Band
3

Rüdiger Buchkremer / Oliver Koch / Andreas Lischka (Hrsg.)

Implicit Decision Voting Made by Humans as Normative and Implementable Rules with the Help of Language Models

~

Patrick Hedfeld

ifid Schriftenreihe
Beiträge zu IT-Management & Digitalisierung

FOM Hochschule

ifid

Institut für IT-Management & Digitalisierung
der FOM University of Applied Sciences

Patrick Hedfeld

Implicit Decision Voting Made by Humans as Normative and Implementable Rules with the Help of Language Models

ifid Schriftenreihe der FOM, Band 3
Beiträge zu IT-Management & Digitalisierung

Essen 2025

ISBN (Print) 978-3-89275-394-0 ISSN (Print) 2699-562X
ISBN (eBook) 978-3-89275-395-7 ISSN (eBook) 2699-5638

Dieses Werk wird herausgegeben vom ifid Institut für IT-Management & Digitalisierung
der FOM Hochschule für Oekonomie & Management gGmbH

Verlag:
MA Akademie Verlags- und Druck-Gesellschaft mbH, Leimkugelstraße 6, 45141 Essen
info@mav-verlag.de

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie;
detaillierte bibliographische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Rüdiger Buchkremer / Oliver Koch / Andreas Lischka (Hrsg.)

Implicit Decision Voting Made by Humans as Normative and Implementable Rules with the Help of Language Models

Patrick Hedfeld

Correspondence:

E-Mail: patrick.hedfeld@gmx.de

Foreword

Artificial intelligence (AI) holds extraordinary potential to transform our world in profoundly positive ways. Whether in healthcare, education, or other critical domains, it provides innovative tools to tackle some of humanity's most pressing challenges. I am deeply convinced that AI can and should be a force for good, but this vision hinges on one critical factor: the integration of ethics into every stage of its design and application. Ethics cannot be an afterthought; it is the bedrock upon which trust, fairness, and progress in AI systems must be built.

This work delves into the fascinating interplay between implicit moral decision-making and AI. Specifically, it investigates how collective human decision votes can inform the development of ethical AI systems. By analyzing moral decision data, this research highlights the potential for AI to function as an implicit moral advisor—a system that honors human agency while fostering solutions that benefit all stakeholders. Through the application of generative language models, the study demonstrates how implicit moral preferences can be made transparent, shaped into normative principles, and positioned within broader societal dialogues.

One of the most striking insights from this research is that even imperfect moral data, often shaped by human biases, can guide AI toward consensus-based ethical rules that advance societal well-being. By addressing critical challenges, such as algorithmic bias and the need for accountability, this work offers a framework for designing AI systems that function as ethical collaborators—partners in achieving shared human values, rather than neutral or purely utilitarian tools.

As a professor and director of an AI institute, I view this work as an important contribution to an urgent conversation. It inspires us to develop AI systems that are not only technically innovative but also deeply aligned with humanity's moral aspirations. Together, we can create AI technologies that empower individuals, strengthen communities, and act as a guiding force for ethical progress in our shared future.

Essen, February 2025

Prof. Dr. Rüdiger Buchkremer

Research Director of the FOM Institute for IT-Management and Digitalization (ifid)

Abstract

Can an ethically justifiable decision-making process be facilitated through a machine moral agent in the form of advice? The premise revolves around decision votes explicitly solicited from human individuals and based on scenarios such as the Trolley Problem, reflected in data and processed through generative language models. These advisories can then be formulated in a general manner and discussed within a societal context, emerging implicitly from individual decisions. Furthermore, we discuss the concept of an implicit moral agent and an honorable AI advisor.

Keywords: AI, Generative Language Models, Decision Votes, Trolley Problem, Implicit Agent

Table of Contents

List of Figures

1 Introduction

In contrast to racist texts, moral data often contain implicit decision voting made by humans. This can be demonstrated by general language models, which leverage brain-inspired and generative approaches. The results produced by these models have the potential to be accepted by humans as normative and implementable rules. Language models, acting as implicit moral advisors, utilize moral data derived from win-lose decisions to address issues. By making the results of implicit voting visible, they pave the way for win-win decisions and situations.

However, it is important to engage in thorough discussions because moral data created by humans encompass various factors, including feelings, erroneous decisions, and egoism. Despite this complexity, a theoretical implicit moral advisor has the capacity to lead to normative and implementable rules through careful consideration and discussion.

2 Problem and Motivation

2.1 Implicit Rules for Ethics

Individual surveys provide implicit starting points for consensus-capable and implementable social ethics by offering insights into the collective preferences and moral judgments of a diverse range of individuals. For example, the survey conducted at Moral Machine, which involved millions of participants, solicited responses regarding their decisions on ethical dilemmas such as the trolley problem (cf. Bonnefon et al., 2016). These surveys reveal common patterns, preferences, and ethical considerations across diverse populations, providing valuable data for shaping ethical norms and principles that are widely accepted and implementable within society. The trolley problem was discussed early on (cf. Welzel, 1951) and later rose to prominence with the help of *Philippa Foot* (cf. Foot, 1978). She discussed the trolley problem, known as the principle of the double effect, where individuals must make a decision regarding the allocation of harm. On one side of the track, there are people (and possibly animals), while on the other side, there are also individuals. This dilemma presents a win-lose situation in decision-making. However, the advantage lies in the fact that decisions, including those from the moral machines survey, are based on numerous individual responses. This survey accumulated a total of 40 million answers from various countries worldwide up to the year 2018 (cf. Awad, 2018b).

In a pure win-lose situation, among other things, there is a lack of human interpretation of decisions or the possibility of shaping our coexistence. Based on *Homann*'s ideas, business ethics can be discussed at three levels, first the level of concrete judgements (Urteile), then the level of the so-called middle principles *Mittlere Grundsätze* and finally the top level of normative principles *Normative Prinzipien. "*Es beginnt auf der Stufe konkreter Urteile, dann folgt die Stufe mittlerer Grundsätze, bevor schließlich die Stufe der obersten Grundsätze, der normativen Prinzipien, diskutiert wird." (Homann, 2014: 214). Homann's ideas are grounded in individual-centric principles, emphasizing that only Pareto superior rule enhancements and win-win scenarios are deemed acceptable and feasible for implementation. "Wir [Homann and Lütge] sind der Auffassung, dass eine öffentliche Diskussion und eine Verständigung über ein fruchtbares Denkschema einen entscheidenden Beitrag zur Zukunftsfähigkeit der Marktwirtschaft und zu ihrer moralischen Qualität wird leisten können. Nur paretor-superiore Regelverbesserung, also Win-win Situationen, sind zustimmungsfähig und implementierbar" (cf. Homann, 2013: 64). They are about questioning the prisoner dilemma situation: "Dies [Argument] zeigt, dass im Umgang der Menschen miteinander die

Gefangenendilemma-Struktur immer stärker intuitiv wahrgenommen wird. Die Frage: Warum soll ich moralisch sein, ist daher durchweg so zu verstehen: Warum soll ich moralisch sein, wenn ich dann von anderen ausgebeutet werde?“ (Homann, 2014: 216) and are accepted in consensus by all individuals in society also to be an implementable rule or a universalization. “Nur solche Normen, Regeln gelten als moralisch gerechtfertigt, die universalisierbar sind.“ (Homann 2013: 83). Only those norms and rules that can be universalized are considered morally justified. At the end of a win-win rule should be human acceptance.

If one seeks to derive actions directly from principles such as freedom or justice, they may be labeled as fundamentalist. Conversely, those who believe that actions should be determined solely by circumstances without any normative orientation are subject to the ideology of pure practical constraints. Acting morally must integrate both sides: actions and principles. “Moralisches Handeln hat immer beides [Handlungen und Prinzipien] zu integrieren.“ (Homann, 2020: 9). The discussion might parallel the acceptance of democracy or capitalism, even if one experiences personal disadvantages in a specific situation. Despite these drawbacks, individuals may still endorse the overall systems of democracy or capitalism.

The concept proposed in this paper is to develop a moral data-driven consultant that incorporates moral data and facilitates win-win situations in controlled human-machine interactions. It is important to acknowledge that within the theory of human-machine interaction, various types of morally data-driven advisors are conceivable (cf. Misselhorn, 2018: 70-74; cf. Riek / Howard, 2014).

Level one represents an ethical impact agent, which influences humans through its use, such as a simple watch or clock. The moral aspects are solely dependent on how the object is utilized. *Level two* entails the implicit ethical agent, where the method of construction holds moral implications, as seen in a warning system, for instance. *Level three* is the explicit moral agent, capable of recognizing moral behavior, processing it, and making decisions accordingly. Finally, *level four* encompasses the full ethical agent, which possesses consciousness and other advanced abilities.

The concept proposed in this work is an implicit moral advisor, corresponding to *level two* according to *Moor* (Moor, 2016). This advisor has the potential to manifest as a human interface and process moral decision data in a manner that, while respecting human freedom, leads to a consensus-driven win-win situation. It is essential to recognize the distinction between being in a concrete dilemma situation and being outside of it, in terms of time or thought, when making decisions.

Here reference should be made to the *Regelbefolgungsmodell* (choice within the rules) and the *Regeletablierungsmodell* (choice of rules) by *James Buchanan* (cf. Homann, 2013: 61; cf. Buchanan, 1984). For this reason, we will look at what *moral data* means and when an algorithm or a statistical evaluation can be *racist* at all.

2.2 Racist Algorithms, the Importance and the Moral Side of Data

The idea of finding a moral algorithm that can make decisions is among other things a topic of machine ethics (cf. Bartneck et al., 2019: 34) and one of the greatest challenges for this discipline is the question: which is the correct moral theory? Only two examples should be mentioned here for different approaches that already exist: After *Gips* (cf. Gips, 1994), robots or artificial intelligence can have different implementations, among others, the idea of utilitarianism (cf. Bentham, 2004; Cf. Ramge, 2008) is discussed here as a form of consequentialist theory.

One approach involves making evaluations and assigning weights to different factors, ultimately leading to decisions based on a hierarchy of priorities. This hierarchy determines who is allowed to live and who must die, based on a simple "more or less" comparison. In the moral machines survey, a similar hierarchy was observed (cf. Bonnefon et al., 2018a). Participants were presented with scenarios akin to the trolley problem, involving various characters such as women, men, cats, dogs, homeless individuals, athletes, and criminals, among others.

Statistical analysis of the survey results revealed certain trends: participants tended to prioritize protecting young or unborn life, women were preferred over men in most categories, and in some cases, animals were deemed more valuable than certain individuals. For instance, in a pure utilitarian approach, children might be prioritized over adults, but intriguingly, the survey results indicated that the value of women diminishes with age, as elderly men were preferred over elderly women. Additionally, even though criminals are human beings, they were often deemed less valuable than animals, such as dogs.

If, as a second example, you now carry out an implementation according to *Kantian* ethics (cf. Kant, 1870), you quickly get the problem that you do not know which maxim has to be implemented and how (cf. Powers, 2006).

On the flip side, algorithms have the potential to make racist decisions or be influenced by racially biased data (cf. Sandvig, 2016; Mittelstadt et al., 2016). It is crucial to recognize that we are increasingly entrusting algorithms with significant decision-making power, whether it is in lending or scoring practices, workforce management in large corporations, or various other domains. The mere existence of theoretical categories such as *skin color* or *race* can have far-reaching implications for selections and classifications (cf. Bowker / Star, 2000), which in turn affects the behavior of algorithms. Consequently, there is an ongoing discussion surrounding the need to eliminate racist elements from both the data used by algorithms and the algorithms themselves (cf. Ananny, 2016).

On the other hand, data and its dimensions often have more implicit structures than we realize (cf. Hedfeld, 2019). In medicine, for example, there has long been a demand for a medical monitor who can help to identify diseases at an early stage (cf. Del Rosario, 2015). With the introduction of the smartphone, this medical monitor was practically acquired. In this way, the smartphone can practically know in advance, based on the pure movement data, when the owner or wearer will become ill, before you notice it yourself, since the movement profile changes. It has long been recognized in e-commerce how important data and its collection are in order to offer the right product to the right customer (cf. Haque et al., 2018).

Using a language model, we can now input moral decisions to generate an implicit moral advisor. GPT, or Generative Pre-trained Transformer, is a type of model inspired by the functioning of the brain. It operates as a connectionistic and contextual system, capable of both long-term and short-term thinking. Acting as an interface to human language or human behavior, the model has the capacity to generate generic or general statements when provided with moral decisions stored in its short-term memory, facilitated by attention algorithms.

3 Language Models in Machine Learning

3.1 From Simple Statistics to GPT Model

What is the reason for using a language model and what is the advantage to put moral data into GPT?

Simply expressed in one sentence: GPT is a generalizing model that is connectionist and contextual, long-term *thinking* and serves as an interface to human language or humans and aims for normative and implementable acceptance. From a literary point of view, the modern digital literature can be described as connectionist in contrast to sequential computer generated texts (cf. Davis, 1958), as a very previous attempt by *Lutz* regarding the stochastic text (Stochastische Texte) of 1959 (cf. Lutz, 1959) and so called *konnektionistisches Paradigma* (cf. Bajohr, 2022: 152) or the idea that connectionism is a theory that AI can explain the brain and emotional states as well (cf. Rayburn / Diederich, 2013).

Let us look at the history of neuronal networks and let us see. The first nets where motivated by biology (cf. Rojas, 1997: 11-13). The concept behind neural networks is to mimic the functioning of axons in the brain (cf. Kaffka, 2017: 25f.). The biological components called dendrites cell bodies and axons work exactly like: input processing and output in computer science. As early as 1958, *Rosenblatt* presented the perceptron, which he implemented in a computer program for the first time in 1960 (cf. Kaffka, 2017: 34). If, in a further step, you build individual perceptrons or so-called neurons into networks, you get powerful generic models with one input and one output layer. These neuronal networks can now be used in various fields of application and areas (cf. Rey/Wender, 2011: 14). This leads to a new field of science so called machine learning where learning is *a specific body of knowledge and an associated set of techniques* (cf. Mitchell, 1997; cf. Bring et al., 2017: 5). Various types of neural networks have been developed, each suited for different tasks. Building upon the initial architecture of neural networks (NN), which includes input, hidden, and output layers, specialized networks have emerged, particularly for language-related tasks. These networks can process diverse forms of input data, contributing to their widespread application across various fields today. For instance, images can be converted into numerical data through a grid, language can be parsed into individual letters and subsequently numerical representations, and even musical elements, such as pitch, can be encoded as numbers. This versatility allows neural networks to model a wide range of inputs effectively.

It is essential to differentiate between supervised and unsupervised learning, with our focus directed towards unsupervised learning for our purposes. In supervised learning, a neural network (NN) is trained on labeled data, allowing it to associate specific inputs with corresponding outputs consistently. For instance, in image recognition tasks, images are inputted into the network, which learns to categorize or classify them into predefined categories, enabling recognition of objects or patterns within the images (cf. Sharma et al., 2018).

Unsupervised learning is an algorithm that learns patterns from unlabeled data. The goal is imitation, an important method of human learning that forces machines to construct concise representations of their world and generate imaginative content from them (Rojas, 1997: 97). As a next step for language analysis there are recurrent neural networks (RNN) which have loops in them, allowing information to persist. In 1986 the RNN was developed further for parallel processing and in a next step for language modelling (Rumelhart et al., 1986: 28; Jordan, 1986: 5; cf. Hopfield, 1982). The concept of a Recurrent Neural Network (RNN) is rooted in its ability to retain information from previous epochs and apply it to the learning process. For instance, when processing a sentence or a longer text, an RNN can recall the preceding elements and utilize this context to continue its analysis. The problem with RNN, however, is that there are difficulties in processing the *context* of a text over a longer *distance* (cf. Sherstinsky, 2020 ; Bengio et al., 1994).

This is the reason why the LSTM (long short-term memory) nets were developed (Hochreiter / Schmidhuber, 1997: 6-8). Through small concatenation of elements, it was possible to increase the context of the texts. This makes the approach long-term. So- called gates were introduced in this type of neural network: Input Gate, Forget Gate and Output Gate. (ibid.) It is possible to think of it this way non-mathematically, the input gate and forget gate store the information in the system that is needed for the context or needs to be removed accordingly (cf. Oinkina / Hakyll, 2015).

In the next phase, Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs) faced three significant obstacles: sequential computation inhibited parallelization, there was no explicit modeling of long and short-range dependencies, and the linear word distance between positions posed a challenge. To overcome these limitations, the attention algorithm was devised, playing a pivotal role in the advancement of language modeling (cf. Vaswani et al., 2017). The idea behind this is that each word of a sentence can have associated information. So, in order to decode accurately, you need to consider every word you enter and pay attention. We would call this context of the information e.g. the sentence: *I am a student* will be translated to French: *Je suis étudiant*. This means that *I am a* maps to *je suis* without considering two or three words but giving attention to the whole struct. In short, the attention algorithms are introducing queries, keys and values for every word (Vaswani et al., 2017: 3). With these three new parameters it is possible to give *more or less attention* to every word.

As a last step transformers were established and they used attention algorithms including self-attention algorithms. The idea of self-attention is to figure out how the sentence is mapped to itself e.g. the word *thinking* might have a bigger score on the *I* than on all the other words in the same sentence. Or in other words: In the sentence *I was thinking of a new house.* – you will have scores on every connection and relation between the words like: Who was thinking? What was it? and so on (Vaswani et al., 2017: 6).

As a last step the general pre-trained transformer were established. We would like it to use it in the investigation (cf. Radford / Narasimhan, 2018a). The more parameters the model has the more language it can produce (cf. Tunstall et al., 2022). We will use GPT-2 which has 1.5 billion parameters in the network for use which can be used for implicit language generation (cf. Huggingface, 2022; cf. Radford et al., 2018b). The API has still 124 million parameters based on *a very large corpus of English data* (cf. Huggingface, 2022). The number of parameters, the volume of data, and the linguistic reference are entirely adequate for the objectives of this paper. In future deliberations, it may be possible to opt for more advanced models or networks with increased parameters and linguistic input.

You can use the GPT2 model transformer as interface to humans because it generates text which can be read by humans and it learns very fast even compared to other language approaches (cf. Kojima et al., 2022).

3.2 The Language Transformer Model as an Enabler

Based on the ideas of *G.W.F. Hegel*, one of the most crucial human abilities is conceptual work and its influence on our thinking and social interactions. In *Hegel*'s philosophy, freedom holds a unique significance, as it emerges through the evolution of the spirit in more advanced forms of thought and action (cf. Seeberger, 1961). In systems philosophy, the term psychology appears in the subjective geist and is established there by intelligence and will, which unfold on ever new conceptual levels (Drüe et al., 2000: 274-283). In the will gradually unfold or developed *Genuss* (pleasure), *Neigung* (tendency), *freiere Aktivitäten* (free activities), which reach into the *objektiver Geist* (objective geist) where one finds, among other things, the legal concept in the science of the human beings *Philosophie des Zwischenmenschlichen* (cf. Jaeschke, 2010: 363). This method of self-movement of the term is also discussed in the literature (cf. Röttges, 1976; Drüe, 2000: 283).

The philosophy of reflection of the german idealism now assumes that on the one hand it is about the *Sollen* (should) but on the other hand also about the implementation (Lütge, 2002: 243; Hegel, 1986 [1801]: §68) because the philosophy should not stop at the pure term (cf. Hegel, 1986 [1821a]; 1986 [1821b]) because philosophy has to do with the idea which is not powerless in order to only ought and not to really be (cf. Hegel, 1986 [1830] §6).

Hegel's philosophy demonstrates the dialectical process that is crucial for connecting individuals within the subjective spirit (subjektiver Geist) as they transition to institutionalized forms within the objective spirit (objektiver Geist) and other societal structures. This influenced, among other things, *Homann*'s ideas on business ethics.

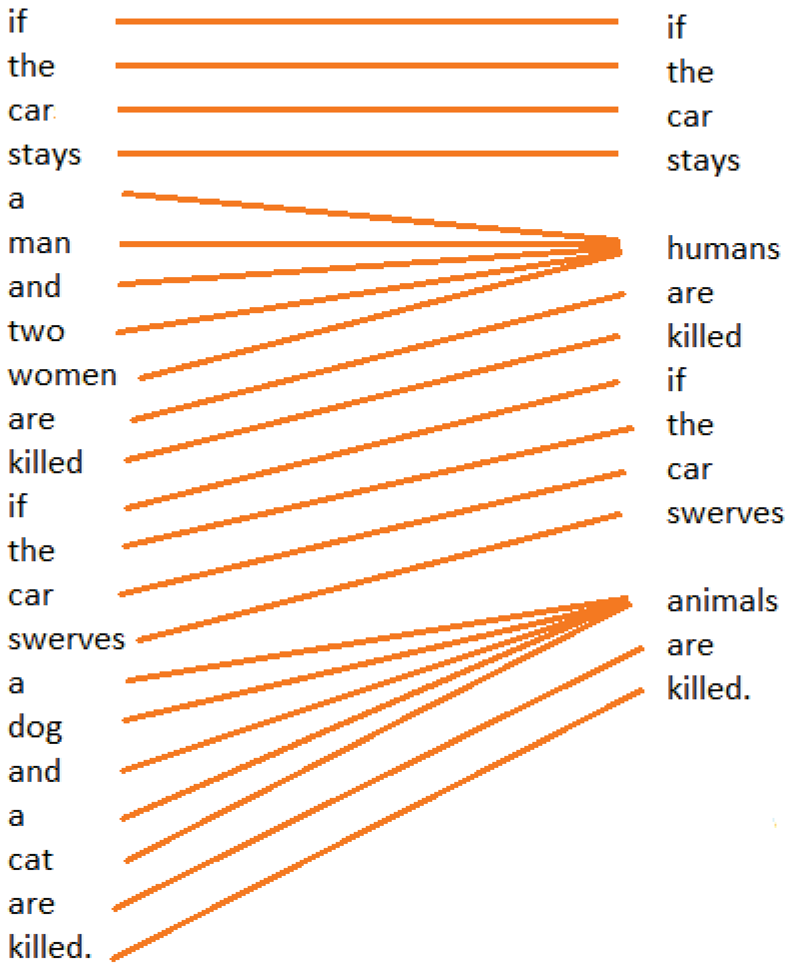
In the concept of a conversion paradigm, as proposed by Homann (Homann, 2002: 189), the emphasis is on individual actions. It revolves around the idea of the canon of duties and virtues, aiming to overcome tendencies to act solely out of duty by providing arguments and good reasons, even in cases of weak motivation to act. This involves three main aspects: firstly, recognizing (or rejecting) morality; secondly, defining actions communicatively; and thirdly, motivating oneself through informal coercion to adhere to these actions (cf. Homann, 2002: 190). A morally data-driven consultant can now do various things: propose normative ideas and principles, which are based on the implicit decision votes of the individually made decisions in the trolley problem e.g. *people are preferable to animals* or *the larger group should usually be spared.* If I'm not in a specific situation but have some time before making a decision, then these general guidelines can be unanimously accepted by everyone. This is because, on one hand, we all share a common humanity, and on the other hand, there's a high likelihood that a larger group of people would agree with these principles (e.g. in the simulations of consensus Homann/Lütge 2013: 83). The problem on the moral data or the implicit decisions might be that everything should be occurrences within the data: drives, inclinations, wrong decisions, in the idea on the development of the different stages etc. The moral computer-based consultant can also be seen as a novel tool that addresses existing limitations (cf. Homann, 2003) and opens up new avenues of thought. This is because the additional resources provided by the computer can now be integrated into the decision-making process, expanding the possibilities for consideration.

Implicit decision-making votes can be made visible in extreme situations and the action is not always justifiable for the acting actor because of its implicit structure (Minnameier 2016: 79). The idea of giving incentives is like an iceberg floating under the surface of the water (Homann, 2014: 232; cf. Minnameier, 2005).

However, this should serve a free decision-making (Homann, 2014: 14) and not be nudgy (cf. Thaler / Sunstein, 2009). It should serve about a connection between empiricism (the decision votes prepared by the moral data-driven consultant) and the normative values, which should lead to *normative judgements* (Normative Urteile) (Homann, 2014: 14).

In addition to *Handlungsethik* (action ethics) and *Ordnungsethik* (normative ethics), there is also the possibility of a third level, which can be understood as a human level, which in turn is available as a discourse about the norms or rules of the game (Pies, 2009: 11). The moral advisor could display decision votes or even offer default settings in certain situations, with the understanding that individuals have the opportunity to justify their decisions against these defaults for valid reasons. One goal could be to foster moral character development through habituation or the cultivation of virtues. This development is facilitated by the visible decision votes provided by the moral advisor (Homann, 2014: 242), without rigidly enforcing norms or declaring them universally valid indefinitely.

Figure 1: Mapping words in the moral data



Source: Illustration Based on (Vig / Belinkov, 2019).

The data provided by moral machines (cf. Awad, 2018b) can be used to establish a moral data-driven *level 2 implicit consultant* (or agent) according to *Moor* (cf. Moor, 2006). The advantages of this approach are manifold: Firstly, the advisor operates implicitly, allowing individuals the freedom to reject or modify its suggestions as desired (Homann, 2014: 214). Secondly, the data provided is based on decision votes that may initially appear as win-lose situations, but can be evolved into win-win outcomes through generalization. Thirdly, the language model is inspired by the human brain, enabling it to generate both simple judgments or decisions, as well as guidelines and general statements. Ultimately, human acceptance is crucial to ensure consensus and acceptance of the corresponding rules among people. The moral consultant represents a solution to scarcity by leveraging computer-controlled performance (cf. Homann, 2003).

The structure that now follows is based on the GPT (See Moral Machines file: MMdataReadMe.txt for a detailed description at Bonnefon 2018) API and is trained using the data from the trolley problem.

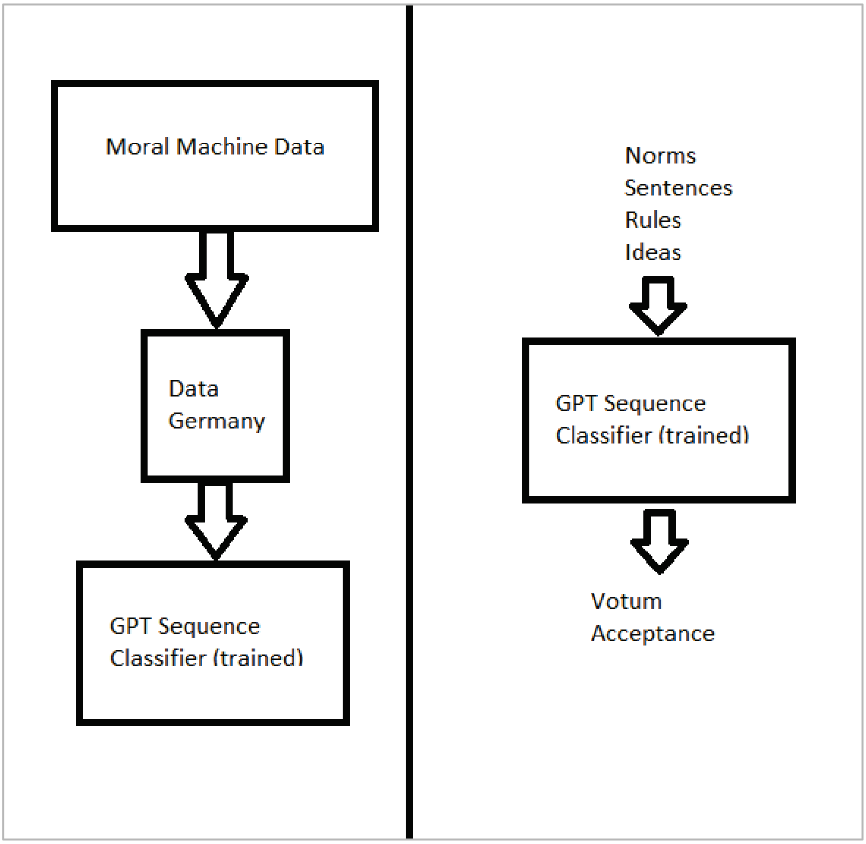
The data provide different scenarios, these are: *Utilitarian*, *Gender, Fitness, Age, Social Value, Species* and *Random*. Utilitarian represents a *more* or *less* problem, *gender* a different gender situation (male and female situations), *fitness* a difference between man and athlete man for example, *age* represents situations of different ages, for example old woman and woman, *social value* includes, among other things, comparisons with the an *executive* e.g., *species* compares humans and animals with each other (also in combination) and *random* was a remnant of a first data collection and contains several variations.

The data for the set *random* is not used for training the data and the whole data is divided into three groups: First: training data, Second: valuation data and Third: test data. The data is first processed and the answers are brought from two lines (the work's storage method) onto one line and is sorted. In a second step, the very structured data is translated into text. An advantage of language models is usually unstructured data, in this case the data is highly structured and prepared. A pre-trained GPT2 model is loaded via the API and then trained with initially ten thousand moral decisions in three epochs. For the test and the validation data, a few variants with two and three people in the difference are taken out of the *utilitarian* scenarios.

To give an example of the mapping: (data fields: Saved = 1, Intervention = 0) One men, two women and two dogs *in the data* will be in this in text: If the car stays one men, two women and a dog are saved. In order to test the model, the validation and the accuracy are measured and a mathematical confusion matrix is generated (a confusions matrix tests the different states for correctness, in this case it is a 2x2 matrix).

In the first field, test data are used to test the first state in the second field (stay on stay) then cross (swerve on stay). Due to the highly structured nature of the data, the accuracy values approach one hundred percent and the losses are less than one percent. Additionally, the confusion matrix closely resembles the identity matrix (You might think of a matrix with four fields in it. The first field represents all stay actions regarding the stay actions and so on. This means stay = stay is very high and swerve = swerve possibly but stay = swerve is very low in both cases).

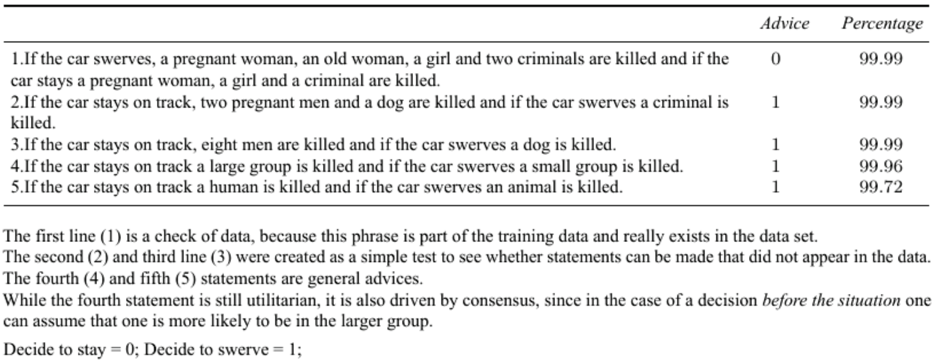
Figure 2: First results, potentials and human acceptance



The initial results were divided into three groups. Firstly, logical statements (1) were included in the moral data to verify if the model is functioning correctly. The expectation was for the majority of statements in the trained model to reproduce these logical statements accurately. Secondly, logical tests (2) and (3) aimed to assess if the model could generalize effectively. These tests involved scenarios not present in the original data, such as a pregnant man or a group larger than five individuals. The results showed a clear trend consistent with moral decision-making principles, including a preference for saving more people, as observed in evaluations by moral machines. This suggests that the model successfully reinforces logical connections during training. In simpler terms, if there's a connection between a pregnant woman and a woman, the model also establishes a stronger connection between a woman and a man, thereby implying a connection between a pregnant woman and a man.

Thirdly (4), (5), the model was asked about general norms. Since there must be a connection between man, woman etc. and human or humanity (or a connection to dog and cat to animal) and the fact of *more* or *less* is also shown, the two statements come about. In this way, a utilitarian approach is basically preserved but on a generic level regarding *Buchanan* as choice within the rules/choice of rules (cf. Buchanan, 1984).

Figure 3: First Results (own representation)



	Advice	*Percentage*
1.If the car swerves, a pregnant woman, an old woman, a girl and two criminals are killed and if the car stays a pregnant woman, a girl and a criminal are killed.	0	99.99
2.If the car stays on track, two pregnant men and a dog are killed and if the car swerves a criminal is killed.	1	99.99
3.If the car stays on track, eight men are killed and if the car swerves a dog is killed.	1	99.99
4.If the car stays on track a large group is killed and if the car swerves a small group is killed.	1	99.96
5.If the car stays on track a human is killed and if the car swerves an animal is killed.	1	99.72

The first line (1) is a check of data, because this phrase is part of the training data and really exists in the data set.
The second (2) and third line (3) were created as a simple test to see whether statements can be made that did not appear in the data.
The fourth (4) and fifth (5) statements are general advices.
While the fourth statement is still utilitarian, it is also driven by consensus, since in the case of a decision *before the situation* one can assume that one is more likely to be in the larger group.

Decide to stay = 0; Decide to swerve = 1;

It provides valuable data for refining the moral advisor. In addition, other language models (e.g. BERT) can also be used to check sentences for their implicit acceptance by the model (cf. Devlin et al., 2018). These sentences can also be examined humanely. This approach may look similar to training a harmless counselor, but it serves the purpose of finding or discussing general rules (cf. Bai et al., 2022) in respect of the idea of RLHF (Reinforcement Learning with Human Feedback). The concept of making probabilities visible also enhances moral education and further development, as it introduces the decision votes of individuals into the discussion, making them impossible to ignore. This transparency encourages engagement and fosters deeper understanding of the moral implications of decisions. Making it impossible to study a technology without the *value-system of the community* (cf. Martin / Freeman, 2004) maybe in a triangle of business, ethics and technology. The idea is that our quick decision system the so-called *system 1* is more at risk of having a racist bias than the slow *system 2* (cf. Kahneman, 2011; Agan et al., 2023). For this reason, education can also serve to use a moral, data-driven advisor or bring it at least into the discussion based on human acceptance, freedom and even a strong system 2 (slow system).

4 The Term Implicit

The term implicit in this paper occurs three times. First: implicit data. Second: implicit decisions and decision votes. Third: implicit agents. These terms must be kept apart. I would like to start with the concept of implicit knowledge based on the thoughts of *Nonaka* and *Takeuchi* (cf. Nonaka / Takeuchi, 2012).

Implicit knowledge here is an unused resource or something that can be made visible or explicit; it is described as premonition, intuition or a mental or cognitive understanding of the world (ibid.: 24). For the sake of simplicity, you can also write: Implicit is everything that does not appear explicitly.

4.1 Implicit Data

Modern neural networks require data for training. This paper addresses the question: What constitutes moral data? It is relatively straightforward to identify what constitutes racist data – examples abound, such as discrimination based on skin color, height, religion, or gender. But what defines moral data, essential for making morally sound, or at least better, decisions? (cf. Floridi / Taddeo, 2016). Furthermore, data can reveal implicit elements that may not be immediately apparent. My movement data reveals more about me than I might want to reveal. From Monday to Friday, I follow a routine of going to work and almost always sleeping at home in the evenings. Occasionally, I might meet my friends every Saturday at a sports stadium, where you might even discern my favorite club and more. Over time, observing me reveals implicit aspects in data that can offer insights into me or society beyond what we consciously realize. This understanding can increase the likelihood of accepting rules derived from our own data, as they reflect our behaviors and preferences implicitly.

4.2 Implicit Decisions and Decision Votes

For many years, researchers have studied whether elections could be predicted by better analyzing undecided voters (cf. Lundberg / Payne, 2014). Attitudes, ideas, or unconscious actions can provide valuable insights into preferences, including the choice of a political party or candidate (cf. Friese et al., 2012). Implicit decisions made by individuals have the potential to evolve into societal norms or rules that are accepted with consensus. The focus shifts from “What would I do in a situation?“ to “What would be beneficial for a person if this were to occur in this situation?“.

In this context, decision data or data from individuals can serve as the basis for statistical studies or large language models. The resulting rules can then be subject to social discussion or evaluation of their ability to garner consensus. Ultimately, the establishment of these rules or norms relies on free human acceptance.

4.3 Implicit (Moral) Agents

According to Moor, there exist different moral agents, among which the implicit ethical agent is worth considering. Moor suggests that if one wishes to instill ethics into a machine, one way is to constrain the machine’s actions to prevent unethical outcomes. In this approach, machine ethics is achieved by creating software that implicitly promotes ethical behavior, rather than explicitly containing ethical maxims. The machine behaves ethically because its internal functions naturally lead to ethical behavior or, at the very least, prevent unethical behavior. Ethical behavior becomes inherent to the machine’s nature, embodying virtues to some extent.

Computers can serve as implicit ethical agents when their design prioritizes safety or critical reliability concerns. For instance, automated teller machines and web banking software act as agents for banks, performing many tasks of human tellers and sometimes more. Given the ethical significance of transactions involving money, these systems must adhere to ethical standards.

Machines must be carefully constructed to give out or transfer the correct amount of money every time a banking transaction occurs. A line of code telling the computer to be honest won’t accomplish this. *Aristotle* suggested that humans could obtain virtue by developing habits. But with machines, we can build in the behavior with- out the need for a learning curve. Of course, such machine virtues are task specific and rather limited. (Cohn et al., 2022; Moor, 2006: 19)

Indeed, in 2006, *Moor* could not have foreseen major language models or their potential benefits. It can be argued that modern chat software exists somewhere between the implicit and explicit agent, as it engages explicitly with human interaction. *Moor* outlines four levels: the ethical impact agent, implicit ethical agent, explicit ethical agent, and full ethical agent. It is important to note that the implicit agent can only make suggestions and cannot decide on norms; only human acceptance or consensus can establish norms.

Statistics could be optionally included in a particular situation to indicate, for example, that a majority of people might have made a certain decision. However, this approach would compromise the generalizing nature of language models. Language models excel at generalization, offering the advantage of facilitating discussions about norms in light of consensus.

5 The Honorable AI Consultant

5.1 Potential

Let’s envision an honorable AI consultant—how about developing a transparent advisor guided by data-driven moral principles? We’re talking about an AI system designed to help with decisions, and while sources might not directly call it an “honorable AI consultant, “ we’ll use that term to highlight its advantages. These benefits aren’t exclusive to this type of AI; other similar systems can benefit too.

An outstanding feature of AI in general and therefore also of the *honorable AI-consultant* is that the system completes tasks in a significantly shorter time. It can handle and is able to process significantly larger data sets other than human capacities allow (cf. Lepri / Pentland, 2021: 1). Furthermore, it is unaffected by human limitations such as exhaustion, hunger, or boredom, as well as external influences like corruption or conflicts of interest (cf. Danziger / Avnaim-Pesso, 2011: 6892; Lepri et al., 2021: 1). Using such a consultant within companies offers cultural advantages. Teams can confidently manage tasks they’ve agreed upon, leading to more efficient collaboration and potentially strengthening team cohesion (cf. Ransbotham et al., 2021: 3-5). In her study, which involved 2, 197 managers and interviews with 18 managers implementing AI to enhance efficiency and decision-making, Ransbotham (Ransbotham et al., 2021: 1) found that 75 percent of respondents reported a positive impact on team morale due to AI usage.

One of the key strengths of the honorable AI consultant is its support for human decision-making. By utilizing this tool, individuals can enhance their analytical and decision-making abilities through access to precise information provided by the system at the most opportune moments. (cf. Wilson / Daugherty, 2018: 6).

Through this support, people can continue their activities on a larger scale, with higher quality and with increased speed (cf. Lepri et al., 2021: 1). The use of the *honorable AI consultant* for decision-making support can also lead to decisions being made more objectively. Fewer errors could occur (cf. Lepri et al., 2021: 1). At this point, it is crucial to demonstrate why human decision-making is not necessarily optimal and how the use of an honorable AI consultant can be both necessary and advantageous. Human decision-making processes are often influenced by external factors and often lead to decisions in a context of incomplete information and resort to the use of heuristics.

Decisions are often influenced by individual characteristics of the decision maker such as origin, personal experience, level of knowledge, acceptance of uncertainty safety and risk as well as cognitive limitations (cf. Aharoni / Tihanyi / Connelly, 2011: 137). Emotions also often play an important role of human decision making (cf. Kouchaki / Desai, 2015: 369). In numerous scenarios, emotions may undoubtedly be beneficial. But in other situations, for example, where an unbiased perspective is required, such emotions have a detrimental effect (cf. Wallach et al., 2008: 567).

Furthermore, it is important to acknowledge that individuals frequently lack the comprehensive information required to make rational decisions. This deficiency may arise from various factors, such as time constraints or limited cognitive resources. Alternatively, even when information is accessible, emotional biases or intuitive influences may prevent its incorporation into the decision-making process (cf. Giubilini / Savulescu, 2018: 170).

Heuristics also play an important role in human decision-making. Heuristics primarily serve to make decisions as quickly and with as little effort as possible (cf. Gigerenzer / Brighton, 2009: 109). Efficient cognitive processes, both conscious and unconscious, play a role in decision-making. Often, some information is disregarded to expedite the decision-making process. Individuals may accept a trade-off between speed and accuracy, prioritizing swiftness and efficiency in their cognitive processes. The conventional perspective suggests that decisions based on heuristics may result in greater errors compared to decisions made using logical or statistical models (cf. Gigerenzer / Gaissmaier, 2011: 451).

Recent research however, in certain circumstances, reveals a *less is more effect* wherewith a reduction in information and calculations, the accuracy starts at one can increase to a certain point (cf. Artinger, Petersen, Gigerenzer & Weibler, 2015: 35-36; cf. Gigerenzer / Brighton, 2009: 107-110). In particularly unpredictable or complex situations, omitting certain information often results in more accurate judgments compared to attempting to consider and weigh all available information (cf. Gigerenzer / Gaissmaier, 2011: 455-473).

In addition to general decision making, AI can also be used as a support serve as a basis for ethical decision-making processes. An artificial system like that *honorable AI advisors* based on pre-learned moral guidelines can make moral decisions with increased speed and efficiency exceed the capabilities of the human brain. Such AI enables initially informed and optimized moral decisions (cf. Giubilini / Savulescu, 2017: 185). There is a possibility that such AI systems could even make more moral decisions than humans because they are not irrational emotional influences, personal interests or stress (cf. Allen, 2002: 2-3; Wallach et al., 2008: 567). It is worth exploring why human decisions often deviate from consistent adherence to moral guidelines. Firstly, individuals frequently fail to consistently apply the ethical principles they consciously endorse. Secondly, weaknesses in willpower or specific neurophysiological conditions may hinder their ability to do so. Nonetheless, the aim is to make optimal judgments based on well-founded information, despite these challenges (cf. Giubilini / Savulescu, 2018: 170). An individual may fundamentally reject violence, but human motivation is influenced by various physiological factors that often lie beyond conscious control and can trigger aggressive tendencies, such as low blood sugar levels (cf. Bushman / DeWall / Pond / Hanus, 2014: 6254-6255; Giubilini / Savulescu, 2018: 170-171). Thirdly, individuals shape their decisions, and consequently, moral discussions, through discussions, conversations, and debates (cf. Weiss, 2021: 96-97).

Finally, the assessment of moral aspects reveals itself to be a complex skill, which many do not learn fully or only practice with limited success (cf. Wallach et al., 2008: 565).

It is evident that human decisions are frequently not purely rational. Therefore, decisions shouldn’t rely solely on emotions or instinctive inclinations, even if they occasionally yield more accurate outcomes. It is essential to incorporate an objective perspective wisely to potentially arrive at the best possible decision. Furthermore, it is been demonstrated that making morally correct decisions can be challenging for individuals. The *honorable AI consultant* can serve as a solution by assuming the role of an objective observer, free from emotional, psychological influences, and cognitive limitations inherent in human decision-making processes.

The *honorable AI advisor* presents an opportunity to harness the strengths of integrating AI systems with human efforts. Through its exclusive advisory role, it fosters collaboration, crucial for maximizing the impact of AI technologies in augmenting and complementing human capabilities. (cf. Wilson / Daugherty, 2018: 4). Thus, the goal is to enhance human capacity to act, without doing so at the expense of replacing human responsibility (cf. Floridi et al., 2018: 691-692; Wilson / Daugherty, 2018: 4). The collaborative effort between humans and machines facilitates the maximum performance enhancement. Human strengths such as leadership, teamwork, creativity, and social understanding complement the speed, scalability, and quantitative capabilities of an honorable AI consultant. Human involvement is essential for training machines, interpreting results, and ensuring responsible usage, making a difference in the outcome (cf. Wilson & Daugherty, 2018: 4-6). As mentioned earlier, *the honorable AI advisor* functions solely as a consultant, offering recommendations and decisions that aren't obligatory for implementation but serve to provide support and suggestions. This approach aims to enhance trust in such a system, as people are generally more receptive to machines that offer advice rather than independently making (moral) decisions. (Misselhorn, 2018: 73; cf. Misselhorn, 2021).

Ultimately, the greatest opportunity lies in the thoughtful implementation of the *honorable AI advisor's* role, which encompasses both economic and ethical considerations in corporate decision-making. Integrating ethical principles is particularly crucial as it enhances trust in a company, contributing significantly to consumer and societal confidence in its operations. By incorporating contemporary criteria, the AI advisor can play a pivotal role in a company's success and achievement of its goals. (cf. Ferrell et al., 2021: 3-4).

The *honorable AI consultant* can make a significant contribution to improving the perspective to significantly improve long-term success and company stability.

5.2 Challenges and Limitations

The implementation of AI comes with a range of challenges and limitations. The *honorable AI consultant*, in particular, faces specific difficulties inherent in the design and structure of AI systems. One of the most significant challenges lies in the fact that artificial systems are initially perceived as black boxes by users. Initially, it is unclear how these systems make decisions and on what basis. Therefore, the current lack of transparency in AI systems poses a significant problem. (cf. Deckert / Meyer, 2020: 30). This lack of transparency can stem from two main factors. Firstly, a significant portion of society lacks the technical expertise needed to comprehend the inner workings of such systems. Secondly, certain machine learning algorithms are inherently complex and challenging to interpret. This inherent opacity of the systems further contributes to the overall lack of transparency (cf. Burrell, 2016: 1-5). The use of *honorable AI consultants* to support decision-making processes has proven to be highly demanding, particularly in ensuring the reliability of the system's suggestions. This reliability would only be achievable if users possess sufficient technical knowledge to evaluate the situation. However, even in such cases, the decision-making process may remain partially opaque and not fully traceable. (cf. Deckert / Meyer, 2020: 30).

Another difficulty lies in the data, which is the fundamental basis for each serve AI. Data is essential for training and testing such systems. These later serve as the basis for recommendations and decisions. Included the quality of data plays a crucial role (cd. Deckert / Meyer, 2020:30).

However, the quality is inadequate due to inconsistencies, missing variables, sufficient scope or outdated data sets are often not guaranteed (cf. Vollhardt et al., 2021: 121). Accordingly, it involves considerable effort and costs high-quality data for the conception, implementation and use of an *honorable AI consultants* (cf. Decker / Meyer, 2020: 30). The data-protection is of great importance. Records can contain sensitive information contain characteristics and behaviors of people. Behavioral data from new other sources, such as social media, enable learning algorithms draw conclusions about private information that may never have been disclosed (cf. Lepri et al., 2021: 4).

Another problem with AI systems is that they can contain biases. This bias is also often cited as one of the main risks of called AI (cf. Golbin / Rao, 2019). In connection with computer systems or in this case, AI-supported systems, the term *bias* is used to describe the system thematic and unjust discrimination against certain individuals or groups.

To express in favor of others (cf. Friedman / Nissenbaum, 1996: 332). This highlights the concern that AI systems can generate biased outcomes. When *honorable AI consultants* are involved in supporting, advising, and making decisions that impact individuals, there is a potential challenge of producing discriminatory results. Such outcomes could contravene principles of justice and equality, while also adversely affecting specific individuals or communities (cf. Kordzadeh / Ghasemaghaei, 2022: 1). Such distortion can manifest at various stages of algorithm development, often through the (unintentional) integration of biases. Examples include contaminated or distorted input data and the design of the algorithm itself, which may incorporate biased properties, weightings, or objectives (Golbin / Rao, 2019; Kordzadeh / Ghasemaghaei, 2022: 3). An AI system may also overlook distorted data, potentially leading to discriminatory outcomes. For instance, in a scenario where an AI system is used to screen job applicants based on historical company data, it could inadvertently perpetuate gender or ethnic biases against groups that have historically been underrepresented (cf. Golbin / Rao, 2019). In particular, the strong dependence of such systems on their training data makes them susceptible to prejudice (Lenzen, 2023: 886). Real-world examples often reveal biases, such as gender discrimination in hiring programs or decisions regarding credit limits (Golbin / Rao, 2019; Kordzadeh / Ghasemaghaei, 2022:1-11).

Another challenge in developing the *honorable AI consultant* is the evolution of morality over time. What is currently considered to conform to ethical standards may be questioned and deemed morally problematic in the future. Therefore, continuous monitoring and updating of the system's moral values are necessary to support moral development and progress in a measured manner, ensuring that the AI reflects evolving ethical perspectives (cf. Lara / Deckers, 2020: 279).

Another difficulty arises from the chosen method of moral implementation. In experiments such as *the Moral Machine*, which served as inspiration, it became evident that even seemingly clear preferences exhibited cultural differences. (Awad et al., 2018a: 59-63). Given that companies often operate within multicultural environments or have the potential to do so, diversification of ethical perspectives may arise in this context. Differences in ethical stances can be expected. However, predicting such diversification is challenging, as the moral decisions made in experiments like the Moral Machine differ fundamentally from those applicable to the honorable AI consultant. In the proposed concept, decisions would pertain to moral dilemmas within an economic context rather than decisions concerning life-threatening situations. Therefore, this presents a potential challenge that may or may not materialize. It is worth mentioning worth that Awad et al. (2018a: 59-63) have found that large parts of the world show some agreement in their ethical preferences.

It is important to understand how the honorable AI consultant affects human decisions. There's a danger that people might get hurt if they rely too much on its recommendations without questioning or thinking carefully about them (cf. Busuioc, 2021: 26). This excessive dependence on AI system recommendations is known as automation bias—the tendency to uncritically trust automated decisions over human judgment. This phenomenon occurs when users overly trust the decisions made by automated support systems, leading to decreased vigilance in seeking and processing information (Busuioc, 2021: 26; Lyell / Coiera, 2017: 423). Krügel / Ostermaier / Uhl, 2022: 1-3) also found that users trust the ethical advice of an AI even without information about the training data. Interestingly, this trust remains even when users have information that could potentially raise doubts about the system's reliability. Their study suggests that people are more likely to place excessive trust in an AI than to distrust it. (Krügel et al., 2022:1-20).

5.3 Rules, Guidelines and Responsibility

The preceding section highlighted the challenges and constraints inherent in the development and utilization of an honorable AI consultant, as well as those encountered with AI systems in general. AI is becoming increasingly integrated into various aspects of human life, bringing about significant transformations with profound impacts on numerous social domains. The importance of cultivating ethically guided programming and implementation of such systems is becoming increasingly evident. In recent years, various educational institutions, private companies, and public sector organizations have developed and published guidelines for ethical AI.

Jobin, *Ienca* and *Vayena* (Jobin / Ienca / Vayena, 2019: 389) came to the conclusion that there is global agreement on five ethical fundamental principles. These are (1) transparency, (2) fairness and justice, (3) non-harm, (4) responsibility and (5) data protection (cf. Jobin et al., 2019: 389). In their study, they emphasize the substantial divergence in interpretations of these principles, their applicability to different topics, domains, or stakeholders, the appropriate methods for implementation, and why adherence to these principles is deemed important (cf. Jobin et al., 2019: 389). To ensure adherence to the five principles despite differences, we can refer to *Sarah Spiekerman*'s explanation (2021: 248). A crucial aspect is the requirement for transparency in the functionality and decision-making processes of artificial systems. This necessitates documentation and communication about their operations and decision-making to be appropriate, accessible, comprehensive, meaningful, and truthful (cf. Spiekermann, 2016: 59; cf. Spiekermann, 2019). Ensuring equal treatment of all users and avoiding systematic bias through AI are central requirements within the framework of ethical principles such as fairness and justice. It is also vital to ensure equal rights and accessibility of the system for all users. The ethical principle of non-harm is particularly relevant to system security, emphasizing the importance of preventing any damage or negative effects on users or society. Responsibility entails that those overseeing AI systems are accountable and responsible for their actions. Integrity is crucial to ensure that responsible actions are carried out ethically and with integrity. Finally, the ethical principle of data stewardship pertains to data handling, encompassing principles such as those outlined in the European General Data Protection Regulation, including the right to be forgotten, data portability, informed access, among others (cf. Spiekermann, 2021: 248).

Ethical guidelines for AI offer a framework to strike a balance between harnessing the diverse capabilities of AI and ensuring oversight over its development and impact. The social acceptance of AI technologies hinges on whether the benefits are deemed significant and whether the risks are perceived as avoidable, reducible, or controllable. (cf. Floridi et al., 2018: 694). Through the introduction and implementation principles and thus the creation of an ethical AI – also responsible called conscious AI – the hope is to build trust in the systems as well to limit negative effects (cf. Eitel-Porter, 2021: 73). It is also important meaning to take responsibility through such guidelines, which type of AI is developed and how it is used (cf. Floridi et al., 2018: 692). In the future there might be a need for standardization of such ethical guidelines to ensure the safe use of AI (cf. Jobin et al., 2019: 396-397). In addition to ethical guidelines, there is a strong emphasis on enhancing digital skills and promoting the responsible use of algorithms (cf. Krügel et al., 2022: 21). The implementation of education initiatives for users could be considered as a possible solution (cf. Burrell, 2016: 10) to have robust governance and compliance mechanisms in place to integrate corporate structures to avoid unwanted negative effects of AI systems (cf. Eitel-Porter, 2021: 73). A typical recommendation for one appropriate leadership structure to ensure responsible AI is to establish a two-tier structure at the top. On one hand, it is recommended to establish an ethics council or advisory group to incorporate external contributions from society. On the other hand, it is proposed to establish an ethics committee or review board internally to guide and monitor the focus on responsible AI (cf. de Laat , 2021: 163). Companies also need support by employees who continually work to ensure that AI systems are operate in a measured, safe and responsible manner (cf. Wilson / Daugherty, 2018: 11). There are concrete steps to implement a robust leadership structure for example *Ray Eitel-Porter* (cf. 2021: 73-80) in his article *Beyond the promise: implementing ethical AI*.

In conclusion, adhering to ethical guidelines throughout the development and utilization stages of the *honorable AI consultant*, alongside the implementation of suitable leadership and control structures, can assist in mitigating the challenges that arise and ensure responsible usage of the system.

An initial step could involve integrating an implicit moral agent using data from human decision-making, which could implicitly align with societal rules and norms, leading to a win-win situation. Subsequently, the introduction of an *honorable AI consultant* could further contribute to the establishment of acceptable and implementable rules for society.

References

Agan, A. Y. / Davenport, D. / Ludwig, J. / Mullainathan, S. (2023): Automating Automaticity: How the Context of Human Choice Affects the Extent of Algorithmic Bias, National Bureau of Economic Research, Cambridge, No. w30981.

Aharoni, Y. / Tihanyi, L. / Connelly, B. L. (2011): Managerial decision-making in international business: A forty-five-year retrospective. Journal of World Business, 46(2), 135–142. https://doi.org/10.1016/j.jwb.2010.05.001.

Ananny, M. (2016): Toward an ethics of algorithms: Convening, observation, probability, and timeliness. In: Science, Technology, & Human Values, 41(1), 93-117.

Artinger, F. / Petersen, M. / Gigerenzer, G. / Weibler, J. (2015): Heuristics as adaptive decision strategies in management. Journal of Organizational Behavior, 36 (S1), 33-52. https://doi.org/10.1002/job.1950.

Awad, E. / Dsouza, S. / Kim, R. / Schulz J. / Henrich, J. / Shariff, A. / Bonnefon, J.F. / Rahwan, I. (2018a): The Moral Machine experiment. In: Nature 563, 59–64. https://doi.org/10.1038/s41586-018-0637-6.

Awad, E. (2018b): Moral Machine. https://osf.io/3hvt2/.

Bai, Y. / Jones, A. / Ndousse, K. / Askell, A. / Chen, A. / DasSarma, N. / Kaplan, J. (2022): Training a helpful and harmless Assistant with reinforcement Learning from human Feedback. arXiv preprint arXiv: 2204.05862.

Bartneck, C. / Lütge, C. / Wagner, A. / Welsh, S. (2019): Ethik in KI und Robotik. München: Carl Hanser Verlag.

Bajohr, H. (2022): Schreibenlassen. Texte zur Literatur des Digitalen. Berlin: August Verlag.

Bengio Y. / Simard, P. / Frasconi, P. (1994): Learning long-term dependencies with gradient descent is difficult. In: IEEE Transactions on Neural Networks Vol. 5 (2), 157-166, March 1994, https://ieeexplore.ieee.org/document/279181.

Bentham, J. / Mill, J.S.(2004): Utilitarianism and other essays. London: Penguin UK.

Bonnefon, J.-F. / Shariff, A. / Rahwan, I. (2016): The social dilemma of autonomous vehicles. In: Science, 352(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654.

Bowker, G. C. / Star, S. L. (2000): Sorting things out: Classification and its consequences. MIT press.

Buchanan, J.M. (1984): Die Grenzen der Freiheit: Zwischen Anarchie und Leviathan, Tübingen: Mohr.

Burrell, J. (2016): How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 2053951715622512. https://doi.org/10.1177/2053951715622512.

Bushman, B. J. / DeWall, C. N. / Pond, R. S. / Hanus, M. D. (2014): Low glucose relates to greater aggression in married couples. Proceedings of the National Academy of Sciences, 111(17), 6254–6257. https://doi.org/10.1073/pnas.1400619111.

Busuioc, M. (2021): Accountable Artificial Intelligence: Holding Algorithms to Account. Public Administration Review, 81(5), 825–836. https://doi.org/10.1111/puar.13293.

Cohn, A. / Gesche, T. / Maréchal, M. A. (2022): Honesty in the digital age. In: Management Science 68(2), 827-845.

Danziger, S. / Levav, J. / Avnaim-Pesso, L. (2011): Extraneous factors in judicial decisions. Proceedings of the National Academy of Sciences, 108(17), 6889 6892. https://doi.org/10.1073/pnas.1018033108.

Davis, M.(1958): Computability and Unsolvability, New York: McGraw-Hill.

de Laat, P. B. (2021). Companies Committed to Responsible AI: From Principles towards Implementation and Regulation? Philosophy & Technology, 34(4), 1135–1193. https://doi.org/10.1007/s13347-021-00474-3.

Deckert, R. / Meyer, E. (2020): Digitalisierung und Künstliche Intelligenz: Kooperation von Menschen und Maschinen aktiv gestalten. Springer Fachmedien. https://doi.org/10.1007/978-3-658-31795-9.

Del Rosario, M.B. / Redmond, S.J. / Lovell, N.H. (2015): Tracking the Evolution of Smartphone Sensing for Monitoring Human Movement. In: Sensors, 15, 18901-18933, Österreichische Artificial-Intelligence-Tagung (KONNAI) Salzburg, Österreich, 18.–21. September 1990 Proceedings, 252, 191. https://doi.org/10.3390/s150818901.

Devlin, J. /Chang, M.W. /Lee, K. /Toutanova, K. (2018): Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Drüe H. / Gethmann-Siefert A. / Hackenesch, C. / Jaeschke, W. / Neuser W. / Schnädelbach H. (2000): Hegels Enzyklopädie der philosophischen Wissenschaften (1830), Frankfurt am Main: Suhrkamp.

Eitel-Porter, R. (2021): Beyond the promise: Implementing ethical AI. AI and Ethics, 1(1), 73–80. https://doi.org/10.1007/s43681-020-00011-6.

Ferrell, O. C. / Fraedrich, J. / Ferrell, L. (2021): Business Ethics: Ethical Decision Making and Cases. Cengage Learning.

Friedman, B. / Nissenbaum, H. (1996): Bias in computer systems. ACM Transactions on Information Systems, 14(3), 330–347. https://doi.org/10.1145/230538.230561.

Friese, M. / Smith, C. T. / Plischke, T. / Bluemke, M., / Nosek, B. A. (2012): Do implicit attitudes predict actual voting behavior particularly for undecided voters?. PLoS ONE 7(8): e44130. https://doi.org/10.1371/journal.pone.0044130

Floridi, L. / Taddeo, M. (2016): What is data ethics?. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083), 20160360.

Floridi, L. / Cowls, J. / Beltrametti, M. / Chatila, R. / Chazerand, P. / Dignum, V. / Luetge, C. / Madelin, R. / Pagallo, U. / Rossi, F. / Schafer, B. / Valcke, P. / Vayena, E. (2018): AI4 People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Foot, P. (1978): The Problem of Abortion and the Doctrine of the Double Effect, In: Philippa Foot: Virtues and Vices and Other Essays in Moral Philosophy, Oxford: Oxford University Press, 19–32.

Gigerenzer, G. / Brighton, H. (2009): Homo Heuristicus: Why Biased Minds Make Better Inferences. Topics in Cognitive Science, 1(1), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x.

Gigerenzer, G. / Gaissmaier, W. (2011): Heuristic Decision Making. Annual Review of Psychology, 62(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346.

Giubilini, A. / Savulescu, J. (2018): The Artificial Moral Advisor. The “Ideal Observer” Meets Artificial Intelligence. Philosophy & Technology, 31(2), 169–188. https://doi.org/10.1007/s13347-017-0285-z

Gips, J.(1994). Toward the ethical robot. https://philarchive.org/rec/GIPTTE.

Golbin, I. / Rao, A. (2019): What is fair when it comes to AI bias? Strategy+business. https://www.strategy-business.com/article/What-is-fair-when-it-comes-to-AI-bias.

Haque T. U. / Saber N. N. / Shah F. M.(2018): Sentiment analysis on large scale Amazon product reviews. IEEE International Conference on Innovative Research and Development (ICIRD) Bangkok, Thailand, 1-6, https://ieeexplore.ieee.org/document/8376299.

Hedfeld, P. (2019): Die neuen Glasperlen. Neue Gesellschaft Frankfurter Hefte, Jg. 2019, Nr. 10, 30-32, Berlin: Dietz Verlag.

Hegel, G.W.F. (1801, 1986): Jenaer Schriften 1801-1807 Berlin: Suhrkamp Verlag.

Hegel, G.W.F. (1821a, 1986): Grundlinien der Philosophie des Rechts Berlin: Suhrkamp Verlag.

Hegel, G.W.F. (1821b, 1986): Vorlesungen über die Geschichte der Philosophie III Berlin: Suhrkamp Verlag.

Hegel, G.W.F. (1830, 1986): Enzyklopädie der philosophischen Wissenschaften I Berlin: Suhrkamp Verlag.

Hochreiter, S. / Schmidhuber, J. (1997): Long Short-term Memory. In: Neural computation. 9. p.1735-1780. 10.1162/neco.1997.9.8.1735.

Homann, K. / Lütge, C. (2003): Anreize und Moral. Gesellschaftstheorie–Ethik-Anwendungen. Philosophie und Ökonomik Band 1. Münster: LIT Verlag.

Homann, K. / Lütge C. (2013): Einführung in die Wirtschaftsethik (3rd. Ed.) Berlin: LIT Verlag Dr. W. Hopf.

Homann, K. (2014): Sollen und Können Wien: Ibera Verlag.

Homann, K. (2020): Praktische Philosophie und ökonomische Theorie – Aufsätze und Vorträge Berlin: LIT Verlag Dr. W. Hopf.

Hopfield, John. (1982): Neural Networks and Physical Systems with Emergent Collective Computational Abilities. In: Proceedings of the National Academy of Sciences of the United States of America. Vol. 79. p. 2554-2558. 10.1073/pnas.79.8.2554.

HuggingFace (2022): GPT2. https://huggingface.co/gpt2.

Jaeschke, W. (2016): HEGEL-Handbuch: Leben–Werk–Schule (2nd. edition). Stuttgart: Springer-Verlag.

Jobin, A. / Ienca, M. / Vayena, E. (2019): The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2.

Jordan, M I. (1986): Serial order: a parallel distributed processing Approach. Technical report, June 1985-March 1986. United States: N. p.

Kahneman, D. (2011): Thinking, fast and slow. Macmillan.

Kant, I. (1870): Grundlegung zur Metaphysik der Sitten (Vol. 28). L. Heimann.

Kaffka, T. (2017): Neuronale Netze Grundlagen. mitp: Frechen.

Kordzadeh, N. / Ghasemaghaei, M. (2022): Algorithmic bias: Review, synthesis, and future research directions. European Journal of Information Systems, 31(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212 .

Kouchaki, M. / Desai, S. D. (2015): Anxious, threatened, and also unethical: How anxiety makes individuals feel threatened and commit unethical acts. Journal of Applied Psychology, 100(2), 360–375. https://doi.org/10.1037/a0037796 .

Kojima, T. / Gu, S. / S., Reid M. / Matsuo, Y. / Iwasawa, Y.): Large language models are zero-shot reasoners. https://doi.org/10.48550/arXiv.2205.11916.

Krügel, S. / Ostermaier, A. / Uhl, M. (2022): Zombies in the Loop? Humans Trust Untrustworthy AI-Advisors for Ethical Decisions. Philosophy & Technology, 35(1), 17. https://doi.org/10.1007/s13347-022-00511-9 .

Lara, F. / Deckers, J. (2020): Artificial Intelligence as a Socratic Assistant for Moral Enhancement. Neuroethics, 13(3), 275–287. https://doi.org/10.1007/s12152-019-09401-y .

Lepri, B. / Oliver, N. / Pentland, A. (2021): Ethical machines: The human-centric use of artificial Science. 2021 Mar 3;24(3):102249. doi: 10.1016/j.isci.2021.102249. PMID: 33763636; PMCID: PMC7973859 url: https://pubmed.ncbi.nlm.nih.gov/33763636/

Lundberg, K. B., / Payne, B. K. (2014): Decisions among the undecided: Implicit attitudes predict future voting behavior of undecided voters. PloS one, 9(1), e85680.

Lütge, C. (Ed.) /Karl Homann (2002): Vorteile und Anreize. Tübingen: Mohr Siebeck.

Lutz, T. (1959): Stochastische Texte. https://auer.netzliteratur.net/0_lutz/lutz_original.html.

Lyell, D. / Coiera, E. (2017): Automation bias and verification complexity: A systematic review. Journal of the American Medical Informatics Association, 24(2), 423–431. https://doi.org/10.1093/jamia/ocw105.

Martin, K. E. / Freeman, R. E. (2004): The separation of technology and ethics in business ethics. Journal of Business Ethics, 53, 353-364.

Minnameier, G. (2005): Wer Moral hat, hat die Qual, aber letztlich keine Wahl!: Homanns (Wirtschafts) Ethik im Kontext der Wirtschaftsdidaktik. Zeitschrift fürs Berufs- und Wirtschaftspädagogik 101, 19-42.

Minnameier, G. (2016): Moralische Motivation und ökonomische Rationalität. In: Ethik und Beruf: Interdisziplinäre Zugänge, 73. Bielefeld: wbv Media GmbH & Company KG.

Misselhorn, C. (2018): Grundfragen der Maschinenethik. Ditzingen: Reclam.

Misselhorn, C. (2021): Künstliche Intelligenz und Empathie. Ditzingen: Reclam.

Mitchell T. (1997): Machine Learning. New York: McGraw-Hill Education International Edition.

Mittelstadt, B. D. / Allo, P. / Taddeo, M. / Wachter, S. / Floridi, L. (2016): The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

Moor, J.H. (2006): The Nature, Importance, and Difficulty of Machine Ethics. Machine Ethics IEEE, August 2006, 18-21.

Nonaka, I. / Takeuchi, H. (2012): Die Organisation des Wissens, wie japanische Unternehmen eine brachliegende Ressource nutzbar machen. Frankfurt am Main: Campus Verlag.

Oinkina / Hakyll (2015): Understanding LSTM Networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Pies, I. (2009): Moral als Heuristik. Ordonomische Schriften zur Wirtschaftsethik Berlin: WVB, Wissenschaftlicher Verlag.

Powers, T. M.(2006): Prospects for a kantian machine. Intelligent Systems, IEEE Volume 21(4), August 2006, 46-51.

Radford, A. / Narasimhan, K.(2018a): Improving Language Understanding by Generative Pre-Training.

Radford, A. / Wu, J. / Child, R. / Luan, D. / Amodei, D. / Sutskever, I. (2018b): Language Models are Unsupervised Multitask Learners.

Ramge, T. (2018): Mensch und Maschine: Wie Künstliche Intelligenz und Roboter unser Leben verändern. Ditzingen: Reclam.

Ransbotham, S. / Candelon, F. / Kiron, D. / LaFountain, B. / Khodabandeh, S. (2021): The cultural benefits of artificial intelligence in the Enterprise. MIT Sloan Management Review and Boston Consulting Group: Cambridge, MA, USA.

Rayburn, W. M., / Diederich, J. (1990): Some Remarks on Emotion, Cognition, and Connectionist Systems. Konnektionismus in Artificial Intelligence und Kognitionsforschung: 6.Österreichische Artificial-Intelligence-Tagung (KONNAI) Salzburg, Österreich, 18.–21. September 1990 Proceedings, 191-195. Berlin & Heidelberg : Springer Verlag.

Rojas, R. (1993): Theorie der neuronalen Netze. Eine systematische Einführung. Berlin: Springer-Lehrbuch.

Röttges, H. (1976): Der Begriff der Methode in der Philosophie HEGELs. Meisenheim am Glan: Verlag Anton Hain.

Riek, L. / Howard, D. (2014): A Code of Ethics for the Human-Robot Interaction Profession. Proceedings of We Robot 2014. https://ssrn.com/abstract=2757805.

Rey, G. / Wender, K. (2011): Neuronale Netze. (2nd Ed.) Bern: Huber.

Rumelhart, D. / Hinton, G. / Williams, R. (1985): Learning internal representations by error propagation. Tech. rep. ICS 8504. San Diego, California: Institute for Cognitive Science, University of California.

Sandvig, C. / Hamilton, K. / Karahalios, K. / Langbort, C. (2016): Automation, Algorithms, and Politics when the Algorithm itself is a racist: Diagnosing ethical

Harm in the basic Components of Software. International Journal of Communication, 10, 19.

Seeberger, W.(1961): HEGEL oder die Entwicklung des Geistes zur Freiheit Stuttgart: Klett Verlag.

Sharma N. / Jain V. / Mishra A.(2018): An Analysis Of Convolutional Neural Networks For Image Classification. Procedia Computer Science Volume 132, 2018, 377-384, https://doi.org/10.1016/j.procs.2018.05.198.

Sherstinsky, A. (2020): Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. Physica D: Nonlinear Phenomena journal, Volume 404, March 2020: Special Issue on Machine Learning and Dynamical Systems,1-43. https://doi.org/10.1016/j.physd.2019.132306.

Spiekermann, S. (2016): Ethical IT Innovation. Boca Raton: CRC Press.

Spiekermann, S. (2019): Digitale Ethik. Pößneck: Droemer Verlag.

Spiekermann, S. (2021): Value-based Engineering: Prinzipien und Motivation für bessere IT-Systeme. Informatik Spektrum, 44(4), 247–256. https://doi.org/10.1007/s00287-021-01378-4.

Thaler, R. H., / Sunstein, C.R. (2009): Nudge: Wie man kluge Entscheidungen anstößt. Remscheid: Ullstein eBooks.

Tunstall, L. / von Werra L. / Wolf, T. (2022): Natural Language Processing with Transformers Revised Edition. Sebastopol: O'Reilly Media, Inc.

Vaswani, A. / Shazeer, N. /Parmar, N. /Uszkoreit, J. /Jones, L. /Gomez, A. N. / Kaiser, Ł. / Polosukhin, I. (2017): Attention is all you need. Advances in Neural Information Processing Systems. 5998–6008.

Vig, J., / Belinkov, Y. (2019): Analyzing the structure of attention in a transformer language model. arXiv preprint arXiv: 1906.04284.

Vollhardt, S. / Schmidt, K. / Kask, S. / Noga, M. (2021). Das intelligente Unternehmen: Effiziente Prozesse mit Künstlicher Intelligenz von SAP – Wie Unternehmen die hohen Erwartungen an die KI erfüllen können. In Buxmann, P. / Schmidt, H. (2021), Künstliche Intelligenz, 2. Aufl. 2021 ed., Springer.

Wallach, W. / Allen, C, / Smit, I. (2008): Machine Morality: Bottom-up and Top-down Approaches for Modeling Human Moral Faculties. AI & Soc., 22, 565–582. https://doi.org/10.1007/s00146-007-0099-0 .

Weiss, J. W. (2021): Business Ethics, Seventh Edition: A Stakeholder and Issues Management Approach. Berrett-Koehler Publishers.

Welzel, H.(1951): Zum Notstandsproblem. ZStW. Zeitschrift für die gesamte Strafrechtswissenschaft 63, 47–56.

Wilson, H. J. & Daugherty, P. R. (2018): Collaborative intelligence: Humans and AI are joining forces. Harvard Business Review, 96(4), 1-11. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces.

Folgende Bände sind bisher in dieser Reihe erschienen:

Band 1 (2021)
Bähren, T. / Braka, D. / Burchard, P. / Cyron, S. / Demary, M. / Dragieva, M. / Eis, L. / Farid, A. T. / Gomes, D. / Hacker, M. / Kaiser, J. / Krüger, R. / Luu, S. / Maasjosthusmann, R. / Marks, A. / Pachocki, C. / Pongratz, M. / Schade, J. C. / Urban, P. / Walter, A. / Winter, V. / Yesilyurt, E. / Buchkremer, R.
Der Einsatz von grünem Tee und anderen Polyphenolen in der Medizin – eine Big-Data-Analyse der medizinischen Fachliteratur
Essen 2021
ISBN (Print) 978-3-89275-119-9 – ISBN (eBook) 978-3-89275-120-5
ISSN (Print) 2699-562X – ISSN (eBook) 2699-5638

Band 2 (2023)
Hamacher, K. / Blach, M. / Kozlik, J. / Muster, F. / Nöllenburg, P.-P. / Ohletz, J.-H. / Franken, G. / Hernes, D. / Hinterding, M. / Höveler, P. / Huppertz, M. / Leppkes, N. / Lopez Rodriguez, A. / Maucy, K. / Petrov, A. / Schäfer, D. / Schneider, R. / Spiegel, B. / Stecker, R. / Steinmann, P. / Tembrink, C. / Titze-Wolter, P. / Vishnyakova, L. / Zimmermann, J. / Buchkremer, R.:
Analyse sensorischer E-Commerce-Elemente mittels Big-Data-Methoden und Künstlicher Intelligenz
Essen 2023
ISBN (Print) 978-3-89275-320-9 – ISBN (eBook) 978-3-89275-321-6
ISSN (Print) 2699-562X – ISSN (eBook) 2699-5638

Forschungsstark und praxisnah

FOM. Die Hochschule besonderen Formats

FOM Hochschulzentrum
Düsseldorf

Mehr als 50.000 Studierende, 25 Forschungseinrichtungen und 500 Veröffentlichungen im Jahr – damit zählt die FOM zu den größten und forschungsstärksten Hochschulen Europas. Initiiert durch die gemeinnützige Stiftung BildungsCentrum der Wirtschaft folgt sie einem klaren Bildungsauftrag: Die FOM ermöglicht Berufstätigen, Auszubildenden, Abiturienten und international Studierenden ein qualitativ hochwertiges und finanziell tragbares Hochschulstudium. Als gemeinnützige Hochschule ist die FOM nicht gewinnorientiert, sondern reinvestiert sämtliche Gewinne – unter anderem in die Lehre und Forschung.

Die FOM ist staatlich anerkannt und bietet mehr als 50 akkreditierte Bachelor- und Master-Studiengänge an – im Campus-Studium an 35 Hochschulzentren oder im einzigartigen Digitalen Live-Studium gesendet aus den Hightech-Studios der FOM.

Lehrende und Studierende forschen an der FOM in einem großen Forschungsbereich aus hochschuleigenen Instituten und KompetenzCentren. Dort werden anwendungsorientierte Lösungen für betriebliche und gesellschaftliche Problemstellungen generiert. Aktuelle Forschungsergebnisse fließen unmittelbar in die Lehre ein und kommen so den Unternehmen und der Wirtschaft insgesamt zugute.

Zudem fördert die FOM grenzüberschreitende Projekte und Partnerschaften im europäischen und internationalen Forschungsraum. Durch Publikationen, über Fachtagungen, wissenschaftliche Konferenzen und Vortragsaktivitäten wird der Transfer der Forschungs- und Entwicklungsergebnisse in Wissenschaft und Wirtschaft sichergestellt.

Alle Institute und KompetenzCentren unter
fom.de/forschung

ISBN (Print) 978-3-89275-394-0 ISSN (Print) 2699-562X
ISBN (eBook) 978-3-89275-395-7 ISSN (eBook) 2699-5638

Institut für IT-Management & Digitalisierung
der FOM University of Applied Sciences

FOM Hochschule

Mit rund 50.000 Studierenden ist die FOM eine der größten Hochschulen Europas und führt seit 1993 Studiengänge für Berufstätige durch, die einen staatlich und international anerkannten Hochschulabschluss (Bachelor/Master) erlangen wollen.

Die FOM ist der anwendungsorientierten Forschung verpflichtet und verfolgt das Ziel, adaptionsfähige Lösungen für betriebliche bzw. wirtschaftsnahe oder gesellschaftliche Problemstellungen zu generieren. Dabei spielt die Verzahnung von Forschung und Lehre eine große Rolle: Kongruent zu den Masterprogrammen sind Institute und KompetenzCentren gegründet worden. Sie geben der Hochschule ein fachliches Profil und eröffnen sowohl Wissenschaftlerinnen und Wissenschaftlern als auch engagierten Studierenden die Gelegenheit, sich aktiv in den Forschungsdiskurs einzubringen.

Weitere Informationen finden Sie unter **fom.de**

ifid

Das ifid Institut für IT-Management & Digitalisierung bündelt Kompetenzen in den Forschungsbereichen Künstliche Intelligenz (KI), Systemwissenschaften, IT-Management und digitale Transformation.

Die Aufgaben des Instituts umfassen Forschung und Entwicklung, Wissenstransfer und Innovationsförderung an der Schnittstelle von Wissenschaft und Praxis. Auch der Transfer von Forschungserkenntnissen in die Lehre spielt eine große Rolle.
Um diese Aufgaben zu erfüllen, setzt die Forschergruppe auf den Einsatz modernster Big Data-Architekturen und KI-Analysesysteme. Es bestehen Kooperationen mit den großen Technologie-Unternehmen und Instituten der Branche.
Die Wissenschaftlerinnen und Wissenschaftler beschäftigen sich insbesondere mit folgenden Feldern:

- Künstliche Intelligenz / Machine Learning / Data Science / Big Data
- Natural Language Processing (NLP)
- Enterprise Architekturen (insbesondere Big Data)
- Einsatz von Blockchain-Technologien
- Digitalisierung von Prozessen
- Integration der Forschung in die Lehre

Weitere Informationen finden Sie unter **fom-ifid.de**