das doch an Zeit frisst. Ein Stück weit im Zuge dessen dann auch Kosten. Aber ich denke, wenn da eine gewisse Sensibilisierung stattfindet, gerade auch für Managementebene, die überhaupt keine Ahnung von Machine Learning haben, die eben auch sehr auch utopisch von Machine Learning denken,
B5	lack of understanding	. I would say just like at the first it's just introduce some kind of transparency. Where should we optimize? And I don't think most of the companies has this kind of view already at this point.
B4	lack of understanding	Ich kann mir vorstellen, das wird wichtiger und ich glaube, es wird ganz natürlich da passieren, wo ich irgendwie entweder eine offensichtliche oder eine Argumentation kreieren kann, warum der nachhaltige eigentlich auch der, der wertvollere, wie es ist.
B3	lack of understanding	dass man dazu kommt, dass wirklich beim Modellieren schon darüber nachgedacht wird. Okay, wie kann ich das so aufbauen, dass es am Ende ressourcenschonend ist und dass das, was ich aus dem Modell heraus bekomme, dass der Vorteil dort gegenüber dem Nachteil der verbrauchten Ressourcen auch wirklich überwiegt.

B3	lack of understanding	Was wir aber nicht haben ist der Impact, den die Data Science Modelle als solche haben.
B3	lack of understanding	Was wir allerdings nicht haben, ist eine Gegenüberstellung. Okay, was ist denn dann der positive Effekt, den so ein Modell hat?
B3	lack of understanding	Also wir werden dahin kommen müssen, dass verschiedene Ansätze miteinander verglichen werden müssen, so dass man so eine Art Benchmark bekommt.
B3	lack of understanding	Ich glaube, wenn wir da hinkommen wollen, müssen wir insgesamt das Thema CO2 Ausstoß viel transparenter machen. Also ich glaube im ersten Schritt wird das Ganze erst mal über den rein monetären Faktor gespielt werden.
B3	lack of understanding	. Aber solange es nicht Standard geworden ist, dass man für alle Bereiche die eine Firma abdeckt, so was wie den CO2 Ausstoß transparent macht, werden wir es bei dem gesamten Thema Data Science auch nicht sehen. Also der der Bereich ist so tief im Unternehmen drin, dass der häufig erst relativ spät gesehen wird. Und ich glaube man muss das CO2 Thema viel weiter nach vorne schieben, um dann halt auch in die einzelnen Gewerke innerhalb eines Unternehmens runter zu kommen und dann zum Beispiel auch für den gesamten Data Bereich eine CO2 Bilanz für jedes Unternehmen zu bekommen.
B2	lack of understanding	ich weiß gar nicht, ob und wie wir das generiert bekämen, aber zuerst muss du ja transparent machen. Und wenn du jetzt so was wie 1,2 nennst, wenn ich das jetzt irgendwo platzieren würde, wüsste ich auch, dass die das aufgreift.
B2	lack of understanding	Ich sag ja, Nachhaltigkeit muss vor allem mit der Transparenz starten.
B2	lack of understanding	Deshalb sage ich, wenn CO2 mehr Transparenz bekäme, also wirklich, dann glaube ich, wird das noch viel stärker.Transparenz schafft erst diesen Wandel, dieses Bewusstsein. Ja, und dann wird das noch stärker hineinfließen
B2	lack of understanding	Ja, da fühlt man sich fast schon so wie ein Umweltschützer. Allein weil dann steht wie viele Tonnen CO2 oder wie viele Kilos - sind es ja noch bei dem. Aber alleine dieses - ich sage ja - dieses transparent machen von einer Zahl ist viel wichtiger als /. Damit beginnt die Reise.
B2	lack of understanding	Ich sage, es wird sich ändern, wenn die Transparenz da ist und wenn die Transparenz von vielen Kleinen hoch wandert auf Managementebene, wo sie dann sehen, wie viel Kilo an CO2 haben wir jetzt verursacht, dann glaube ich, wird sich die Diskussion drehen
B2	lack of understanding	Also wie gesagt, die Transparenz, die steht für mich an erster Stelle und dann kommt der Rest automatisch.
B2	lack of understanding	dieses Thema Nachhaltigkeit politisch verkaufen auf dem Finanzmarkt transparent zu machen. Du musst den Leuten zeigen, dass Du auf einem Trend bist, der nach unten folgt, dass du deine Ziele erreichen kannst wenn du das dann fortschreitest. Und deshalb gehört es dazu, das auch festzuhalten im Sinne von Vorher-Nachher-Effekt.
B12	lack of understanding	aber zumindest nutzen wir die Möglichkeit, dass wir uns da ein Bild verschaffen und dass wir da ein Bewusstsein herstellen.
B10	lack of understanding	Aber was äquivalent in Energie ist, das wissen die meisten gar nicht. Wie lange man davon seine Wohnung heizen kann, wäre vielleicht ein guter Vergleichswert.
B1	lack of understanding	Wo vielleicht der Haken ein bisschen immer dahinter steht. Die Leute wissen nicht, wo sie anfangen sollen. Also wie sie eigentlich die besten Use Cases identifizieren? Wo sind eigentlich die größten Sustainability Potenziale zu heben?
B1	lack of understanding	Wo fängt man denn an? Wo ist denn das größte Potenzial zu heben? Wo soll man eigentlich anfangen? Daran scheitert es häufig ein bisschen.
B1	lack of understanding	Ich finde diese das ist ja hochtechnisch und auch nicht unbedingt einfach messbar, welches Modell mehr Energie verbraucht oder weniger. Denn das wird ganz schnell wiederum komplex.

B1	lack of understanding	Also generell (...) erstmal muss man es messbar machen. Du musst es überhaupt erst mal tracken. Du musst eine Benchmark schaffen, dass du sagen kannst "Hier für diesen Typ von Modell auf dieser Hardware für ein Training, das so und so lange dauert, mit so und so viel Parameter, wird so und so viel Energie verbraucht"
B1	lack of understanding	Das ist ja nur "Energy Invested" Anteil. Aber sie hat nicht den "Energy Return"-Anteil gesehen. Du musst beide Seiten der Medaille IMMER betrachten.

A.3.4 *Necessary steps for a sustainable AI*

Table A.8: document statistics necessary steps

	Category Name	Absolute Count	% of SUM	N of Documents	% of Documents
transparency		*22*	*24*	*10*	*83*
	measure and track emissions	22	24	10	83
relief of the developer		*30*	*32*	*9*	*75*
	Behavior control	17	18	9	75
	dedicated person for sustainability	4	4	2	16
	guidenance of employees	9	9	4	33
Balancing of sustainability		*15*	*16*	*5*	*41*
	Balancing sustainability	15	16	5	41
Knowledge generation		*24*	*26*	*11*	*91*
	create and share knowledge	24	26	11	91

Table A.9: coded passages necessary steps

Interviewee	Category Title	Marked Text
B1	Balancing sustainability	Deswegen muss man, wie gesagt wieder den Standard schaffen. Dann musst du sagen, wie viel Verlust in der Vorhersagegenauigkeit entspricht eigentlich später - in der Implementierung - eingesparte Energie zum Beispiel.
B2	Balancing sustainability	Ich finde, dass funktioniert nur dann glaube ich, wenn du tatsächlich schaffst, Nachhaltigkeit auch mit einem Kostenfaktor zu preisen und dann wirklich auch gegeneinander zu so ein Stück weit auszuspielen.
B3	Balancing sustainability	Am Ende geht es aus meiner Sicht immer um die Incentivierung des Einzelnen. Und wir kommen, glaube ich, auch da immer wieder zu dem Punkt, dass es um Kosten und Nutzen geht.
B3	Balancing sustainability	Also ich denke, das Allerwichtigste ist, dass man sich erst einmal sehr genau Gedanken darüber macht, an welchen Stellen gerade so was wie künstliche Intelligenz überhaupt Anwendung finden sollte und an welchen nicht. Dass ich immer noch die die größte He,rausforderung.
B4	Balancing sustainability	Ich kann mir vorstellen, das wird wichtiger und ich glaube, es wird ganz natürlich da passieren, wo ich irgendwie entweder eine offensichtliche oder eine Argumentation kreieren kann, warum der nachhaltige eigentlich auch der, der wertvollere, wie es ist.
B4	Balancing sustainability	Und wo du natürlich Data Science benutzen kannst, um ein Attributierungsmodell zu entwickeln.
B4	Balancing sustainability	Wenn ich bemessen kann, wie sich ein bestimmter Faktor jetzt finanziell auswirkt, dann hilft das sicherlich. Leute zu überzeugen, da rein zu investieren. Macht es immer einfacher, diese Gespräche zu führen und auch solche entscheidungen zu bewegen.
B4	Balancing sustainability	Ich sehe es so ein bisschen wie ein zweischneidiges Schwert, weil es sozusagen auch zu einem gewissen Grad konstruiert ist.
B4	Balancing sustainability	Ich glaube, wenn ich den Trade-off habe, ich kriege jetzt irgendwie 2 % mehr circa aber dafür rechnet das eine Stunde mehr. Ja, dass man auf den Einzelfall gucken
B4	Balancing sustainability	Also ich sage, ich habe halt ein Modell, das rechnet eine Stunde und wenn ich da jetzt viel Hirnschmalz reinsteckt, dann kriege ich das auf fünf Minuten runter.
B4	Balancing sustainability	Ich glaube, da kommt es wahrscheinlich eher und schneller in den Bereich, wo du ganz, ganz konkrete Kostenvorteile siehst.
B4	Balancing sustainability	Also wenn ich jetzt mein Modell von einer Stunde auf zehn Minuten runter bringe und dafür aber zwei Wochen Entwicklungsressource aufbringen muss, dann ist das vielleicht kein guter Trade für mich. Also Unternehme
B4	Balancing sustainability	Es wird sehr interessant sein zu sagen Wo sind die Effizienzgewinne, die ich habe, wenn ich so ein Modell effektiv einsetze? Gegen was sind die Kosten für die. Für den Betrieb davon.
B8	Balancing sustainability	muss man das neuronale Netz noch verzehnfachen für 0,05 % Accuracy oder so?
B8	Balancing sustainability	kein Interesse, die Beratung unendlich lange laufen zu lassen. Ja, aber dass man natürlich sagt hier, aber das sind dann gerade noch so Sachen, die ihr dann mitbekommt wie eben die Optimierung, also günstigere, natürlich geringere Rechenzeit und damit auch nachhaltig, nachhaltiger.
B2	Behavior control	dann muss es leicht funktionieren. Wenn das kompliziert wird oder es Aufwand generiert, dann wird man schnell in alte Muster verfallen. Also wenn, dann muss das halt verknüpft sein mit Automatisierung. Mit Leichtigkeit. Ja dann wirst du so was nutzen und davon profitieren
B2	Behavior control	Also Automatisierung und gerade dafür steht ja die IT, die kann, die könnte das dann ja auch. Dass man dann wirklich da hingeht und sagt, man setzt eine Automatisierung auf, die in der Lage ist, nach Nachhaltigkeitsaspekten praktisch die Ausführung von einem Workload so zu steuern und zu koordinieren, dass das nach den Kriterien am besten erfolgt.

B2	Behavior control	Ja, und wie gesagt, das muss man wegnehmen von den Leuten, weil der Data Scientist soll sich ja um sein Kernproblem kümmern, er soll das Business Problem lösen.
B3	Behavior control	die vorhandenen Optimierungsmöglichkeiten nutzen.
B3	Behavior control	Also den besten Hebel haben sicherlich die Cloudbetreiber
B4	Behavior control	ganz anderen Hebel, aber das sehe ich tatsächlich bei den Cloud Betreibern
B4	Behavior control	wo ich ein bestimmter Teil meiner Flotte oder meine Flotte darf nur noch durchschnittlich so viel Emissionen erzeugen. Also wenn ich so einen Anreiz hätte, extern oder ich als Gesamtunternehmen zeigen muss, wie viel einfach Ressourcen ich verbrauche, dann wäre das natürlich natürlich ein interessanter Faktor.
B5	Behavior control	But it just needs something like a badge, something that you see whenever you go to seletc them. For how good is this on an energy level ranking,CO2 level, something like that. So on that perspective, only transparency would definitely help.
B6	Behavior control	. Aber wenn man merkt, dass es ganz, ganz wenig Schritte braucht, um vieles zu einzusparen, dann würde man einfach diesen Schritt natürlich gehen können.
B8	Behavior control	Also ich persönlich denke schon, dass man das wenigstens so semi automatisieren kann, wenigstens in dem Sinne von Systemen oder so, dass du, auch wenn du durch einen Code gehst, das scannst.
B8	Behavior control	Git hat ja ein Haufen, so dass wenn du einen Code commitest im Repository, dass da ein Haufen so Pre-Checks, die dann durchlaufen. Die dann Redundanzen quasi im Endeffekt einfach erkennen.
B8	Behavior control	Ob das aber irgendwelche Institutionen geben wird, die sagen, wie Green Coding, wird es unser Ding. Und wir bieten es an für Software Developer so weiter, um die Leute zu schulen
B09	Behavior control	For me, the cloud first strategy and every company should actually have this cloud strategy because this is the future.
B10	Behavior control	Sie sagen es ja, wenn es etwas gibt, wie man es einfacher macht oder wie mehr nachhaltig machen es die Leute dann nicht. Und dann ist meistens die Antwort weil es halt unkomfortabel ist und schlecht zu benutzen.
B10	Behavior control	Oder so ein Kubernetes-Flag wo es gerade am nachhaltigsten ist
B12	Behavior control	Ja, ganz ehrlich, es müsste mit einem Pricetag oder mit gewissen Hürden verbunden sein besonders energieintensive Algorithmen, Programmiersprachen, Abläufe zu nutzen.
B12	Behavior control	Und ich denke erst wenn man spürt, auf irgendeine Weise angezeigt bekommt, wie gesagt, irgendwelche Eintrittsbarrieren zu überwinden hat bei besonders energieintensiven Sprachenalgorithmen. Dann wird wirklich was passieren.
B1	create and share knowledge	Du willst ja eigentlich mit dem kleinsten möglichen Modell [arbeiten], was genau groß genug ist, um alle Informationen, die du hast, in den Daten zu lernen - in ihrer Komplexität, wie sie da sind. Das heißt, das ist eigentlich auch ein aktives Forschungsgebiet. Es ist im Vornherein ja nicht klar, wie groß dieses Modell sein muss, weil du noch gar nicht weißt, welche Information in deinen Daten nicht drinstecken.
B1	create and share knowledge	Du musst eine Benchmark schaffen, dass du sagen kannst "Hier für diesen Typ von Modell auf dieser Hardware für ein Training, das so und so lange dauert, mit so und so viel Parameter, wird so und so viel Energie verbraucht"
B2	create and share knowledge	ich weiß gar nicht, ob und wie wir das generiert bekämen, aber zuerst muss du ja transparent machen. Und wenn du jetzt so was wie 1,2 nennst, wenn ich das jetzt irgendwo platzieren würde, wüsste ich auch, dass die das aufgreift.
B3	create and share knowledge	Das wird noch einiger Forschung bedürfen, damit wir dahin kommen langfristig.
B3	create and share knowledge	Also wir werden dahin kommen müssen, dass verschiedene Ansätze miteinander verglichen werden müssen, so dass man so eine Art Benchmark bekommt. (
B3	create and share knowledge	Das heißt, was wir eigentlich brauchen, ist wirklich so eine Art Benchmark im Vergleich dessen okay, was, was funktioniert gut und was funktioniert weniger gut? Klassische Forschung sicherlich, die man in anderen Bereichen auch hat, so dass man da Grundlagen hat, auf denen man aufbauen kann, sodass man das Rad nicht in jedem Unternehmen immer wieder selber neu erfinden muss.
B3	create and share knowledge	Und vor diesem Hintergrund ist es ja auch dann wirtschaftlich und nicht nur idealistisch angezeigt, auch mit Universitäten zusammenzuarbeiten und da gemeinsam in die Forschung zu gehen, um dann für beide Seiten positiv am Ende heraus zu bekommen. Insofern ja, kann ich mir sehr gut vorstellen.
B3	create and share knowledge	Wir werden das noch viel tiefer erforschen müssen, dieses gesamte Gebiet, um diese Expertise zu haben, die mir dann auch erlaubt zu sagen, okay, ich möchte jetzt meine Umsetzung auf Nachhaltigkeit auslegen. Ich habe auch das entsprechende Hintergrundwissen dazu, um das überhaupt zu ermöglichen. Ich glaube, da brauchen wir noch weitere Forschung.
B3	create and share knowledge	u Anfang werden wir natürlich etwas sehen, wo einzelne Unternehmen erst mal selber davon profitieren wollen, wenn sie in der Richtung etwas erforscht haben. Aber auf lange Sicht wird das sicherlich Gemeingut werden. Aber das wird halt wirklich noch deutlich länger dauern.
B3	create and share knowledge	Wie kann man das Thema schon mal sinnvoll angehen und wir werden da sicherlich noch in Zukunft weiter lernen müssen
B4	create and share knowledge	Es wird sehr interessant sein zu sagen Wo sind die Effizienzgewinne, die ich habe, wenn ich so ein Modell effektiv einsetze? Gegen was sind die Kosten für die. Für den Betrieb davon. (
B5	create and share knowledge	Transparency would definitely help on the cloud level, like giving some kind of ranking of the algorithm and implementation with their corresponding memory footprint, resource footprint, CO2 footprint, energy consumption.

B6	create and share knowledge	Was, wenn es coole Best Practices gibt, die dann auch veröffentlicht werden und quasi Einsicht haben? Dann herzlich gern.
B6	create and share knowledge	Und ich glaube, eine Kombination von Best Practice und stetiger Erinnerung würde, würde, würde zumindest initiieren, dass man daran daran ein Bewusstsein entwickelt.
B7	create and share knowledge	Transparenz. Oder vielleicht kann man es auch anders sagen, vielleicht kann man es Kompetenz sagen. Ich glaube, wir brauchen einfach auch. Mehr Experten, also mehr Leute, die sich wirklich gut auskennen in dem Bereich, die gut ausgebildet werden.
B7	create and share knowledge	Also ich glaube, dass das Bildung in dem Bereich auch eine ganz große Rolle spielt und dass da auch das Thema Nachhaltigkeit und solche Aspekte vermittelt werden müssen.
B8	create and share knowledge	e, inwieweit wird sich quasi das unterscheiden, dass man sagt nachhaltiger Code ist gleich Runtime optimaler Code, oder? Also da bin ich mir selber auch noch nicht sicher, ob es da irgendwo, also ob es da überhaupt Unterschiede geben wird, quasi im Endeffekt
B8	create and share knowledge	Wiel der Rest kommt, Green Coding und so weiter, wenn es mal im Bewusstsein der Leute ist
B8	create and share knowledge	h glaube also, die Data Scientist, wenn man sich da ein bisschen aktiv mit beschäftigt mit den aktuellen Architekturen, weiß man ja ungefähr, wie groß die sind, also wie diese Sprachmodelle, die dann noch zwei, 3 % haben, aber dafür eine Größenordnung mehr Parameter haben.
B09	create and share knowledge	. I think one problem is also maybe in different departments people are working on similar applications. So transparency is lacking.
B10	create and share knowledge	Aber dann so redundante Arbeit oder so in der Verantwortung, dass man schaut, dass man Sachen nicht doppelt macht oder aber wirklich gut miteinander teilt, sodass nicht da jeder so sein eigenes Süppchen braut.
B10	create and share knowledge	Also ich glaube, das ist eigentlich das größte Problem, zum Beispiel in der Akademie oder in der Forschung, weil niemand negative Ergebnisse teilen, die man teilt
B10	create and share knowledge	Also so in diese Richtung macht es schon Sinn, dass es quasi Best Practices wird, dass man sein seine Ergebnisse Open Source stellt.
B12	create and share knowledge	Und vielleicht braucht es dazu auch noch mehr Best Practice Sharing.
B2	dedicated person for sustainability	Ich glaube, das ist so die erste bedeutsame Verantwortung, dass wir da wirklich einen Zentral Bereich haben, der das bündeln wird
B2	dedicated person for sustainability	Ja, ich glaube schon, dass die aus diesen Nachhaltigkeitsprogrammen dann raus purzeln.
B2	dedicated person for sustainability	Also ich will es jetzt nicht den Data Scientist überlassen, diese Antwort zu finden, sondern ja
B6	dedicated person for sustainability	In diesem Zusammenhang dachte ich immer, es wäre echt cool, wenn es eine dedizierte Person im Unternehmen gibt, die eben genau so was macht. Also wirklich alles in Form von kostentechnisch - jetzt nicht Verwaltung und auch nicht Buchhaltung - sondern sich bewusst über die Nachhaltigkeit befasst und Kosten quantifiziert, also in Form von wirklich Abteilung zu Abteilung mal rein sichtet und guckt
B3	guidenance of employees	Ich glaube aber, dass zu diesem Thema nachhaltig Data Science machen, langfristig gehören muss, dass man dazu kommt, dass wirklich beim Modellieren schon darüber nachgedacht wird.
B6	guidenance of employees	Oder denkst du, dass sollte auch definitiv von Unternehmen gesteuert sein, dass diese Maßnahmen dann auch geteilt werden und auch in gewisser Art und Weise dann auch durchgesetzt? (18:28) B6 Ja am liebsten das letztere, weil dadurch hast du auch die Gewissheit, dass das quasi komplett im kompletten Unternehmen auch aufgenommen wird.
B6	guidenance of employees	Ähm, in erster Linie das Bewusstsein. Ich finde, es ist ein extrem neues und junges Thema und hoffentlich gewinnt es auch mehr an Aufwind aufwärts.
B6	guidenance of employees	Und ähm, ganz konkrete Best Practices würden uns sehr helfen
B6	guidenance of employees	h Management Unterstützung letztendlich würde dazu auch führen? (26:47) B6 So was. Ja.
B8	guidenance of employees	Ob das aber irgendwelche Institutionen geben wird, die sagen, wie Green Coding, wird es unser Ding. Und wir bieten es an für Software Developer so weiter, um die Leute zu schulen.
B8	guidenance of employees	Ja, ich denke. Mitarbeiter schulen in gewissen Sinne, dass man sagt "Was ist die Relevanz davon?"
B8	guidenance of employees	Dass man sagt, Nachhaltigkeit ist ein Teil der Beratung.
B12	guidenance of employees	Es gibt bei uns Development Guidelines, die sind aber noch nicht Nachhaltigkeit optimiert. Auch da sind wir gerade am Anfang und im Moment sind es noch die Individuen, die davon wissen oder nicht.
B1	measure and track emissions	Du musst es überhaupt erst mal tracken.
B2	measure and track emissions	Ich sag ja, Nachhaltigkeit muss vor allem mit der Transparenz starten. Ich glaube, dann beginnt es ja auch zu wirken.
B2	measure and track emissions	Deshalb sage ich, wenn CO2 mehr Transparenz bekäme, also wirklich, dann glaube ich, wird das noch viel stärker.Transparenz schafft erst diesen Wandel, dieses Bewusstsein.
B2	measure and track emissions	Allein weil dann steht wie viele Tonnen CO2 oder wie viele Kilos - sind es ja noch bei dem. Aber alleine dieses - ich sage ja - dieses transparent machen von einer Zahl ist viel wichtiger als /. Damit beginnt die Reise. (
B2	measure and track emissions	Ja, erst erst, nachdem du es transparent gemacht hast und du kannst es auch eins zu eins vielleicht einem Verantwortungsbereich zuordnen.
B2	measure and track emissions	ber wenn du das feingranularer zerlegen kannst oder dann auch wirklich Teilaspekte beziffern könntest. Also wie gesagt, die Transparenz, die steht für mich an erster Stelle und dann kommt der Rest automatisch.

B3	measure and track emissions	Was wir allerdings nicht haben, ist eine Gegenüberstellung. Okay, was ist denn dann der positive Effekt, den so ein Modell hat? Also wenn ich jetzt zum Beispiel hingehe und sage, okay, ich versuche jetzt über ein Modell verstärkt die Wegstrecken, die unsere Fahrer zurücklegen, zu reduzieren, dann habe ich im Moment keine Gegenüberstellung.
B3	measure and track emissions	Ich glaube, wenn wir da hinkommen wollen, müssen wir insgesamt das Thema CO_2 Ausstoß viel transparenter machen. Also ich glaube im ersten Schritt wird das Ganze erst mal über den rein monetären Faktor gespielt werden. Aber wie schon mehrfach gesagt, die beiden hängen halt bei Data Science einfach super eng zusammen.
B3	measure and track emissions	Und ich glaube man muss das CO_2 Thema viel weiter nach vorne schieben, um dann halt auch in die einzelnen Gewerke innerhalb eines Unternehmens runter zu kommen und dann zum Beispiel auch für den gesamten Data Bereich eine CO_2 Bilanz für jedes Unternehmen zu bekommen.
B4	measure and track emissions	Und ich glaube, dass das wahrscheinlich der Startpunkt eine bewusste Entscheidung dafür ist, zu sagen "wir wollen messen, wie nachhaltig wir sind und wie unser Handeln sich darauf auswirkt.".
B5	measure and track emissions	Transparency would definitely help on the cloud level, like giving some kind of ranking of the algorithm and implementation with their corresponding memory footprint, resource footprint, CO_2 footprint, energy consumption.
B6	measure and track emissions	aber ich denke die meisten meisten können eher was mit dieser CO_2 Emissionen auch anfangen. Besonders wenn du das dann in Relation setzt.
B6	measure and track emissions	stetiger Erinnerung
B8	measure and track emissions	Oder so eine Art Energiekostenvoranschlag
B10	measure and track emissions	Also ich glaube schon, dass es wichtig ist, weil es ja. Vieles ist einfach so weg abstrahiert, dass man es gar nicht so versteht. Ja.
B10	measure and track emissions	Ja eigentlich da erst mal zu zeigen, dass es da ein Problem gibt.
B10	measure and track emissions	Du, verbrauchst jetzt so viel. Du könntest nur so viel verbrauchen mit diesen Lösungsansätzen.
B11	measure and track emissions	Ja. Transparenz hilft immer. Glaube ich das? Das könnte ich mir schon vorstellen. Ja. Und wenn man das noch mit einem attraktiven Angebot verbindet, dann. Dann auf jeden Fall
B12	measure and track emissions	Da sind wir dran, dass wir unseren Anwendern bewusst machen, wenn wir verschiedene Algorithmen zur Verfügung stellen, dann gibt es ja deutliche Unterschiede in der Performance.
B12	measure and track emissions	Wir haben uns tatsächlich Strom Messgeräte besorgt und haben verschiedene Teams Einstellungen ausprobiert. ,das ist unser Kommunikations oder Tool. Wir haben unser selbst geschriebenes Nano Production Programm unter verschiedenen Einstellungen getestet oder vermessen.
B12	measure and track emissions	Ja so wie man ohne Probleme im Task Manager sehen kann wie viel CPU Auslastung gerade bei welchem Prozess liegt. Dass man zu jedem Prozess auch angezeigt bekommt, wie viel CO_2 da aktuell ausgestoßen wird oder in den letzten zehn 20 Stunden oder so was. Oder auf dem Smartphone sieht man ja auch sofort, wie weit es mein Akku runter.

A.3.5 *Future of sustainable AI*

Table A.10: document statistics outlook

	Category Name	Absolute Count	% of SUM	N of Documents	% of Documents
AI vs. climate change		*12*	*34*	*6*	*50*
	AI for advanced sustainability	10	28	4	33
	AI no solution for everything	2	5	2	16
Development of topic in companies		*23*	*65*	*11*	*91*
	Importance of sustainability of AI is raising	14	40	6	50
	no widespread of the topic in the next years	9	25	7	58

Table A.11: coded passages outlook

Interviewee	Category Title	Marked Text
B1	AI for advanced sustainability	Ich bin davon überzeugt, für die Zukunft, weil Sustainability ist ein komplexes (.) Problem mit ganz, ganz vielen Einflussfaktoren, die sich gegenseitig beeinflussen
B1	AI for advanced sustainability	Natürlich, für ein KI-Modell / Sagen wir mal so, da ist ein großer strategischer Anteil. Ich glaube, da müssen Länder zusammenkommen und sagen "Okay, wenn wir wirklich als Menschheit, als Civilization, als Society, sustainable werden wollen"
B1	AI for advanced sustainability	Aber das musst du ja dann voll planen, musst die richtigen Energiespeicher haben und und und. Die Kosten, die das kostet, das alles aufzubauen an Energie, das musst du alles mit einkalkulieren. Und dazu braucht es, glaube ich, KI. Da wird KI eine große Rolle spielen. Nicht nur in den einzelnen Use-Cases, sondern auch in der Strategie
B1	AI for advanced sustainability	wie manage ich das dann alles in Realtime optimal. Alles das musst du mit einer KI machen. Das wird kein Mensch machen. Da sehe ich den riesen riesen Benefit der KI. Also große Skala -Big Scale.
B2	AI for advanced sustainability	Dass man dann wirklich da hingeht und sagt, man setzt eine Automatisierung auf, die in der Lage ist, nach Nachhaltigkeitsaspekten praktisch die Ausführung von einem Workload so zu steuern und zu koordinieren, dass das nach den Kriterien am besten erfolgt.
B2	AI for advanced sustainability	Also ich sage AI, als wirklich als Anwendungsfall zur Einsparung von Nachhaltigkeit. Bei uns ist es zum Beispiel Autonome Verfahren oder Optimierung von Rollwegen
B2	AI for advanced sustainability	KI wird eher bei uns dort Nachhaltigkeit fördern, indem es wirklich Betriebsabläufe optimiert, autonome Systeme schafft, die effizienter, effektiver laufen und damit wirklich Nachhaltigkeit generieren.
B2	AI for advanced sustainability	Ich glaube, gerade bei der Werkstoffherstellung erwarte ich mir da große Durchbrüche, dass man tatsächlich in den Produktionsketten/Lieferketten, dass das durch KI auch auch vieles optimiert werden kann
B8	AI for advanced sustainability	Nein, das glaube ich geht nicht ohne. Die Systeme sind einfach viel zu kompliziert. Also ich ganz persönlich bin der Meinung, wir Menschen sind nicht mehr in der Lage, viele dieser- wenn ich mir vorstelle, was weiß ich, ich soll als einzelner Mensch den Hamburger Hafen koordinieren oder so? Das ist einfach, das sind zu viele Variablen, das kriegen wir im Kopf nicht mehr hin. Das müssen wir über Computer machen. Und analytische Lösungen da sind die Computer nicht schnell genug. Insofern müssen wir das immer so approximative Systeme machen wie KIs oder so.
B09	AI for advanced sustainability	, I think at least in the near future we will probably have lots of companies already producing some applications regarding the sustainability reports with NLP technology.
B3	AI no solution for everything	Aber ich habe ganz häufig gerade bei diesem Thema Künstliche Intelligenz den Eindruck, gerade in den klassischen Medien wird das so als Allheilmittel gesehen und es werden da Wunderdinge erwartet. Und ich finde, man muss halt schon sehr klar sehen, dass Künstliche Intelligenz eigentlich weniger mit Intelligenz zu tun hat, als vielmehr damit zu tun hat, klassische, ich nenne es mal Programmier oder Entwicklungsaufgaben auf anderem Wege zu lösen. Wo ich hin will, ist : Künstliche Intelligenz wird nur in den seltensten Fällen von sich aus neue Lösungswege finden. Das heißt, da werden wir als die Nutzer der künstlichen Intelligenz oder als die Entwickler der künstlichen Intelligenz immer noch den Lösungsweg finden müssen.
B11	AI no solution for everything	h würde mir mehr natürliche Intelligenz als künstliche Intelligenz wünschen, ehrlich gesagt. Und dass das gilt für viele bereiche und für mich ist das eher ein Bewusstseinswandel der natürlichen Intelligenz als der künstlichen Intelligenz und künstliche Intelligenz ist nur irgendwie ein Hilfsmittel dabei, das umzusetzen
B1	Importance of sustainability of AI is raising	Das kann durchaus so kommen. Ich kann mir durchaus vorstellen, dass wir hier in Deutschland da irgendwie Vorreiter werden oder in anderen Ländern
B2	Importance of sustainability of AI is raising	Aber man geht davon aus, dass wenn Sie intelligente Projekte steuern wollen, wenn Sie intelligent Maßnahmen identifizieren wollen, wenn Sie vielleicht so was wie Treiber, also Nachhaltigkeitstreiber identifizieren wollen, dass Sie dann eben auch auf diesen Punkt des Data Science stoßen

B2	Importance of sustainability of AI is raising	Wie gesagt, ich gehe jetzt auch davon aus, dass das bei uns durch dieses Literacy Programm viel, viel stärker in die Breite gerät. Und dann kommst du auch an die Punkte, wo diese kritischen Massen so langsam entstehen, wo dann Cluster, also Computer Ressourcen oder Compute Ressourcen allokiert werden in einem größeren Maß, wo dann Kosten sich subsumieren, die dann wirklich so langsam überall auch so zeigen "Oh, das nimmt jetzt größere, größere Effekte an". Und dann wird dieses Thema auch kommen. Dann glaube ich, dann wird man sich tatsächlich um dieses effizientes Rechnen/ langes Rechnen viel mehr bewusster machen.
B2	Importance of sustainability of AI is raising	Und wahrscheinlich werden wir dann auch so ein bisschen überwachen und gucken, dass sie das dann effizient und so tun werden. Das wird so ein bisschen stärker. Vielleicht kann ich mir gut vorstellen, unsere Rolle dann auch einnehmen, als Enabler befähigen wir sie im ersten Schritt. Aber das Thema Governance ist so ein Thema, das werden wir uns stärker vornehmen und dann kommen solche Aspekte.
B2	Importance of sustainability of AI is raising	Aber ich glaube, du brauchst so eine kritische Masse. Also du musst so eine gewisse Masse erreichen, weil du mit dieser Masse dann auch die Awareness schaffst.
B2	Importance of sustainability of AI is raising	Das ist ja alles Effekte von Nachhaltigkeit und ich glaube, das wird sich überall durchtragen. KI wird sowieso so bedeutsam werden wie Strom. Ja, idealerweise wird man es auch so wie Strom nicht merken.
B3	Importance of sustainability of AI is raising	Man hat es, finde ich, mit der Datenschutzgrundverordnung in Europa wunderbar gesehen, was erst sehr negativ angekommen ist und mittlerweile sogar eigentlich zu einem Standortvorteil geworden ist. Und ich kann mir gut vorstellen, dass man mit ähnlichen Initiativen auch gerade dieses Thema Nachhaltigkeit in der Datenverarbeitung gut voranbringen kann.
B3	Importance of sustainability of AI is raising	Wenn ich mir im Moment angucke, was Energie und Ressourcenverbrauch angeht, ist natürlich dieser Bereich Künstliche Intelligenz zunehmend daran beteiligt. Sicherlich kann man das mit anderen Industriezweigen vergleicht immer noch relativ wenig. Aber dadurch, dass das gerade so einen großen Hype erfährt, haben wir da schon ein sehr großes Potenzial in die positive wie auch in die negative Richtung.
B6	Importance of sustainability of AI is raising	Also mein Gefühl ist eher, dass aufgrund der Corona Pandemie auch viele Kosten ein Stück weit eingespart werden mussten. Und ich denke im Zuge dessen wird man doch ein bisschen genauer hinschauen, wie Rechenkapazität in welcher Art auch immer verwendet werden. Und ich denke auch, dass das stimmt, wird es irgendwann auch cooles Tracking Tools geben oder Software, die eben genau das quantifizieren, also wirklich alles vom Strom bis hin zur Lebenserwartung von gewissen Kabeln die in diesen GPU Servern liegen bis hin zum zum zum Endprodukt
B7	Importance of sustainability of AI is raising	Also erst mal finde ich ja, die Tatsache, dass ich jetzt zu diesem Thema gerade interviewt werde, zeigt doch zeigt doch das, dass man sich offensichtlich anfängt, sich intensiver mit dem Thema zu beschäftigen. Es passt ja auch voll in den Zeitgeist
B7	Importance of sustainability of AI is raising	Wie können wir es irgendwie optimieren, weil es ja wahrscheinlich jetzt Frau von der Leyen gesagt ist, man of the Moon Projekt für Europa, aber auch für die ganze Welt ist. Insofern ich glaube, es wird auf jeden Fall eine größere Bedeutung bekommen in der Zukunft
B7	Importance of sustainability of AI is raising	Du meinst quasi KI Anwendung, die dann positive nachhaltige Effekte erzielen? Genau. Ja, das ist natürlich. Das ist natürlich ein Riesenwachstums Bereich. Weil wir ja wie gesagt, als Menschheit darauf angewiesen sind, nachhaltiger einfach zu leben.
B7	Importance of sustainability of AI is raising	kann jetzt nur aus meiner Erfahrung sprechen, was ich mir auf irgendwelchen Konferenzen oder so gesehen habe, dass Unternehmen wie [anonymisier]beispielsweise sich schon sehr intensiv in dem Bereich engagieren, da schon viele Daten, Produkte oder Produkte entwickelt haben, die Sie, die sie da in Produktionsbereich einsetzen, dass sie da sogar eigene Firmen für gegründet haben, um dieses Thema sozusagen dediziert anzugehen.
B09	Importance of sustainability of AI is raising	Definitely there will be lots of projects, use cases coming from sustainability into this field. But this sustainability field, I think is still a bit young. It also needs to be established. But once it is established, there will be lots of projects coming, but also from AI perspective,
B3	no widespread of the topic in the next years	aber ich glaube. Im Moment noch nicht daran, dass wir das in absehbarer Zeit als Standard in allen Unternehmen haben werden.
B4	no widespread of the topic in the next years	ch sehe Nachhaltigkeit nicht als ein sehr prominentes Thema. Also in den Foren, in denen ich unterwegs bin, ist das sehr ein totales Nischenthema, würde ich sagen
B5	no widespread of the topic in the next years	don't know if any other branches, but in B2B, we are talking still about digitalization and we are still on the baby steps of the digitalization.
B5	no widespread of the topic in the next years	I mean in B2B organizations here, it's not the case honestly. So I don't see that within at least the next five years. I don't see this to be a topic. Maybe from that's my perspective, maybe I'm just like I don't see it.
B7	no widespread of the topic in the next years	wie gesagt, bei uns im Unternehmen ist das noch sehr weit weg, weil wir aber auch einfach allgemein noch relativ weit weg sind davon, überhaupt solche Anwendungen, welche sehr energieintensiven sind zu entwickeln. Wahrscheinlich ist es dann am Ende des Tages sind so die ersten, die sich die sich wirklich damit. Wahrscheinlich auch heute schon beschäftigen, sind natürlich die Googles, Amazons, Facebooks der Welt.
B10	no widespread of the topic in the next years	Aber in Filmen ist es momentan noch nicht so wichtig. Also die Filme selber würden das niemals tun. Die machen es nur, wenn sie der Kunden von ihnen wollen. Für alle, die direkt an Kunden verkaufen, also an Privatpersonen, die beschäftigen sich damit. Aber wir verkaufen es an Firmen quasi als. Es ist kein Ziel da oder kein Incentive was zu tun.
B11	no widespread of the topic in the next years	nd ja, da sind Banken und Versicherungen was was wir beraten noch total weit vorne weg und von daher ist das überhaupt erst mal ein Prozess dahin zu kommen was machen wir überhaupt mit KI und was machen wir mit Data Science? Und der nächste Schritt ist dann eigentlich erst mal Was müssen wir denn da alles beachten dabei? Und das ist unter anderem dann Nachhaltigkeit.
B11	no widespread of the topic in the next years	as wir auch täglich machen, da machen wir uns auch wenig Gedanken darum, dass das WhatsApp vielleicht dann doch nicht immer so gut ist, wenn man nur noch chattet. Und genauso ist es eigentlich auch zurzeit bei den bei den Banken und Versicherungen oder bei den Kunden. Und das wird sich auch auf Sicht glaube ich wenig wenig ändern.

B12	no widespread of the topic in the next years	Ich hoffe, dass es mehr ist als ein Trend. Und ich bin mir bewusst, dass das kein einfacher Weg sein wird. Wenn wir wirklich nachhaltiger KI machen wollen.

B

MODEL RANKING MATERIAL

The code of the execution is available at
https://github.com/LuWiesa/ranking_classification_models

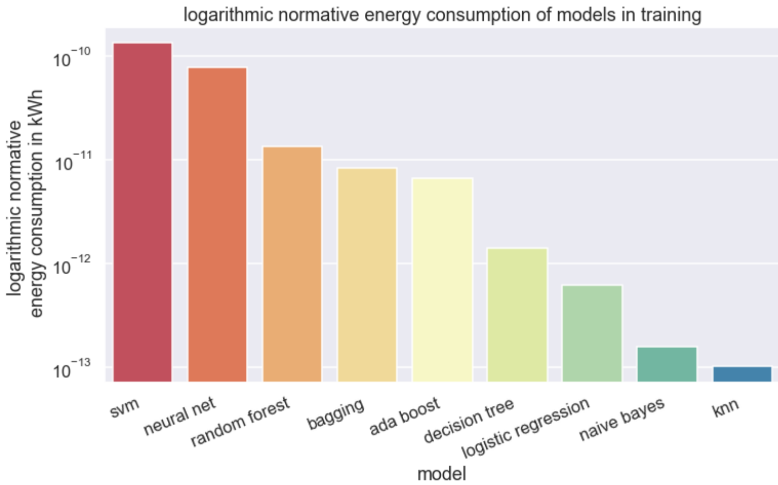

Figure B.1: logarithmic mean normative energy consumption of models in training

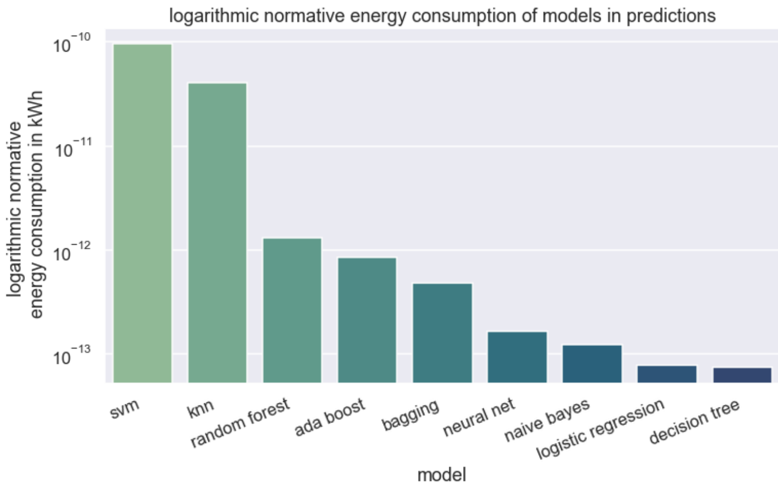

Figure B.2: logarithmic mean normative energy consumption of models in prediction

[Abd+15] Sarah Abdulsalam, Ziliang Zong, Qijun Gu, and Meikang Qiu. "Using the Greenup, Powerup, and Speedup metrics to evaluate software energy efficiency." In: *2015 Sixth International Green and Sustainable Computing Conference (IGSC)*. 2015, pp. 1–8. DOI: 10.1109/IGCC.2015.7393699.

[AENEN20] Tasneem Alsaati, Samir El-Nakla, and Darin El-Nakla. "Level of Sustainability Awareness among University Students in the Eastern Province of Saudi Arabia." In: *Sustainability* 12.8 (2020). ISSN: 2071-1050. DOI: 10.3390/su12083159. URL: https://www.mdpi.com/2071-1050/12/8/3159.

[ACM18] Sulaiman Alsheibani, Yen Cheung, and Chris Messom. "Artificial Intelligence Adoption: AI-readiness at Firm-Level." In: *22nd Pacific Asia Conference on Information Systems, PACIS 2018, Yokohama, Japan, June 26-30, 2018*. Ed. by Masaaki Hirano, Michael D. Myers, Kyoichi Kijima, Motonari Tanabu, and Dai Senoo. 2018, p. 37. URL: https://aisel.aisnet.org/pacis2018/37.

[ACM19] Sulaiman Alsheibani, Yen Cheung, and Chris Messom. "Towards An Artificial Intelligence Maturity Model: From Science Fiction To Business Facts." In: *23rd Pacific Asia Conference on Information Systems, PACIS 2019, X'ian, China, July 8-12, 2019*. Ed. by Kwok Kee Wei, Wei Wayne Huang, Jae Kyu Lee, Dongming Xu, James J. Jiang, and Hee-Woong Kim. 2019, p. 46. URL: https://aisel.aisnet.org/pacis2019/46.

[Amo+19] Dario Amodei, Danny Hernandez, Girish Sastry, Jack Clark, and Ilya Sutskever. *ai and compute*. 2019.

[AKS20a] Lasse F. Wolff Anthony, Benjamin Kanding, and Raghavendra Selvan. "Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models." In: *CoRR* abs/2007.03051 (2020). arXiv: 2007.03051. URL: https://arxiv.org/abs/2007.03051.

[AKS20b] Lasse F. Wolff Anthony, Benjamin Kanding, and Raghavendra Selvan. "Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models." In: *CoRR* abs/2007.03051 (2020). arXiv: 2007.03051. URL: https://arxiv.org/abs/2007.03051.

[ALV14] Nima Asadi, Jimmy Lin, and Arjen P. de Vries. "Runtime Optimizations for Tree-Based Machine Learning Models." In: *IEEE Transactions on Knowledge and Data Engineering* 26.9 (2014), pp. 2281–2292. DOI: 10.1109/TKDE.2013.73.

[Bar19] Jeff Bar. *AWS News Blog*. 2019. URL: https://aws.amazon.com/de/blogs/aws/amazon-ec2-update-inf1-instances-with-aws-inferentia-chips-for-high-performance-cost-effective-inferencing.

[BAK22] Firas Bayram, Bestoun S. Ahmed, and Andreas Kassler. *From Concept Drift to Model Degradation: An Overview on Performance-Aware Drift Detectors*. 2022. DOI: 10.48550/ARXIV.2203.11070. URL: https://arxiv.org/abs/2203.11070.

[Ber+16] Luca Bertinetto, Joao Henriques, Jack Valmadre, Philip Torr, and Andrea Vedaldi. "Learning feed-forward one-shot learners." In: (June 2016).

[Bro+20] Tom B. Brown et al. "Language Models are Few-Shot Learners." In: *CoRR* abs/2005.14165 (2020). arXiv: 2005.14165. URL: https://arxiv.org/abs/2005.14165.

[CS16] David Casey and Sebastian Sieber. "Employees, sustainability and motivation: Increasing employee engagement by addressing sustainability and corporate social responsibility." In: *Research in Hospitality Management* 6 (Jan. 2016), pp. 69–76. DOI: 10.2989/RHM.2016.6.1.9.1297.

[Che+22] Zhenghua Chen, Min Wu, Alvin Chan, Xiaoli Li, and Yew-Soon Ong. *A Survey on AI Sustainability: Emerging Trends on Learning Algorithms and Research Challenges*. 2022. DOI: 10.48550/ARXIV.2205.03824. URL: https://arxiv.org/abs/2205.03824.

[Chi+21] Mihai Cristian Chirodea, Ovidiu Constantin Novac, Cornelia Mihaela Novac, Nicu Bizon, Mihai Oproescu, and Cornelia Emilia Gordan. "Comparison of Tensorflow and PyTorch in Convolutional Neural Network - based Applications." In: (2021), pp. 1–6.

[CCS12] Adele Cutler, D. Richard Cutler, and John R. Stevens. "Random Forests." In: *Ensemble Machine Learning: Methods and Applications*. Ed. by Cha Zhang and Yunqian Ma. Boston, MA: Springer US, 2012, pp. 157–175. ISBN: 978-1-4419-9326-7. DOI: 10.1007/978-1-4419-9326-7_5. URL: https://doi.org/10.1007/978-1-4419-9326-7_5.

[DJ15] Qi Deng and Shaobo Ji. "Organizational Green IT Adoption: Concept and Evidence." In: *Sustainability* 7.12 (2015), pp. 1–19. URL: https://EconPapers.repec.org/RePEc:gam:jsusta:v:7:y:2015:i:12:p:15843-16755:d:60842.

[Dev+19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423. URL: https://aclanthology.org/N19-1423.

[Dre16] Thorsten Dresing. *Praxisbuch Interview, Transkription und Analyse - Anleitungen und Regelsysteme für qualitativ Forschende*. 2016. URL: http://nbn-resolving.de/urn:nbn:de:101:1-2015100817897.

[Dun15] Stephan Dunkel. *Eine Anforderungsanalyse für ein mobiles Patienteninformationssystem unter Berücksichtigung aktuellen Datenschutzes*. 2015.

[Ell11] Steve Elliot. "Transdisciplinary Perspectives on Environmental Sustainability: A Resource Base and Framework for IT-Enabled Business Transformation." In: *MIS Quarterly* 35.1 (2011), pp. 197–236. ISSN: 02767783. URL: http://www.jstor.org/stable/23043495 (visited on 05/25/2022).

[FKH18] Stefan Falkner, Aaron Klein, and Frank Hutter. "BOHB: Robust and Efficient Hyperparameter Optimization at Scale." In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, 2018, pp. 1437–1446. URL: https://proceedings.mlr.press/v80/falkner18a.html.

[FD+14] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. "Do we need hundreds of classifiers to solve real world classification problems?" In: *The journal of machine learning research* 15.1 (2014), pp. 3133–3181.

[Fre+22] Nathan C. Frey, Baolin Li, Joseph McDonald, Dan Zhao, Michael Jones, David Bestor, Devesh Tiwari, Vijay Gadepally, and Siddharth Samsi. *Benchmarking Resource Usage for Efficient Distributed Deep Learning*. 2022. DOI: 10.48550/ARXIV.2201.12423. URL: https://arxiv.org/abs/2201.12423.

[Gup21] Abhishek Gupta. "The current state of affairs and a roadmap for effective carbon-accounting tooling in AI." In: (2021). URL: https://devblogs.microsoft.com/sustainable-software/the-current-state-of-affairs-and-a-roadmap-for-effective-carbon-accounting-tooling-in-ai/.

[Hel11] Cornelia Helfferich. *Die Qualität qualitativer Daten : Manual für die Durchführung qualitativer Interviews*. 2011. URL: https://doi.org/10.1007/978-3-531-92076-4.

[Hen+20] Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. "Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning." In: *Journal of Machine Learning Research* 21.248 (2020), pp. 1–43. URL: http://jmlr.org/papers/v21/20-312.html.

[HB20] Danny Hernandez and Tom B. Brown. "Measuring the Algorithmic Efficiency of Neural Networks." In: *CoRR* abs/2005.04305 (2020). arXiv: 2005.04305. URL: https://arxiv.org/abs/2005.04305.

[HVD15] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. *Distilling the Knowledge in a Neural Network*. 2015. DOI: 10.48550/ARXIV.1503.02531. URL: https://arxiv.org/abs/1503.02531.

[Ins21] Capgemini Research Institute. "Sustainable it." In: (2021).

[Jac+17] Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew G. Howard, Hartwig Adam, and Dmitry Kalenichenko. "Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference." In: *CoRR* abs/1712.05877 (2017). arXiv: 1712.05877. URL: http://arxiv.org/abs/1712.05877.

[JS19] Aled James and Daniel Schien. *A low carbon kubernetes scheduler*. English. Other. Jan. 2019.

[Jol15] Antonio Jolov. *Nachhaltigkeitsinnovation durch Nutzerintegration in KMU*. 2015.

[JHW22] Wandri Jooste, Rejwanul Haque, and Andy Way. "Knowledge Distillation: A Method for Making Neural Machine Translation More Efficient." In: *Information* 13.2 (2022). ISSN: 2078-2489. DOI: 10.3390/info13020088. URL: https://www.mdpi.com/2078-2489/13/2/88.

[Kai14] Robert Kaiser. *Qualitative Experteninterviews : Konzeptionelle Grundlagen und praktische Durchführung*. 2014. URL: https://doi.org/10.1007/978-3-658-02479-6.

[KM20] Saif M. Khan and Alexander Mann. "AI Chips: What They Are and Why They Matter." In: (2020). DOI: https://doi.org/10.51593/20190014. URL: cset.georgetown.edu/research/ai-chips-what-they-are-and-why-they-matter/.

[Kle+17] Aaron Klein, Stefan Falkner, Simon Bartels, Philipp Hennig, and Frank Hutter. "Fast Bayesian Optimization of Machine Learning Hyperparameters on Large Datasets." In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. Ed. by Aarti Singh and Jerry Zhu. Vol. 54. Proceedings of Machine Learning Research. PMLR, 2017, pp. 528–536. URL: https://proceedings.mlr.press/v54/klein17a.html.

[Lac+19a] Alexandre Lacoste, Alexandra Sasha Luccioni, Victor Schmidt, and Thomas Dandres. "Quantifying the Carbon Emissions of Machine Learning." In: *ArXiv* abs/1910.09700 (2019).

[Lac+19b] Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. "Quantifying the Carbon Emissions of Machine Learning." In: *CoRR* abs/1910.09700 (2019). arXiv: 1910.09700. URL: http://arxiv.org/abs/1910.09700.

[LNH09] Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. "Learning to detect unseen object classes by between-class attribute transfer." In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 951–958. DOI: 10.1109/CVPR.2009.5206594.

[LMS19] Thomas Lauf, Michael Memmler, and Sven Schneider. "Emissionsbilanz erneuerbarer Energieträger 2018." In: *Climate Change* 37 (Nov. 2019), p. 158.

[LL19] Petro Liashchynskyi and Pavlo Liashchynskyi. "Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS." In: *CoRR* abs/1912.06059 (2019). arXiv: 1912.06059. URL: http://arxiv.org/abs/1912.06059.

[Lig+21] Anne-Laure Ligozat, Julien Lefèvre, Aurélie Bugeau, and Jacques Combaz. "Unraveling the hidden environmental impacts of AI solutions for environment." In: *CoRR* abs/2110.11822 (2021). arXiv: 2110.11822. URL: https://arxiv.org/abs/2110.11822.

[May14] Philipp Mayring. *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. 2014. URL: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173.

[McD+22] Joseph McDonald, Baolin Li, Nathan C Frey, Devesh Tiwari, Vijay N. Gadepally, and Siddharth Samsi. "Great Power, Great Responsibility: Recommendations for Reducing Energy for Training Language Models." In: *ArXiv* abs/2205.09646 (2022).

[MS20] Swaroop Mishra and Bhavdeep Singh Sachdeva. "Do We Need to Create Big Datasets to Learn a Task?" In: *Proceedings of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing*. Online: Association for Computational Linguistics, Nov. 2020, pp. 169–173. DOI: 10.18653/v1/2020.sustainlp-1.23. URL: https://aclanthology.org/2020.sustainlp-1.23.

[MWH18] Felix Mohr, Marcel Wever, and Eyke Hüllermeier. "ML-Plan: Automated machine learning via hierarchical planning." In: *Machine Learning* 107 (2018), pp. 1495–1515.

[Moh+21] Felix Mohr, Marcel Wever, Alexander Tornede, and Eyke Hüllermeier. "Predicting Machine Learning Pipeline Runtimes in the Context of Automated Machine Learning." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.9 (2021), pp. 3055–3066. DOI: 10.1109/TPAMI.2021.3056950.

[Muh15] Ravil Muhamedyev. "Machine learning methods: An overview." In: *Computer modelling & new technologies* 19.6 (2015), pp. 14–29.

[Nas17] Vladimir Nasteski. "An overview of the supervised machine learning methods." In: *Horizons. b* 4 (2017), pp. 51–62.

[Par+17] Aniruddha Parvat, Jai Chavan, Siddhesh Kadam, Souradeep Dev, and Vidhi Pathak. "A survey of deep-learning frameworks." In: *2017 International Conference on Inventive Systems and Control (ICISC)*. 2017, pp. 1–7. DOI: 10.1109/ICISC.2017.8068684.

[Pat+22] David Patterson, Joseph Gonzalez, Urs Hölzle, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. *The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink*. 2022. DOI: 10.48550/ARXIV.2204.05149. URL: https://arxiv.org/abs/2204.05149.

[Per+17] Rui Pereira, Marco Couto, Francisco Ribeiro, Rui Rua, Jácome Cunha, João Paulo Fernandes, and João de Sousa Saraiva. "Energy efficiency across programming languages: how do energy, time, and memory relate." In: *Proceedings of the 10th ACM SIGPLAN International Conference on Software Language Engineering* (2017).

[Pet00] Johann Petrak. "Fast Subsampling Performance Estimates for Classification Algorithm Selection." In: *Meta-learning: Building Automatic Advice Strategies for Model Selection and Method Combination* (May 2000).

[Rad+21] Alec Radford et al. "Learning Transferable Visual Models From Natural Language Supervision." In: *CoRR* abs/2103.00020 (2021). arXiv: 2103.00020. URL: https://arxiv.org/abs/2103.00020.

[SS21] Timo Schick and Hinrich Schütze. "It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners." In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, June 2021, pp. 2339–2352. DOI: 10.18653/v1/2021.naacl-main.185. URL: https://aclanthology.org/2021.naacl-main.185.

[Sch+19] Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. "Green AI." In: *CoRR* abs/1907.10597 (2019). arXiv: 1907.10597. URL: http://arxiv.org/abs/1907.10597.

[Sha+21] Omar Shaikh, Jon Saad-Falcon, Austin P Wright, Nilaksh Das, Scott Freitas, Omar Asensio, and Duen Horng Chau. "EnergyVis: Interactively Tracking and Exploring Energy Consumption for ML Models." In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI EA '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380959. DOI: 10.1145/3411763.3451780. URL: https://doi.org/10.1145/3411763.3451780.

[SV22] Dimitar Sht. Sherionov and Eva Vanmassenhove. "The Ecological Footprint of Neural Machine Translation Systems." In: *CoRR* abs/2202.02170 (2022). arXiv: 2202.02170. URL: https://arxiv.org/abs/2202.02170.

[SH19] Chance Simmons and Mark Holliday. "A comparison of two popular machine learning frameworks." In: vol. 35. Oct. 2019, pp. 20–25.

[Sin+21] Nikhil Singh, Brandon Kates, Jeff Mentch, Anant Kharkar, Madeleine Udell, and Iddo Drori. *Privileged Zero-Shot AutoML*. 2021. arXiv: 2106.13743 [cs.LG].

[So+21] David R. So, Wojciech Manke, Hanxiao Liu, Zihang Dai, Noam Shazeer, and Quoc V. Le. "Primer: Searching for Efficient Transformers for Language Modeling." In: *CoRR* abs/2109.08668 (2021). arXiv: 2109.08668. URL: https://arxiv.org/abs/2109.08668.

[SU21] Andreas Streim and Merle Uhl. *Künstliche Intelligenz kommt in Unternehmen Allmählich Voran*. 2021. URL: https://www.bitkom.org/Presse/Presseinformation/Kuenstliche-Intelligenz-kommt-in-Unternehmen-allmaehlich-voran.

[SGM19] Emma Strubell, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP." In: *CoRR* abs/1906.02243 (2019). arXiv: 1906.02243. URL: http://arxiv.org/abs/1906.02243.

[Tam+20] Thierry Tambe, Coleman Hooper, Lillian Pentecost, En-Yu Yang, Marco Donato, Victor Sanh, Alexander M. Rush, David Brooks, and Gu-Yeon Wei. "EdgeBERT: Optimizing On-Chip Inference for Multi-Task NLP." In: *CoRR* abs/2011.14203 (2020). arXiv: 2011.14203. URL: https://arxiv.org/abs/2011.14203.

[Tha+20] Urmish Thakker, Jesse Beu, Dibakar Gope, Ganesh Dasika, and Matthew Mattina. *Rank and run-time aware compression of NLP Applications*. 2020. DOI: 10.48550/ARXIV.2010.03193. URL: https://arxiv.org/abs/2010.03193.

[Tor+21] Tanja Tornede, Alexander Tornede, Jonas Hanselle, Marcel Wever, Felix Mohr, and Eyke Hüllermeier. "Towards Green Automated Machine Learning: Status Quo and Future Directions." In: *CoRR* abs/2111.05850 (2021). arXiv: 2111.05850. URL: https://arxiv.org/abs/2111.05850.

[Van18] Joaquin Vanschoren. "Meta-Learning: A Survey." In: *CoRR* abs/1810.03548 (2018). arXiv: 1810.03548. URL: http://arxiv.org/abs/1810.03548.

[Wan+19] Yaqing Wang, Quanming Yao, James Kwok, and Lionel M. Ni. *Generalizing from a Few Examples: A Survey on Few-Shot Learning*. 2019. DOI: 10.48550/ARXIV.1904.05046. URL: https://arxiv.org/abs/1904.05046.

[Wat+20] Cody Watson, Nathan Cooper, David Nader Palacio, Kevin Moran, and Denys Poshyvanyk. *A Systematic Literature Review on the Use of Deep Learning in Software Engineering Research*. 2020. DOI: 10.48550/ARXIV.2009.06520. URL: https://arxiv.org/abs/2009.06520.

[Wie+21] Philipp Wiesner, Ilja Behnke, Dominik Scheinert, Kain Kordian Gontarska, and Lauritz Thamsen. "Let's Wait Awhile: How Temporal Workload Shifting Can Reduce Carbon Emissions in the Cloud." In: *CoRR* abs/2110.13234 (2021). arXiv: 2110.13234. URL: https://arxiv.org/abs/2110.13234.

[Xu+21] Jingjing Xu, Wangchunshu Zhou, Zhiyi Fu, Hao Zhou, and Lei Li. *A Survey on Green Deep Learning*. 2021. DOI: 10.48550/ARXIV.2111.05193. URL: https://arxiv.org/abs/2111.05193.