The research presented in this paper was developed within the project “SkillExtract” (HA-Projekt-Nr.: 628/18-51), financed with funds of LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz, Förderlinie 3: KMU-Verbundvorhaben (State Offensive for the Development of Scientific and Economic Excellence).

Please quote as: Dellermann, D.; Calma, A.; Lipusch, N.; Weber, T.; Weigel, S. & Ebel, P. (2019):The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. In: Hawaii International Conference on System Sciences (HICSS). Hawaii, USA.

The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems

Dominik Dellermann
vencortex
dominik.dellermann@vencortex.com

Adrian Calma
vencortex
adrian.calma@vencortex.com

Nikolaus Lipusch
Kassel University
lipusch@uni-kassel.de

Thorsten Weber
Kassel University
weber@uni-kassel.de

Sascha Weigel
Kassel University
weigel@uni-kassel.de

Philipp Ebel
University of St. Gallen
ebel@unisg.ch

Abstract

Recent technological advances, especially in the field of machine learning, provide astonishing progress on the road towards artificial general intelligence. However, tasks in current real-world business applications cannot yet be solved by machines alone. We, therefore, identify the need for developing socio-technological ensembles of humans and machines. Such systems possess the ability to accomplish complex goals by combining human and artificial intelligence to collectively achieve superior results and continuously improve by learning from each other. Thus, the need for structured design knowledge for those systems arises. Following a taxonomy development method, this article provides three main contributions: First, we present a structured overview of interdisciplinary research on the role of humans in the machine learning pipeline. Second, we envision hybrid intelligence systems and conceptualize the relevant dimensions for system design for the first time. Finally, we offer useful guidance for system developers during the implementation of such applications.

1. Introduction

Recent technological advances especially in the field of deep learning provide astonishing progress on the road towards artificial general intelligence (AGI) [1, 2]. Artificial intelligence (AI) are progressively achieving (super-) human level performance in various tasks such as autonomous driving [3], cancer detection [4], or playing complex games [5, 6]. Therefore, more and more business applications that are based on AI technologies arise. Both research and practice are wondering when AI will be able to solve complex tasks in real-world business applications apart from laboratory settings in research. However, those advances provide a rather one-sided picture on AI, denying the fact that although AI is capable to solve certain tasks with quite impressive performance, AGI is far away from being achieved. There are lots of problems that machines can not yet solve alone [7], such as applying expertise to decision making, planning, or creative tasks just to name a few. In particular, machine learning systems in the wild have major difficulties with being adaptive to dynamic environments and self adjusting [8], lack of what humans call common sense. This makes them highly vulnerable for adversarial examples [9]. Moreover, AGI needs massive amounts of training data compared to humans, who can learn from only few examples [10], and fails to work with certain data types (e.g. soft data). Nevertheless, a lack of control of the learning process might lead to unintended consequences (e.g. racism biases) and limit interpretability, which is crucial for critical domains such as medicine [11]. Therefore, humans are still required at various positions in the loop of the machine learning process. While lot of work has been done in creating training sets with human labelers, more recent research point towards end user involvement [12] and teaching of such machines [5], thus, combining humans and machines in ***hybrid intelligence*** systems. The main idea of hybrid intelligence systems is, thus, that socio-technical ensembles and its human and AI parts can co-evolve to improve over time. Therefore, the following central questions are arise: *Which and how should certain design decisions be made for implementing such systems?* The purpose of this paper is to point towards such hybrid intelligence systems. Thereby, we aim at conceptualizing the idea of ***hybrid intelligence*** systems and provide an initial taxonomy of design knowledge for developing such socio-technical ensembles. By following a taxonomy development method [13], we reviewed various literature in interdisciplinary fields and combine those findings with empirical examination of practical business applications in the context of

hybrid intelligence. The contribution of this paper is threefold. First, we provide a structured overview of interdisciplinary research on the role of humans in the machine learning pipeline. Second, we offer an initial conceptualization of the term ***hybrid intelligence*** systems and relevant dimensions for system design. Third, we intend to provide useful guidance for system developers during the implementation of hybrid intelligence systems in real-world applications. Towards this end, we propose an initial taxonomy of hybrid intelligence systems.

2. Related Work

2.1. Machine Learning and AI

The subfield of intelligence that relates to machines is called **artificial intelligence** (**AI**). By this term we mean systems that perform *"[] activities that we associate with human thinking, activities such as decision-making, problem solving, learning []"* [14]. Although, various definitions exist for AI, this term generally covers the idea of creating machines that can accomplish complex goals. This includes facets such as natural language processing, perceiving objects, storing of knowledge and applying it for solving problems, and machine learning to adapt to new circumstances and act in its environment [15].

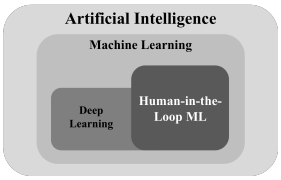

Figure 1. Machine learning and AI.

A subset of techniques that is required to achieve AI is **machine learning (ML)**. Mitchell [16] defines this as: *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."*

A popular approach that drives current progress in both paradigms is **deep learning** [9]. Deep-learning constitutes a representation-learning method that includes multiple levels of representation, obtained by combining simpler but non-linear models. Each of those models transforms the representation of one level (starting with the input data) into a representation at more abstract level [17]. Deep learning is a special machine learning technique.

Finally, **human-in-the-loop learning** describes machine learning approaches (both deep and other) that use the human in some part of the pipeline. Such approaches are in contrast to research on most knowledge-base systems in IS that use rather static knowledge repositories. We will focus on this in the following chapter.

2.2. The Role of Humans-in-the-Loop of Machine Learning

Although, the terms of AI and machine learning give the impression that humans become to some extent obsolete, the machine learning pipeline still requires lot of human interaction such as for feature engineering, parameter tuning, or training. While deep learning has decreased the effort for manual feature engineering and some automation approaches (e.g. AutoML [18]) support human experts in tuning models, the human is still heavily in the loop for sense-making and training. For instance, unsupervised learning requires humans to make sense of clusters that are identified as patterns in data to create knowledge [19]. More obviously, human input is required to train models in supervised machine learning approaches, especially for creating training data, debug models, or train algorithms such as in reinforcement learning [5]. This is especially relevant when divergences of real-life and machine learning problem formulations emerge. This is for instance the case when static (offline) training datasets are not perfectly representative of realist and dynamic environments [20]. Moreover, human input is crucial when models need to learn from human preferences (e.g. recommender systems) and adapt to users or when security concerns require both control and interpretability of the learning process and the output [11]. Therefore, more recent research has focused on interactive forms of learning (e.g. [21, 12] and machine teaching (e.g. [22]). Those approaches make active use of human input (e.g. active learning [23]) and thus learn from human intelligence. This allows machines to learn tasks that they can not yet achieve alone [7], adapt to environmental dynamics, and deal with unknown situations [24].

2.3. Hybrid Intelligence

Rather then using the human just in certain parts and time during the process of creating machine learning models, applications that are able to deal with real-world problems require a continuously collaborating socio-technological ensemble integrating humans and machines, which is contrast to previous research on decision support and expert systems [21, 25].

Therefore, we argue that the most likely paradigm

for the division of labor between humans and machines in the next years, or probably decades, is ***hybrid intelligence***. This concept aims at using the complementary strengths of human intelligence and AI to behave more intelligently than each of the two could be in separation (e.g. [7]). The basic rational is to try to combine the complementary strengths of heterogeneous intelligences (i.e., human and artificial agents) into a socio-technological ensemble. We envision ***hybrid intelligence*** systems, which are defined as systems that have *the ability to accomplish complex goals by combining human and artificial intelligence to collectively achieve superior results than each of the could have done in separation and continuously improve by learning from each other.*

Collectively: means that tasks are performed collectively. This means that the activities conducted by each part are dependent, however, are not necessarily always aligned to achieve a common goal (e.g. teaching an AI adversarial tasks such as playing games).

Superior results: defines that the system achieves a performance that none of the involved actors could have achieved in a certain task without the other. The goal is, therefore, to make the outcome (e.g. a prediction) both more efficient and effective on the level of the whole socio-technical system by achieving goals that could not have been achieved before.

Continuous learning: describes that over time this socio-technological system improves both as a whole and each single component (i.e. humans and machines) learn through experience from each other, thus improving performance in a certain task. The performance of such systems can be thus not only measured by the superior outcome of the whole system but also by the learning of the human and machine agents that are parts of the socio-technical system.

The idea of hybrid intelligence systems is thus that socio-technical ensembles and its human and AI parts can co-evolve to improve over time. The central questions are, therefore, which and how certain design decisions should be made for implementing such hybrid systems rather than focusing.

3. Methodology

3.1. Taxonomy Development Method

For developing our proposed taxonomy, we followed the methodological procedures of Nickerson et al. [13]. In general, a taxonomy is defined as a *"fundamental mechanism for organizing knowledge"* and the term is considered as a synonym to *"classification"* and *"typology"* [13]. The method follows an iterative process consisting of the following steps: 1) defining a meta-characteristic; 2) determining stopping conditions; 3) selecting an empirical-to-conceptual or conceptual-to-empirical approach; and 4) iteratively following this approach, until the stopping conditions are met (see Figure 3).

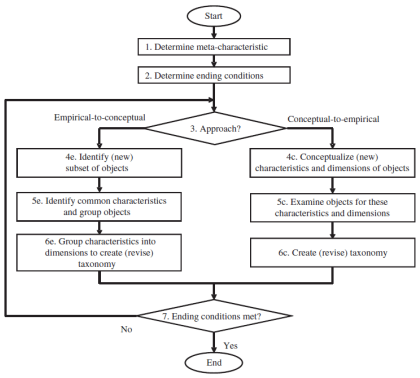

Figure 2. Taxonomy development method [13].

The process of the taxonomy development starts with defining a set of meta-characteristic. This step limits the odds of *naive empiricism* where a large number of characteristics are defined in search for random pattern, and reflects the expected application of the taxonomy [13]. For this purpose, we define those meta-characteristic as generic design dimensions that are required for developing hybrid intelligence systems. Based on our classification from literature, we choose four dimensions: task characteristics, learning paradigm, human-AI interaction, and AI-human interaction. In the second step, we selected both objective and subjective conditions to conclude the iterative process. The following conditions, adapted from Nickerson et al. [13], were selected:

We applied the following **objective conditions**: 1) All papers from the sample of the literature review and empirical cases are examined. 2) Then, at least one object is classified under every characteristic of every dimension. 3) While performing the last iteration, no new dimension or characteristics are added. 4) We treated every dimension as unique. 5) Lastly, every characteristic is unique within its dimension.

The following **subjective conditions** were considered: conciseness, robustness, comprehensiveness, extensibility, explanatory, and information availability. We included no unnecessary dimension or characteristic (conciseness), whereas there are enough dimensions and characteristics to differentiate (robustness). At this point, all design decisions can be classified in the taxonomy (comprehensiveness), while still allowing for new

Table 1. Empirical evidence from business applications of hybrid intelligence systems.

Application	Domain	Reference
Teachable Machine	Image Recognition	[26]
Cindicator	Asset Management	[27]
vencortex	Startup Financing	[28]
Cobi	Conference Sheduling	[29]
Stitch Fix	Fashion	[30]
Alpha Go	Games	[31]
Custom Decision Service	General	[32]
TOTAL	7	

dimensions and characteristics to be subsequently added (extensible). Furthermore, the information is valuable for guiding hybrid intelligence systems design decisions (explanatory) and is typically available or easily interpretable (information availability).

We conducted a total of three iterations so far. The first iteration used a conceptual-to-empirical approach, where we used extant theoretical knowledge from literature in various fields such as computer science, HCI, information systems, and neuro science to guide the initial dimensions and characteristics of the taxonomy. Based on the identified dimensions of ***hybrid intelligence*** systems, we sampled seven real-world applications that make use of human and AI combinations. The second iteration used the empirical-to conceptual approach focuses on creating characteristics and dimensions based on the identification of common characteristics from a sample of AI applications in practice. The third iteration then used the conceptual-to-empirical approach, based on an extended literature review including newly identified search termini.

3.2. Data Sources and Sample

Literature review: For conducting our literature review, we followed the methodological procedures of [33, 34]. The literature search was conducted from April to June 2018. A prior informal literature search revealed keywords for the database searches resulting in the search string (”hybrid intelligence” OR ”human-in-the-loop” OR ”interactive machine learning” OR ”machine teaching” OR ”machine learning AND crowdsourcing” OR ”human supervision” OR ”human understandable machine learning” OR ”human concept learning”). During this initial phase we decided to exclude research on knowledge-base systems such as expert systems or decision support systems in IS [35, 25], as the studies either do not focus on the continuous learning of the knowledge repository or do not use machine learning techniques at all. Moreover, the purpose of this study is to identify and classify relevant (socio-) technical design knowledge for hybrid intelligence systems, which is also not included in those studies. The database search was constrained to title, abstract, keywords and not limited to a certain publication. Databases include AISeL, IEEE Xplore, ACM DL, AAAI DL, and arXiv to identify relevant interdisciplinary literature from the fields of IS, HCI, bio-informatics, and computer science. The search resulted in a total of 2505 hits. Titles, abstracts and keywords were screened for potential fit to the purpose of our study. Screening was conducted by three researchers independently and resulted in 85 articles that were reviewed in detail so far. A backward and forward search ensured the extensiveness of our results. Table 1 lists the number of search results after the review phases.

Empirical cases: To extend our findings from literature and provide empirical evidence (cf. Table 1) from recent (business) applications of ***hybrid intelligence*** systems, we include an initial set of seven empirical applications that was analyzed for enhancing our taxonomy.

4. Taxonomy of Design Knowledge on hybrid intelligence Systems

Our taxonomy of hybrid intelligence systems is organized along the four meta-dimensions *task characteristics, learning paradigm, human-AI interaction*, and *AI-human interaction*. Moreover, we identified 16 sub-dimensions and a total of 50 categories for the proposed taxonomy. For organizing the dimensions of the taxonomy we followed a hierarchical approach following the sequence of the design decisions that are necessary to develop such systems.

4.1. Task Characteristics

The goal of hybrid intelligence is to create superior results through a collaboration between humans

and machines. The central component that drives design decisions for ***hybrid intelligence*** systems is the task, that humans and machines solve collaboratively. Task characteristics focus on how the task itself is carried out [36]. In context of ***hybrid intelligence*** systems, we identify the following four important tasks characteristics.

Type of Task: The task to be solved is the first dimension that has to be defined for developing ***hybrid intelligence*** systems. In this context, we identified four generic categories of tasks: *recognition, prediction, reasoning* and *action.* First, *recognition* defines tasks that recognize for instance objects [17], images [37], or natural language [38]. On an application level such tasks are used for autonomous driving (e.g. [3]) or smart assistants such as Alexa, Siri or Duplex. Second, *prediction* tasks aim at predicting future events based on previous data such as stock prices or market dynamics [39]. The third type of task, *reasoning*, focuses on understanding data by for instance inductively building (mental) models of a certain phenomenon and therefore make it possible to solve complex problems with small amount of data [10]. Finally, *action* tasks are characterized as such that require an agent (human or machine) to conduct a certain kind of action [40].

Goals: The two involved agents, the human and the AI, may have a *common* "goal" like solving a problem through the combination of the knowledge and abilities of both. An example for such common goals are recommender systems (e.g. Netflix [41]), which learn a user's decision model to offer suggestions. In other contexts, the agents goals also be *adversarial*. For instance, in settings where AIs tzry to beat human in games such as IBMs Watson in the game of Jeopardy! [42]. In many other cases the goal of the human and the AI may also be *independent* for example when humans train image classifiers without being involved in the end solution.

Shared Data Representation: The shared data representation is what is the data that is shown to both the human and the machine before executing their tasks. The data can be represented in different levels of granularity and abstraction to create a shared understanding between humans and machines [22, 43]. *Features* describe phenomenas in different kinds of dimensions like height and weight of a human being. *Instances* are examples of a phenomena which are specified by features. *Concepts* on the other hand are multiple instances that belong to one common theme, e.g. pictures of different humans. *Schemas* finally illustrate relations between different concepts [44].

Timing in Machine Learning Pipeline: The last sub-dimension describes the timing in the machine

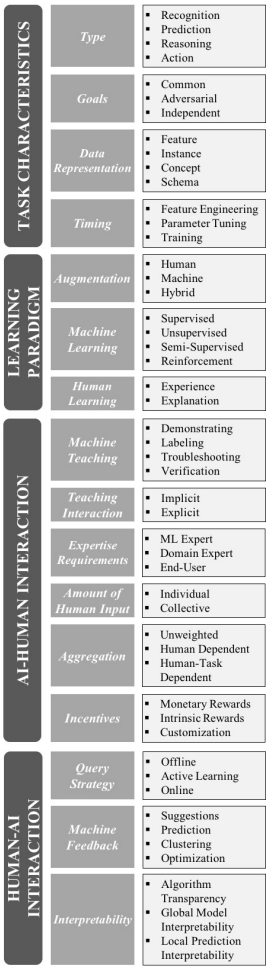

Figure 3. Taxonomy of hybrid intelligence design.

learning pipeline that focuses on hybrid intelligence. For this dimension we identified three characteristics: *feature engineering*, *parameter tuning*, and *training*. First, *feature engineering* allows the integration of domain knowledge in machine learning models. While more recent advances make it possible to fully automatically (i.e. machine only) learn features through deep learning, human input can be combined for creating and enlarging features such in the case of artist identification on images and quality classification of Wikipedia articles (e.g. [45]). Second, *parameter tuning* is applied to optimize models. Here machine learning experts typically use their deep understanding of statistical models to tune hyper-parameters or select models. Such human only parameter tuning can be augmented with approaches such as AutoML [18] or neural architecture search [46, 47] automate the design of machine learning models,thus, making it much more accessible for non-experts. Finally, human input is crucial for *training* machine learning models in many

domains. For instance large dataset such as ImageNet or the lung cancer dataset LUNA16 rely on human annotations. Moreover, recommender systems heavily rely on input of human usage behavior to adapt to specific preferences (e.g. [12]) and robotic applications are trained by human examples [40].

4.2. Learning Paradigm

Augmentation: In general, ***hybrid intelligence*** systems allow three different forms of augmentation: *human*, *machine*, and *hybrid* augmentation. The augmentation of *human* intelligence is focused on typically applications that enable humans to solve tasks through the predictions of an algorithm such as in financial forecasting or solving complex problems [48]. Contrary, most research in the field of machine learning focuses on leveraging human input for training to augment *machines* for solving tasks that they cannot yet solve alone [7]. Finally, more recent work identified the great potential for simultaneously augmenting both at the same time through *hybrid* augmentation [49, 50] or the example of Alpha Go that started by learning from human game moves (i.e. *machine* augmentation) and finally offered hybrid augmentation by inventing creative moves that taught even mature players novel strategies [6, 51].

Machine Learning Paradigm: The machine learning paradigm that is applied in hybrid intelligence systems can be categorized into four relevant subfields: *supervised, unsupervised, semi-supervised*, and *reinforcement* learning [52]. In *supervised learning*, the goal is to learn a function that maps the input data x to a certain output data y, given a labeled set of input-output pairs. In *unsupervised learning*, such output y does not exist and the learner tries to identify pattern in the input data x [16]. Further forms of learning such as reinforcement learning or semi-supervised learning can be subsumed under those two paradigms. *Semi-supervised learning* describes a combination of both paradigms, which uses both a small set of labeled and a large set of unlabeled data to solve a certain task [53]. Finally, *reinforcement learning*. An agent interacts with an environment thereby learning to solve a problem through receiving rewards and punishment for a certain action [5, 6].

Human Learning Paradigm: Humans have a mental model of their environment, which gets updated through events. This update is done by finding an explanation for the event [50, 49, 10]. Human learning can therefore can be achieved from *experience* and comparison with previous experiences [54, 44] and from description and *explanations* [55].

4.3. Human-AI Interaction

Machine Teaching: defines how humans provide input. First, humans can demonstrate actions that the machine learns to imitate [40]. Second, humans can annotate data for training a model for instance through crowdsourcing [56, 57]. We designate that as a *labeling*. Third, human intelligence can be used to actively identify a misspecification of the learner and debug the model, which we define as *troubleshooting* [58, 24]. Moreover, human teaching can take the form of *verification* whereby humans verify or falsify machine output [59].

Teaching Interaction: The input provided through human teaching, can be both *explicit* and *implicit*. While *explicit* teaching leverages active input of the user such as for instance labeling tasks such as image or text annotation [60], *implicit* teaching learns from observing the actions of the user and thus adapts to their demands. For instance, Microsoft uses contextual bandit algorithms to suggest users certain content, using the actions of the user as implicit teaching interaction.

Expertise Requirements: Hybrid intelligence systems can have certain requirements for the expertise of humans that provides input for systems. While by now both most research and practical applications focus on human input from an *ML expert* [61, 62, 63, 24], thus, requiring deep expertise in the field of AI. Moreover, *end users* can provide the system with input for product recommendations and e-commerce or input from human non-experts accessed through crowd work platforms [64, 58, 65]. More recent endeavors, however, focus on the integration of *domain experts* in hybrid intelligence architectures that leverage the profound understanding of the semantics of a problem domain to teach a machine, while not requiring any ML expertise [22, 66, 67].

Amount of Human Input: The amount of human input can vary between those of individual humans and aggregated input from several humans. *Individual* human input is for instance applied in recommender systems for individualization or due to cost efficiency reasons [60]. On the other hand, *collective* human input combines the input of several individual humans by leveraging mechanisms of human computation (e.g. [68, 66, 67]). This approach allows to reduce errors and biases of individual humans and the aggregation of heterogeneous knowledge [45, 69, 65].

Aggregation: When human input is aggregated from a collective of individual humans, different aggregation mechanisms can be leveraged to maximize the quality of teaching. First, *unweighted* methods can be used that use averaging or majority voting to

aggregate results (e.g. [60]). Additionally, aggregation can be achieved by modeling the context of teaching through algorithmic approach such as expectation maximization, graphical models, entropy minimization, or discriminative training. Therefore, the aggregation can be *human dependent* focusing on the characteristics of a the individual human [70, 71, 72], or *human-task dependent* adjusting to the teaching task [73, 72, 74].

Incentives: Humans need to be incentivized to provide input in hybrid intelligence systems. Incentives can be *monetary rewards* such in the case of crowd work on platforms (e.g. Amazon Mechanical Turk), *intrinsic rewards* such as intellectual exchange in citizen science [75], fun in games with a purpose [76] learning [77]. Another incentive for human input is *customization*, which allows to increase individualized service quality for users that provide a higher amount of input to the learner [12, 78].

4.4. AI-Human Interaction

This sub-dimension describes the machine part of the Interaction, the AI-human interaction. At first, which query strategy the algorithm used to learn. Second, we describe the feedback of the machine to humans. Third, we carry out a short explanation of interpretability to show the influence for hybrid intelligence.

Query Strategy: *Offline* query strategies require the human to finish her task completely before her actions are applied as input to the AI (e.g. [79, 80]). Handling a typical labeling task the human would first need to go through all the data and label each instance. Afterwards the labeled Data is fed to an machine learning algorithm to train a model. In contrast, *online* query strategies let the human complete subtasks whose are directly fed to an algorithm, so that teaching and learning can be executed almost simultaneously [64, 58, 72]. Another possibility is the use of *active learning* query strategies [81, 23]. In this case, the human is queried by the machine when more input to give an accurate prediction is required.

Machine Feedback: Those four categories describe the feedback that humans receive from the machine. First, humans can get direct *suggestions* from the machine, which makes explicit recommendations to the user on how to act. For instance recommender systems such as Netflix or Spotify provide such suggestions for users. Furthermore, systems can make suggestions for describing images [58]. *Predictions* as machine feedback can support humans e.g. to detect lies [45], predict worker behaviors [72], or classify images. Thereby, this form of feedback provides a probabilistic value of a certain outcome (e.g. probability of some data x belonging to a certain class y). The third form of machine feedback is *clustering* data. Thereby, machines compare data points and put them in an order for instance to prioritize items [82], or organize data among identified pattern. Furthermore, another possibility of machine feedback is *optimization*. Machines enhance humans for instance in making more consistent decisions by optimizing their strategy [83].

Interpretability: For AI-Human interaction in ***hybrid intelligence*** systems interpretability is crucial to prevent biases (e.g. racism), achieve reliability and robustness, ensure causality of the learning, debugging the learner if necessary and for creating trust especially in the context of AI safety [11]. Interpretability in ***hybrid intelligence*** systems can be achieved through *algorithm transparency*, that allows to open the black box of an algorithm itself, *global model interpretability* that focuses on the general interpretability of a machine learning model, and *local prediction interpretability* that tries to make more complex models interpretable for a single prediction [84, 11].

5. Discussion

Our proposed taxonomy for ***hybrid intelligence*** systems extracts interdisciplinary knowledge on human-in-the-loop mechanisms in ML and proposes initial descriptive design knowledge for the development of such systems that might guide developers. Our findings reveal the manifold applications, mechanisms, and benefits of hybrid systems that might probably become of increasing interest in real-world applications in the future. In particular, our taxonomy of design knowledge offers insights on how to leverage the advantages of combining human and machine intelligence. For instance, this allows to integrate deep domain insights into machine learning algorithms, continuously adapt a learner to dynamic problems, and enhance trust through interpretability and human control. Vice versa, this approach offers the advantage of improving humans in solving problems by offering feedback on how the task was conducted or the performance of a human during that task and machine feedback to augment human intelligence. Moreover, we assume that the design of such systems might allow to move beyond sole efficiency of solving tasks to combined socio-technical ensembles that can achieve superior results that could no man or machine have achieved so far. Promising fields for such systems are in the field of medicine, science, innovation and creativity.

6. Conclusion

Within this paper we propose a taxonomy for design knowledge for ***hybrid intelligence*** systems, which presents descriptive knowledge structured along the four meta-dimensions *task characteristics, learning paradigm, human-AI interaction*, and *AI-human interaction*. Moreover, we identified 16 sub-dimensions and a total of 50 categories for the proposed taxonomy. By following a taxonomy development methodology [13], we extracted interdisciplinary knowledge on human-in-the-loop approaches in machine learning and the interaction between human and AI. We extended those findings with an examination of seven empirical applications of ***hybrid intelligence*** systems.

Therefore, our contribution is threefold. First, the proposed taxonomy provides a structured overview of interdisciplinary research on the role of humans in the machine learning pipeline by reviewing interdisciplinary research and extract relevant knowledge for system design. Second, we offer an initial conceptualization of the term ***hybrid intelligence*** systems and relevant dimensions for developing applications. Third, we intend to provide useful guidance for system developers during the implementation of hybrid intelligence systems in real-world applications.

Obviously this paper is not without limitations and provides a first step towards a comprehensive taxonomy of design knowledge on ***hybrid intelligence*** systems. First, further research should extend the scope of this research to more practical applications in various domains. By now our empirical case selection is slightly biased on decision problem contexts.Second, as we proceed our research we will further condensate the identified characteristics by aggregating potentially overlapping dimensions in subsequent iterations. Third, our results are overly descriptive so far. As we proceed our research we will therefore focus on providing prescriptive knowledge on what characteristics to choose in a certain situation and thereby propose more specific guidance for developers of ***hybrid intelligence*** systems that combine human and machine intelligence to achieve superior goals and driving the future progress of AI. For this purpose, we will identify interdependencies between dimensions and sub-dimensions and evaluate the usefulness of our artifact for designing real-world applications. Finally, further research might focus on integrating the overly design oriented knowledge of this study with research on knowledge-base systems in IS to discuss the findings in the context of those class of systems.

References

[1] B. Goertzel and C. Pennachin, *Artificial general intelligence*, vol. 2. Springer, 2007.

[2] R. Kurzweil, *The singularity is near*. Gerald Duckworth & Co, 2010.

[3] L. Fridman, B. Reimer, B. Mehler, and W. T. Freeman, "Cognitive Load Estimation in the Wild," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, (New York, USA), pp. 1–9, 2018.

[4] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[5] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[6] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. Van Den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of Go without human knowledge," *nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[7] E. Kamar, "Directions in hybrid intelligence: Complementing AI systems with human intelligence," in *IJCAI International Joint Conference on Artificial Intelligence*, 2016.

[8] C. Müller-Schloer and S. Tomforde, *Organic Computing Technical Systems for Survival in the Real World*. Autonomic Systems, Cham: Springer International Publishing, 2017.

[9] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," *arXiv preprint arXiv:1607.02533*, 2016.

[10] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behavioral and Brain Sciences*, vol. 40, 2017.

[11] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

[12] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, "Power to the people: The role of humans in interactive machine learning," *AI Magazine*, vol. 35, no. 4, pp. 105–120, 2014.

[13] R. C. Nickerson, U. Varshney, and J. Muntermann, "A method for taxonomy development and its application in information systems," *European Journal of Information Systems*, vol. 22, no. 3, pp. 336–359, 2013.

[14] R. Bellman, *An introduction to artificial intelligence : can computers think?* Boyd & Fraser Pub. Co, 1978.

[15] S. J. Russell and P. Norvig, "Artificial intelligence: a modern approach," 2016.

[16] T. M. Mitchell *et al.*, "Machine learning. 1997," *Burr Ridge, IL: McGraw Hill*, vol. 45, no. 37, pp. 870–877, 1997.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[18] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter, "Efficient and robust automated machine learning," *Advances in Neural Information Processing Systems*, pp. 2962–2970, 2015.

[19] R. G. Gomes, P. Welinder, A. Krause, and P. Perona, "Crowdclustering," *Advances in neural information processing systems*, pp. 558–566, 2011.

[20] J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan, "Human decisions and machine predictions," *The quarterly journal of economics*, vol. 133, no. 1, pp. 237–293, 2017.

[21] A. Holzinger, "Interactive machine learning for health informatics: when do we need the human-in-the-loop?," *Brain Informatics*, vol. 3, no. 2, pp. 119–131, 2016.

[22] P. Y. Simard, S. Amershi, D. M. Chickering, A. E. Pelton, S. Ghorashi, C. Meek, G. Ramos, J. Suh, J. Verwey, M. Wang, *et al.*, "Machine teaching: A new paradigm for building machine learning systems," *arXiv preprint arXiv:1707.06742*, 2017.

[23] B. Settles, "Active Learning Literature Survey," *Machine Learning*, 2010.

[24] J. Attenberg, P. Ipeirotis, and F. Provost, "Beat the Machine," *Journal of Data and Information Quality*, vol. 6, no. 1, pp. 1–17, 2015.

[25] S. Gregor, "Explanations from knowledge-based systems and cooperative problem solving: an empirical study," *International Journal of Human-Computer Studies*, vol. 54, no. 1, pp. 81–105, 2001.

[26] "Teachable Machine." `https://teachablemachine.withgoogle.com` last access: 2018-06-14.

[27] "Cindicator-Hybrid Intelligence for Effective Asset Management — Cindicator — Cindicator." `https://cindicator.com` last access: 2018-06-14.

[28] "vencortex." `https://www.vencortex.com` last access: 2018-06-14.

[29] "Cobi Communitysourcing Large-Scale Conference Scheduling." `http://projectcobi.com` last access: 2018-06-14.

[30] "Your Online Personal Stylist — Stitch Fix." `https://www.stitchfix.com` last access: 2018-06-14.

[31] "AlphaGo DeepMind." `https://deepmind.com/research/alphago` last access: 2018-06-14.

[32] "Microsoft Custom Decision Service." `https://portal.ds.microsoft.com` last access: 2018-06-14.

[33] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly*, vol. 26, no. 2, pp. 8–23, 2002.

[34] J. vom Brocke, A. Simons, B. Niehaves, K. Riemer, R. Plattfaut, A. Cleven, J. V. Brocke, and K. Reimer, "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," *17th European Conference on Information Systems*, 2009.

[35] U. Kayande, A. De Bruyn, G. L. Lilien, A. Rangaswamy, and G. H. Van Bruggen, "How incorporating feedback mechanisms in a dss affects dss evaluations," *Information Systems Research*, vol. 20, no. 4, pp. 527–546, 2009.

[36] W. M. Reynolds, G. E. Miller, and I. B. Weiner, *Handbook of psychology*. Wiley, 2013.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[38] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[39] P. Choudhary, A. Jain, and R. Baijal, "Unravelling airbnb predicting price for new listing," *arXiv preprint arXiv:1805.12101*, 2018.

[40] R. Mao, J. S. Baras, Y. Yang, and C. Fermuller, "Co-active Learning to Adapt Humanoid Movement for Manipulation," sep 2016.

[41] C. A. Gomez-Uribe and N. Hunt, "The netflix recommender system: Algorithms, business value, and innovation," *ACM Transactions on Management Information Systems (TMIS)*, vol. 6, no. 4, p. 13, 2016.

[42] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, and C. Welty, "Building Watson: An Overview of the DeepQA Project," *AI Magazine*, vol. 31, p. 59, jul 2010.

[43] S. Feldman, "Enforcing social conformity: A theory of authoritarianism," 2003.

[44] D. Gentner and L. Smith, "Analogical Reasoning," in *Encyclopedia of Human Behavior*, 2012.

[45] J. Cheng, J. Teevan, S. T. Iqbal, and M. S. Bernstein, "Break It Down: A Comparison of Macro- and Microtasks," *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pp. 4061–4064, 2015.

[46] E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. Le, and A. Kurakin, "Large-scale evolution of image classifiers," *arXiv preprint arXiv:1703.01041*, 2017.

[47] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," *arXiv preprint arXiv:1802.01548*, 2018.

[48] S. Doroudi, E. Kamar, E. Brunskill, and E. Horvitz, "Toward a Learning Science for Complex Crowdsourcing Tasks," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, (New York, USA), pp. 2623–2634, ACM Press, 2016.

[49] S. Milli, P. Abbeel, and I. Mordatch, "Interpretable and pedagogical examples," *arXiv preprint arXiv:1711.00694*, 2017.

[50] S. Carter and M. Nielsen, "Using Artificial Intelligence to Augment Human Intelligence," *Distill*, vol. 2, no. 12, 2017.

[51] L. Baker and F. Hui, "Innovations of AlphaGo." `https://deepmind.com/blog/innovations-alphago/` last access: 2018-06-14.

[52] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[53] X. Zhu, "Semi-supervised learning literature survey," *Computer Science, University of Wisconsin-Madison*, vol. 2, no. 3, pp. 4–44, 2006.

[54] B. Kim, C. Rudin, and J. A. Shah, "The bayesian case model: A generative approach for case-based reasoning and prototype classification," 2014.

[55] R. M. Hogarth, "On the learning of intuition," *Intuition in judgment and decision making*, pp. 111–126, 2011.

[56] R. Snow, B. O'Connor, D. Jurafsky, and A. Y. Ng, "Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks," *Proceedings of the conference on empirical methods in natural language processing*, pp. 254–263, 2008.

[57] V. C. Raykar, S. Yu, L. H. Zhao, G. Hermosillo Valadez, C. Florin, L. Bogoni, L. Moy, and L. M. Org, "Learning From Crowds," *Journal of Machine Learning Research*, vol. 11, pp. 1297–1322, 2010.

[58] B. Nushi, E. Kamar, E. Horvitz, and D. Kossmann, "On human intellect and machine failures: Troubleshooting integrative machine learning systems.," *AAAI*, pp. 1017–1025, 2017.

[59] K. Pei, Y. Cao, J. Yang, and S. Jana, "Towards practical verification of machine learning: The case of computer vision systems," *arXiv preprint arXiv:1712.01785*, 2017.

[60] G. Li, Y. Zheng, J. Fan, J. Wang, and R. Cheng, "Crowdsourced data management: Overview and challenges," in *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 1711–1716, ACM, 2017.

[61] A. Chakarov, A. Nori, S. Rajamani, S. Sen, and D. Vijaykeerthy, "Debugging machine learning tasks," *arXiv preprint arXiv:1603.07292*, 2016.

[62] T. Kulesza, S. Stumpf, M. Burnett, W.-K. Wong, Y. Riche, T. Moore, I. Oberst, A. Shinsel, and K. McIntosh, "Explanatory Debugging: Supporting End-User Debugging of Machine-Learned Programs," in *2010 IEEE Symposium on Visual Languages and Human-Centric Computing*, pp. 41–48, IEEE, sep 2010.

[63] K. Patel, N. Bancroft, S. M. Drucker, J. Fogarty, A. J. Ko, and J. Landay, "Gestalt: integrated support for implementation and analysis in machine learning," *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*, pp. 37–46, 2010.

[64] J. C. Chang, S. Amershi, and E. Kamar, "Revolt," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 2017.

[65] J. C. Chang, A. Kittur, and N. Hahn, "Alloy: Clustering with Crowds and Computation," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2016.

[66] D. Dellermann, N. Lipusch, and P. Ebel, "Developing Design Principles for a Crowd-Based Business Model Validation System," in *International Conference on Design Science Research in Information Systems*, pp. 163–178, 2017.

[67] D. Dellermann, N. Lipusch, P. Ebel, K. M. Popp, and J. M. Leimeister, "Finding the Unicorn : Predicting Early Stage Startup Success through a Hybrid Intelligence Method," *ICIS 2017 Proceedings*, pp. 1–12, 2017.

[68] A. J. Quinn and B. B. Bederson, "Human computation: a survey and taxonomy of a growing field," pp. 1403–1412, ACM, 2011.

[69] J. Y. Zou, K. Chaudhuri, and A. T. Kalai, "Crowdsourcing feature discovery via adaptively chosen comparisons," 2015.

[70] A. P. Dawid and A. M. Skene, "Maximum likelihood estimation of observer error-rates using the em algorithm," *Applied statistics*, vol. 28, pp. 20–28, 1979.

[71] H.-C. Kim and Z. Ghahramani, "Bayesian Classifier Combination," in *Proceedings of the International Conference on Artifical Intelligence and Statistics (AISTATS)*, 2012.

[72] K. A. M. Kamar, Z. A. Hamid, and N. Dzulkalnine, "Industrialised Building System (IBS) construction: Measuring the perception of contractors in Malaysia," in *BEIAC 2012 - 2012 IEEE Business, Engineering and Industrial Applications Colloquium*, 2012.

[73] M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel, "Manifestations of user personality in website choice and behaviour on online social networks," *Machine learning*, vol. 95, no. 3, pp. 357–380, 2014.

[74] J. Whitehill, T.-f. Wu, J. Bergsma, J. R. Movellan, and P. L. Ruvolo, "Whose vote should count more: Optimal integration of labels from labelers of unknown expertise," 2009.

[75] A. Segal, K. sGal, E. Kamar, E. Horvitz, and G. Miller, "Optimizing interventions via offline policy evaluation: Studies in citizen science," *AAAI 2018*, 2018.

[76] L. Von Ahn and L. Dabbish, "Designing games with a purpose," *Communications of the ACM*, vol. 51, no. 8, pp. 58–67, 2008.

[77] J. W. Vaughan, "Making better use of the crowd: How crowdsourcing can advance machine learning research," *Journal of Machine Learning Research*, vol. 18, no. 193, pp. 1–46, 2018.

[78] A. Bernardo, A. Cervero, M. Esteban, E. Tuero, J. R. Casanova, and L. S. Almeida, "Freshmen Program Withdrawal: Types and Recommendations.," *Frontiers in psychology*, vol. 8, p. 1544, 2017.

[79] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*, pp. 740–755, 2014.

[80] V. S. Sheng, F. Provost, and P. G. Ipeirotis, "Get another label? improving data quality and data mining using multiple, noisy labelers," *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 614–622, 2008.

[81] L. Zhao, Y. Zhang, and G. Sukthankar, "An active learning approach for jointly estimating worker performance and annotation reliability with crowdsourced data," *arXiv preprint arXiv:1401.3836*, 2014.

[82] G. Kou, Y. Peng, and G. Wang, "Evaluation of clustering algorithms for financial risk analysis using mcdm methods," *Information Sciences*, vol. 275, pp. 1–12, 2014.

[83] A. M. Chirkin and R. König, "Concept of Interactive Machine Learning in Urban Design Problems," in *Proceedings of the SEACHI 2016 on Smart Cities for Better Living with HCI and UX - SEACHI 2016*, 2016.

[84] Z. C. Lipton, "The mythos of model interpretability," *Queue*, vol. 16, no. 3, p. 30, 2018.