Fake Intelligence Online Summit 2019

Summit Proceedings

Proceedings of Fake Intelligence Online Summit 2019

Editors Harri Ketamo and Petra O’Rourke

Satakunta University of Applied Sciences
Series D, Other Publications 1/2019
ISSN 2323-8372 | ISBN 978-951-633-275-1

Publisher:
Satakunta University of Applied Sciences
P.O. Box 1001, FIN-28101 Pori, Finland
www.samk.fi

The publications of Satakunta University of Applied Sciences are
free downloadable at theseus.fi

Table of Contents

Prologue

1st Fake Intelligence Summit, May 7, 2019, Pori, Finland

Introduction to Fake Intelligence

PhD Harri Ketamo (chair), Headai, Chairman and Founder

We know the concepts of fake news, data biases and people living in their bubbles. Fake intelligence covers pretty much the same topics but in context of AI: We have to have critical discussions on how data affects to any AI and eventually reveal all kinds of AI bubbles.

In many cases fake intelligence is desired outcome: Game AI has always been there to entertain players. In most cases it is art of stupidity: Perfect opponent is boring, but human like game AI can be entertaining. In games, we have always understood that AI is the art of fake or art of human-like stupidity. This, however, don't take anything away from game AI developers. This only adds the value of game AI developers: you know what you are aiming at and at the same time you know how you fake it. In fact, we should always discuss ai like game AI. We are trying to mimic human behaviour. We know it is not human behaviour, so we can openly discuss the limits, biases and risks related to such approach. In too many cases we are just afraid on saying we don't have a perfecta AI and we know it.

Science fiction is fiction, but we have to understand is as a form of art and we have to be able to connect the message of fiction into reality. We can see scifi also as future research of as "alternative futures thinking" -process. Alternative futures can bring important dimensions into our discussion, but everyone understands we are not talking present. For example, Blade Runner, the movie, brings in front the eternal discussion on what makes human and what is the difference between human consciousness and machine consciousness. The movie does not say 'we should not develop robots, because in future they will turn against people', what is a common misinterpretation on the topic.

> *"This is your last chance. After this, there is no turning back. You take the blue pill – the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill – you stay in Wonderland, and I show you how deep the rabbit hole goes."*
>
> *— Morpheus, from The Matrix*

Fake Intelligence is in the core concept of AI literacy. Without understanding the fakes, biases, limitations behind AI, we can believe whatever we want to believe. If we want to bring a reality check into our current AI discussion, we should check how deep the rabbit hole goes.

However, we should not spend too much time on defining what is AI. All machine performed activities that would require conscious thinking also from people performing the same activity, can be called Artificial Intelligence. No matter what algorithms are used, no matter what technological framework is used. The definition of AI can't be related to technology, that's why the technologies have exact names.

Furthermore, AI is old. As a science it comes from 50's and it has been working with us for more than 60 years. The first-generation AIs were decision trees and state machines, based on programmed rules. In many solutions, the rules were based on scientific work and millions of lines of data, but the rules were programmed after people have found the rules. In second generation AI, there are no pre-programmed rules and machine learn the rules from given data. In other words, the rules how AI works are not programmed, they are (machine) learned. There are excellent use cases for both types of AI, for some cases programmed rules are perfect, because they are fast to perform, e.g. spell checking. Some other cases require machine learning because of complexity of the case, e.g. speech recognition. The third generation is going to be about extending the concept of machine learning. I.e. teaching machines case by case with natural language.

AI can be used like an assistant for people, tightly managed worker or as an autonomous system. In many cases we use AI as co-worker, letting it to do easy, easy but complex, boring, dirty or dangerous tasks. Nowadays, we use more and more AI for assisting professional and enabling professional to focus on higher order thinking while AI performs maybe time consuming and complex, but predictable processes. There are very limited number of autonomous AI's in use but still when we talk about AI we tend to think autonomous AI. It is important to notice that the ethical discussion around AI is very different if we are talking about autonomous systems or assisting/tightly managed worker type of AI. Alongside with AI ethics, we should discuss data ethics: who is allowed to collect data, who owns the data, is the data real and valid, what are biases in the data, what is the role of manipulation, and so on.

Finally, we have to accept we cannot talk AI as an isolated system, it is always connected to other technologies, processes and people and interacting with all of them.

We cannot give a detailed picture on fake, limits and biases behind AI, that's why members of the Fake Intelligence scientific committee have recognized AI bubbles and stated them Fake Intelligence. Originally published at https://www.fakeintelligence.fi/ai-bubbles-2/. In following, the bubbles and short explanations to the conference. Reality check on how deep the rabbit hole goes starts.

Bubble no. 1: Everything can be solved with Deep Learning

> *George Orwell introduces a concept of doublethink in his novel nineteen eighty-four. Doublethink means that person accepts two mutually contradictory beliefs as correct without any mental conflict or contradiction, i.e. without cognitive dissonance.*
>
> *Deep Learning has brought a paradigm shift into AI programming. It has changed the way we see AI nowadays. Definitely one of the most important solution around AI.*
>
> *Our doublethink around Deep Learning is that any intelligent operation can be done with it. No other thought allowed, or you are not a true member of AI society.*
>
> *At the same time, we know for sure (based on e.g. science) that human thinking is way more complex. We also know, that so far only relatively trivial cases are successfully implemented with Deep Learning. And still, as AI society, we doublethink that everything can be solved with Deep Learning.*
>
> *— Harri Ketamo*

As mentioned before, the definition of AI can't be related to technology, that's why the technologies have exact names. Furthermore, if Deep Learning would enable all the cases, we would have Artificial General Intelligence in our hands. Reality is not that simple.

Bubble no. 2: Every software is AI software

> *Artificial intelligence, data analytics and big data rose to public awareness some three years ago. As an immediate consequence, a huge number of software, IoT and robotics companies changed their marketing vocabulary. Now they are all AI experts in their business area.*
>
> *Their web page once advertised their ERP software to support business growth. Now their artificial intelligence solutions optimally secure and guide the business growth. Likewise, they once used statistical methods to explore data, but nowadays they exploit data analytics to reveal hidden patterns. My very good friend Harri Ketamo (bubble 1), the major stakeholder of a real AI company, has portrayed the situation enjoyably: "If one has more rows in Excel than can be seen on the display, it is big data today".*
>
> *— Cimmo Nurmi*

This is a real challenge nowadays. Because every software company is using Google's, Microsoft's or Watson's APIs, every company can say, they are AI company. However, understanding on who is really developing AI and who is applying AI is more difficult case. It's pretty much the same if a company using Java as a programming language, claims they are developing Java. They are Java Developers, not developing Java. Today AI developer is the word used for every developer, also for those who are really developing AI.

Bubble no. 3: Optimization is now easy

> *For over 10 years ago me and my friend started to optimize. Only scientists knew how to. The ordinary man didn't even recognize that it could be done.*
>
> *Today everybody knows how to. Even those who doesn't know they are in the business. They even seem to be the best ones in it. Almost.*
>
> *— Jari Kyngäs*

There are no shortcuts. It is so easy to just throw data in and get a result out, but one has to also understand what result means. If we have a classification system trained to classify photos to belong either cat -category or dog -category, no other options. Such system will classify a fish into cat or dog category, no other options. Most of our systems are like this, way more complex, but full of biases we should discuss.

Bubble no. 4: AI makes suspicious decisions

> *We are afraid of artificial intelligence making non-perfect decisions. We are not worried who trains the AI and with what data. In fact, we are not worried at all if we are fooled by beliefs told by other people, not even in the case when all the facts are against the belief.*
>
> *AI is not figuring out anything on its own, it behaves based on algorithm and data. Both are people made decisions.*
>
> *— Harri Ketamo*

The same in opposite way: some of us close their eyes on manmade biases and at the same time requires we should not develop AI. We should make decision chains and all behind computational decisions visible, no matter if it is AI or algorithm. We all know there are many occasional correlations between random events, like ice cream consumption and drowning. This, however, do not make correlation invalid computational method.

Bubble no. 5: Artificial intelligence changes how we learn in the future – or does it?

> *AI is changing quite a many areas in our society, but one thing is not changing – learning. And I mean learning in a very fundamental way. Learning is something that happens in ones' brain. It is not yet fully understood how this actually happens, but it is known that our memory is associative. We remember things from associations. A smell from your childhood might bring up a memory of a place you used to live. Many couples have a special music that means a lot to them because it was playing when they first met. Playing the song again brings up nice memories of your partner. All this is happening inside our brain and AI is not changing that anytime soon.*
>
> *However, AI will bring (and has already done so) new means to study and teach. It will even bring tools to monitor learning outcomes and even the learning process and can identify those students who are in risk of failing the course. In such, AI is just a tool for the learner and the teacher. While AI may also be used in providing learning content to the masses, teachers will still be needed to support and facilitate learning.*
>
> *— Jari Multisilta*

We could not highlight enough this. The way how people learn is an outcome of millions of years of evolution and it can't be changed with technology. In short, we learn through interaction with observed world. I.e. we make observations, conceptualize our observations and connect them into our existing understanding. When we connect observations we have done before, we just strengthen our current understanding. When adding new observations, we learn. When we make observations that don't fit into our existing understanding, we might end up learning and at the same time changing our understanding radically.

In science, the learning process is studied in e.g. chemistry level, biological level, neural cell level, social level and in psychological level. All science supports this generalized definition of learning, no matter if we call it e.g. conceptual learning, constructivism or connectivism. Claims that AI will change the way we learn is just like a claim that television will change the way we learn. There has been technologies claimed to change the learning in every ten years. Games (2010), mobile devices (2000), Internet (1990), CD-rom (1980), television (1970), programmed instruction (1960), radio (1950) and spirit duplicators (1930) just few to mention.

Bubble no. 6: Artificial Intelligence is Neural Network Intelligence

> *As the years go by, new buzzwords and industry jargon evolve, and their meaning will change. Artificial Intelligence is a good epitome of this. The term AI truly emerged in the early '80s when the first academic AI conferences were held, and the Lisp machines and parallel computing were introduced. In that time, AI was publicly advertised as rule-based expert systems. The expectations were high, but the results were not that convincing. The public interest soon faded.*
>
> *In '90s, old inventions were reinvented with the help of computer power increase. Neural networks (1943), evolutionary computation (1954), fuzzy systems (1965) and other nature-inspired algorithms evolved and started to show their abilities. A new term Computational Intelligence (CI) started to spread in order to cover all of these computational techniques. It was back then and still is debatable, whether CI is a subset of AI or vice versa.*
>
> *Today, almost all AI publicity is centered on neural networks. This had led to the public conclusion that real artificial intelligence is obtained by using neural networks. Are the other intelligent algorithms then artificial artificial intelligence? The truth is that only a small number of current intelligent systems are based on neural networks. Significant number of celebrated intelligent applications actually use*

> *optimization algorithms designed in '80s and '90s. Still, the recent real-world applications using neural networks are very promising and convincing.*
>
> *— Cimmo Nurmi*

Like mentioned many times before, the definition of AI can't be related to technology, that's why technologies and methods do have exact names.

Bubble no. 7: AI will generate jobs

> *Maybe one of the biggest bubbles we can create is about AI generating new jobs. Way too often people say that AI will generate more jobs than it will take. We have to understand that AI itself generates zero jobs. All new jobs are created by people and often enabled or powered by AI. And vice versa, all decisions about giving peoples' work to AI are made by people.*
>
> *— Harri Ketamo*

We are responsible for everything related to AI. We are responsible for all technologies we develop. We are responsible for all the decisions we make; we can't blame AI on decisions, we have to understand every decision we make.

Finally, during next 10 years, fake intelligence will be part of general literacy. This requires we start open and critical discussion also AI fails, biases, limits and non-perfect solutions, unless we choose to take the blue pill, wake up from our beds and continue to believe whatever we want to believe.

HARK Side of Deep Learning - From Grad Student Descent to Automated Machine Learning

OGUZHAN GENCOGLU
Top Data Science Ltd., Helsinki, Finland, oguzhan.gencoglu@topdatascience.com

MARK VAN GILS
VTT Technical Research Centre of Finland Ltd., Tampere, Finland

ESIN GULDOGAN
Huawei Technologies, Tampere, Finland

CHAMIN MORIKAWA
Morpho Inc., Tokyo, Japan

MEHMET SÜZEN
Jülich, Germany

MATHIAS GRUBER
Novozymes, Copenhagen, Denmark

JUSSI LEINONEN
Bayer, Espoo, Finland

HEIKKI HUTTUNEN
Tampere University, Tampere, Finland

Recent advancements in machine learning research, i.e., deep learning, introduced methods that excel conventional algorithms as well as humans in several complex tasks, ranging from detection of objects in images and speech recognition to playing difficult strategic games. However, the current methodology of machine learning research and consequently, implementations of the real-world applications of such algorithms, seems to have a recurring *HARKing* (Hypothesizing After the Results are Known) issue. In this work, we elaborate on the algorithmic, economic and social reasons and consequences of this phenomenon. We present examples from current common practices of conducting machine learning research (e.g., avoidance of reporting negative results) and failure of generalization ability of the proposed algorithms and datasets in actual real-life usage. Furthermore, a potential future trajectory of machine learning research and development from the perspective of accountable, unbiased, ethical and privacy-aware algorithmic decision making is discussed. We would like to emphasize that with this discussion we neither claim to provide an exhaustive argumentation nor blame any specific institution or individual on the raised issues. This is simply a discussion put forth by *us*, insiders of the machine learning field, reflecting on *us*.

1. Introduction

Hypothesizing after the results are known (HARKing) [1] occurs when researchers masquerade one or more post hoc hypotheses as *a priori* hypotheses. This means that instead of following a traditional hypothetico-deductive model [2], in which previous knowledge or conjecture is used to formulate hypotheses that are then tested, the researcher instead looks at the results first and then forms a *post hoc* hypothesis. HARKing can occur in different forms, such as *constructing*, *retrieving* or *suppressing* hypotheses after the results are known [1]. A number of studies in recent years have examined and discussed the incidences, causes and implications of such practices within various fields such as management, psychology as well as natural sciences [3, 4, 5].

In recent years, deep learning (DL) methods have dramatically improved the state-of-the-art (SotA) within the fields of speech recognition, visual object recognition, machine translation and several other domains such as drug discovery and genomics [6]. However, there are certain troubling trends in the current machine learning (ML) research, outlined in [7] as failure to distinguish between explanation and speculation, use of mathematics that obfuscates rather than clarifies, and misuse of language. Unfortunately, HARKing has also been one of those recurrent trends in machine learning and especially in deep learning research. Since much of such research is being eagerly applied to real-world applications in both industry and society, such issues are of utmost importance due to the wide impact of machine learning products and services across all walks of life. Transparent and reliable practices are critical when trying to combat suspicions towards new technologies, and the trust needs to be built over long period of time; as acknowledged recently even on the European Commission level [8].

Our hypothesis is that the recent explosion of advances within the fields of machine learning and in particular deep learning, as well as the hyper-competitive nature of these fields, may potentially be a dangerous breeding ground for various HARKing behaviors, the implications of which are not yet fully explored. At the very least, concerns regarding such behaviours deserve to be critically discussed from different angles so as to encourage best practices when building ML systems and algorithms. It is noted that these issues are not new by themselves. In fact, since as long as data-driven approaches and learning systems have been around, it has been critical, and sometimes difficult, to remain fully objective in analyzing results. Issues have been reported earlier for example, such as self-deception practiced by scientists; finding patterns that are not there [9, 10].

In this paper we discuss HARKing behavior from different angles:

- Section 2 - Competitiveness in DL research leading to questionable improvements of state-of-the-art and claims of novelty.
- Section 3 - Pressure to create reports that are favorable for publication and aversion towards negative results.
- Section 4 - The belief that current training datasets are representative of real-world samples.
- Section 5 - Automated machine learning
- Section 6 - Explainability, ethics, reproducibility and more for AI systems.

2. Grad Student Descent and SotA-hacking

In a typical deep neural network model there are numerous design choices, i.e., tunable parts such as model architecture and hyper-parameters, that affect the predictive performance. Proposing a decent set of these design choices that will result in high generalization ability (relative to the other sets of choices) is difficult mainly due to two reasons. Firstly, due to the inherent non-deterministic and highly non-linear nature of neural networks, it is not trivial to deduce explicit relationships neither between the hyper-parameters and the model performance, nor between the interactions of hyper-parameters themselves. For instance, a large *batch size* is key to speed up neural network training in large distributed computation infrastructures, however, significant degradation in model performance has been observed in practice when large batch sizes are employed [11]. To overcome this issue, typically, hyper-parameters belonging to the *optimizer* need to be tuned. Secondly, as the parameter search space increases exponentially, it is not feasible to apply exhaustive or brute-force search methods. Therefore, a significant portion of deep learning research has been focusing on engineering efficient model architectures and hyper-parameters for specific tasks.

Even though this manual discovery process has been successful for several applications (often empirically), there has been significant divergence from the traditional hypothesis-driven scientific approach in the methodology of such studies. Instead of hypothesis-forming based on theory, extensive research on previous studies and/or reflection against the existing domain knowledge, *grad student descent* (a cheesy pun referring to the well-known *gradient descent* algorithm) is applied.

Grad student descent is a type of optimization scheme in which the task of model architecture or hyper-parameter search is assigned to several graduate students, usually to be performed by trying what works and what does not. This is an iterative approach, where one starts with a baseline architecture or possibly with an earlier SotA, measures its performance and applies various modifications by trial-and-error, without a sound hypothesis. Once marginal improvements are observed, iterations of modifications continue further in that direction until a local optimum (often a publishable result) is reached and an explanation is forged. In essence, this whole process is driven by HARKing. Furthermore, this process is performed with a limited set of data, that is used and re-used again and again to find the "optimal" solution (further discussed in Section 4). Oftentimes, final testing on a completely independent test set that has not been touched or observed at all at any moment is not performed and cross-validation is either not used or used under problematic assumptions and/or executions such as performing model tuning and estimation of model error at the same time [12, 13].

The abovementioned HARKing pattern, consequently, results in increased difficulty in distinguishing and identifying why a proposed method works or not. Lack of thorough hypothesis forming prior to experimentation often leads to negligence of comprehensive discussions on the results as well, especially when accompanied with comparison of a single score or metric. For instance, a recent work by Reimers and Gurevych shows that reporting a single performance score is insufficient to compare non-deterministic approaches such as neural networks [14]. Their study demonstrates that the seed value for the random number generator can result in statistically significant differences in performances of state-of-the-art methods [14].

The negative effects of HARKing are not specific to deep learning research alone, and they can be observed in research dealing with traditional machine learning methods as well. However, as the concept of *state-of-the-art* (a method or a set of methods that outperforms all the previously proposed methods for a given machine learning task in a certain metric such as test accuracy, inference speed, training speed etc.) has been disproportionately promoted in DL, both in academy and industry, presence of HARKing is becoming more likely to be overlooked especially if there are claims of advancing the SotA. This phenomenon has been promoting the concept of *SotA-hacking* and publishing of *marginally SotA* results without in-depth analysis or discussion, similar to *p-hacking*, *data dredging* and prevalence of *marginally significant* results in several other fields [15, 16, 17]. Typical examples of misleading comparisons leading to unfair or inadequate SotA claims include usage of additional training data (the common concept of *transfer learning* in DL), usage of data augmentation, comparison to poorly implemented baselines or ensembling of several models. Similar unjustified claims can be observed in "novelty" of proposed methods as well.

3. Chronic Allergy to Negative Results

Publication bias, the phenomenon occurring when the probability of a scientific study being published is not independent of its results [18], leads to systematic difference in the findings of published tests of a claim from the findings of all tests of the same claim [19]. Often recurring as a *positive outcome bias*, this phenomenon has been observed in several research fields for a long time [20, 21, 22]. For example, in clinical research, studies finding no difference between the study groups were less likely to be published than those with statistically significant results [20]. In fact, there is evidence of negative results being less likely to be published even if they provide corrections of errors in previous studies [23]. A similar troubling trend has been prevalent in ML/DL research and arguably HARKing exacerbates this further.

Publishing a null or negative result in the current ML researchosphere is considerably difficult due to the widespread assumption that "every positive result is scientifically more valuable, or interesting, than any negative one". This is likely even more the case in DL research because of the ever-increasing competition. For instance, the percentage of accepted papers related to deep neural networks in the *Conference on Computer Vision and Pattern Recognition (CVPR)*, one of the most prestigious in its field, has been 1%, 14% and 25% for the years 2013, 2015 and 2017, respectively [24]. Note that the amount of publication submissions to conferences and journals are increasing every year as well, e.g., the number of submissions to *Annual Conference on Neural Information Processing Systems (NeurIPS)* doubled from 2016 to 2018 [25]. Similar trends can be expected to be observed in research funding or scholarship applications. A research proposal is more likely to get a positive review if it builds further on "encouraging results" from previous work. There have been incentives to discuss the importance of negative results and share them in ML research [26] such as the *First Workshop on Negative Results in Computer Vision* in 2017 and we hope more actions towards this direction will be realized in the future.

Current outcome reporting bias in ML/DL research is generated both from the authors' side as a reluctance to report negative results as well as the journals' side in selecting the results worth publishing and it is not trivial to separate the extent of the two. Even in presence of a positive result, authors may not report the negative ones, thinking such reporting will devaluate their work. As stated by Nissen et al., even if authors' behavior is the main contributor to publication bias (there is evidence supporting this in other fields [27, 28]), they may simply be responding to the editorial preferences for positive results [29]. The lack of traditional hypothesis construction before conducting the experiments and the lack of expectation to do so, supports the incentive of avoiding reporting of negative results in ML/DL field.

There are several consequences of such allergy against negative results in deep learning research. First, it eventually creates a bias against disruptive innovative ideas and favors incremental tweaks on well-established methods. Secondly, when negative results are not reported or published, it is essentially more difficult to construct causality and elaborate on the phenomena behind the positive results. As in other aspects of life, after all, we learn from negative results as well as positive ones. Furthermore, it increases the waste of resources and efforts due to unnecessary (re-) implementation of methods that have been shown to be inferior but never reported. Finally, the probability of a negative result being caused simply because of poor implementation exhibits the potential of that work being influential once implemented properly.

The trend of starting from a solution (often somebody else's) instead of from the problem itself and HARKing after minor modifications can be changed by changing our paradigm of publication process. Hereby, we propose a *results-blind* review process for ML/DL research:

- A paper is submitted with a clear hypothesis accompanied with the design of experiments. The hypothesis can be based on extensive analysis of previous studies, mathematical theory with unambiguous assumptions and/or domain knowledge of the specific field.
- The paper gets peer reviewed, preferably double-blind, and the reviewers suggest modifications and improvements on the experimental methods.
- Once accepted, the experiments are run.
- The paper gets published regardless of the results with a comprehensive discussion section.

This approach would increase the likelihood of the study to be informative and influential regardless of the outcome, not only in the case of positive results. Essentially, the review process will give more attention to the experimental design and the hypothesis behind the proposed methods, decreasing the incentive for HARKing significantly. Naturally, this will also encourage researchers to navigate outside the "marginal improvements over the previous SotA" thinking. Similar ideas have been discussed especially in the field of psychology [30, 31, 32]. Note that we do not claim that the abovementioned proposal is applicable for every machine learning research publication process, mostly due to the scarcity of high-quality reviewers. Nevertheless, we believe such discussions are beneficial and may eventually lead to improvements that will decrease the prevalence of HARKing in ML/DL research.

4. "In the Wild" Illusion

Numerous studies in the field of deep learning utilize publicly available annotated datasets for computer vision, natural language processing, audio analysis and various other tasks. Several of these datasets even include the phrase "In the Wild" in their name - an expression to convey the message that the dataset holds no constraints and is representative of real-world circumstances. Even though it is not stated explicitly, the main assumption behind using these datasets is that the observations belonging to these datasets are drawn from the same statistical distribution of all possible observations naturally occurring in real-world.

In 2011, Torralba and Efros proposed to examine dataset bias in twelve popular image datasets by observing if it is possible to train a machine learning model to identify the dataset a given image is selected from [33]. Considering the random guess accuracy is only 1/12 ≈ 8%, the authors found that humans were able to perform at > 75%, while a simple support vector machine classifier performed at 39%. The authors furthermore demonstrated the inability to perform cross-dataset generalization, thereby highlighting how models trained on typical datasets actually *overfit* and thus fail to generalize to other datasets yet alone to real-world settings.

A similar problem of overfitting stems from the hyper-competitive nature of machine learning, where there is little incentive of trying to publish methods that have inferior performance compared to SotA on test datasets (see Section 3). Therefore, we can reasonably expect that effectively most research uses the test set as a validation set, rather than following the standard practice of defining a separate validation set from the training data. Recht et al. show this by creating a new test set for CIFAR10, a widely used image dataset, where they found that there was a significant drop in accuracy (4-15%) from the old test set to the new test set when tested with several DL architectures [34]. In a more recent work, a similar phenomenon is also shown for the well-known ImageNet dataset, suggesting that the accuracy drops are caused by the models' inability to generalize to slightly "harder" images than those found in the original test sets [35].

From the HARKing perspective, formulating hypotheses that are specifically designed to account for the observed results for a specific sample of observations go hand in hand with overfitting and failure of generalization. Furthermore, the selected datasets to run the proposed experiments on have to be in parallel with the hypothesis. For instance, the well-known *Labeled Faces in the Wild* dataset [36] contains images of famous people only, but have been used extensively to test hypotheses of face recognition or person identification in unconstrained settings. And from the implementation perspective, by splitting a dataset into training, validation, and testing sets, we invariably risk giving the false impression that because our model may perform well on the test dataset, it will also generalize to images found in real world applications. In both cases mentioned above (using biased datasets and/or overfitting to specific test sets), it can be argued that hypotheses testing is conditional on the dataset in question, and therefore to convince a reader that HARKing has not occurred, an author should always take great care to demonstrate the generalizability of new methods. Obviously, overfitting is a problem encountered in ML in general and is not specific to neural networks. However, considering:

i. feed-forward neural networks are universal function approximators (by *Universal Approximation Theorem*) as well as convolutional networks, i.e., a single hidden layer network containing a finite number of neurons can approximate continuous functions with arbitrary precision [37, 38]

ii. the complexity of the computed function by a neural network grows exponentially with its depth, i.e., for every additional hidden layer, one needs *exponentially* more parameters to express the same function with a shallower network [39, 40]

deep neural architectures are very likely to suffer from overfitting due to their expressive power.

5. Automated Machine Learning

The traditional data science approach relies on many sequential tasks; i.e. data preprocessing and cleaning, feature engineering and selection, model selection and parameter tuning, postprocessing, and finally critical analysis of results. Often in practice, the human decision-making processes in these tasks are inefficient (see Section 2) or based on heuristics. Furthermore, the combined complexity of these tasks often present an insurmountable barrier for non-experts, and thus automated machine learning (AutoML) is a topic that has become increasingly popular in recent years, promising to automate (at least parts) of this pipeline in order to improve efficiency of machine learning and accelerating research.

Recently, the most popular AutoML task has focused extensively on neural architecture search (NAS) [41, 42, 43, 44, 45, 46, 47], i.e., automating the design of neural network architectures for the search of architectures that are superior to hand-crafted ones. Several other AutoML tasks include automated hyper-parameter optimization [48], activation function search [49], optimizer search [50], data augmentation policy search [51] or even search for better hardware utilization in heterogeneously distributed (mixture of CPUs and GPUs) computing environments [52]. The methods behind such *meta-learning* approaches are mainly based on *Bayesian optimization* [48], *evolutionary algorithms* [43, 46] or more recently on *reinforcement learning* [41, 49, 52]. Some of these methods are available both to the academy as well as to the industry as open source software or in the form of software-as-a-service.

These advancements not only help us discover better DL models and solutions in terms of quantitative metrics than hand-engineered ones, but also carry the possibility to transform the everyday working practices of machine learning researchers and practitioners. With AutoML, data scientists are expected to offload a significant portion of their routine work and focus on tasks that require a higher-level thinking and creativity. However, certain issues have been raised related to AutoML approaches lately. For instance, Scuito et al. demonstrate that the search policies of state-of-the-art NAS techniques are no better than random policies [53]. Similarly, Li and Talwalkar show that random search with early-stopping is a competitive NAS baseline on two benchmark tasks - one from computer vision and one from natural language processing [54]. In addition, they discuss the reproducibility issues of published NAS results by elaborating on the necessity of having a tremendous amount of computation resources, lack of available source material/code and questionable robustness of published results [54].

Interestingly, the pursuit of simplifying machine learning development resulted in a significant increase in algorithmic complexity of AutoML methods including complicated training routines and architecture transformations [54]. This complexity makes it more difficult to pinpoint which components of the found solution is crucial for high performance. In addition, considering the lack of ablation studies (the analysis of systematic removal of components or features of a model in order to identify which of them are the most relevant) in many works, AutoML field creates a dangerous ground for HARKing.

6. The Insert_Adjective_Here AI Wave

6.1 Ethical AI

Ethical issues regarding current developments in machine learning are perhaps much more critical than they currently perceived to be; as we already encounter ethically questionable decisions given by algorithms, sometimes unbeknownst to us. Examples include replacing faces and voices in videos [55], detecting people using WiFi signals [56], deciding whose life to risk in an eminent accident [57] and generating fake news [58]. In various scenarios, ML impacts decisions on legal and ethical issues as well such as insurance, hiring, lending. Therefore, it is crucial to develop models that are fair and unbiased regardless of the biases in the data [59, 60]. This issue has been recently emphasized even by the European Commission in their ethics guidelines report for AI by underlining the importance of paying attention to situations involving more vulnerable groups such as children, persons with disabilities or minorities, or to situations with asymmetries of power or information (e.g. employee-employer or business-consumer) [61].

With established industries (e.g. example firearms), it is common for the researchers and developers to leave the responsibility of ethics to entities that follow them (e.g. arms sellers and legislators). However, most AI-based system have been much faster to deploy than conventional technology. Therefore, it is highly desirable for researchers to discuss ethical implications of their work and create a dialogue about them at the earliest possible stage. While selecting research topics that raise ethical issues itself serves this purpose, the desire to present good results might deter the discussion.

Another important ethical issue revolves around covert AI systems. A human should always know if she/he is interacting with a human being or a machine, and it is the responsibility of us that this is reliably achieved. As AI practitioners, we should ensure that humans are made aware of - or able to request and validate the fact that – they interact with an AI identity [61]. Thus, hypothesis forming process should be clear and unambiguous, and should consider the possible use cases or implications as well. And in this pursuit, HARKing won't do.

6.2 Human-centric AI

At the current stage, ML/DL algorithms are often designed as tools for defined domain experts, thus they need to address human needs and psychology in a realistic manner. To decrease the amount of HARKing, high-level domain experts should be incorporated to the study teams from the beginning as a collective intelligence of domain experts has considerable benefits and should be utilized whenever possible [62]. This will lead to more successful forming of *a priori* hypotheses and in the end should put pressure on scrutinizing results that do not support these hypotheses. Previously, worrying examples of failure in this have surfaced, where there has been only a limited input from the domain experts [63]. High-level expertise is especially relevant to create scientific hypotheses and should be differentiated from defining practical use-cases and training of AI, where a diverse spectrum of possible users should be affiliated to the project.

HARKing is potentially a serious threat especially in AI-driven change in medical practice. This applies mostly to the effect of failing to report *a priori* hypotheses that are unsupported by the current results [5]. The algorithms that will be used in medicine typically need to be clinically validated in laborious and high-cost trials [64]. Suppressing hypotheses after the results are known can lead to wrongly planned clinical trials, as the background scientific literature (meta-analyses) is biased and this can lead to losing credibility in the eyes of physicians and decision makers, together with spending a huge amount of limited human and financial resources available to run these trials.

6.3 Explainable, transparent and interpretable AI

Explainable artificial intelligence (XAI) is not only interesting as an academic curiosity; it is a necessity for the future. Developing explainable and transparent systems, as well as tools to measure transparency, is crucial for ethical AI development (see section 6.1). The main concept of XAI is centered around *causal attribution* as it is in human nature to understand causality naturally. Having such causal explanations will provide substantial leap in reaching human-like perception of AI systems and anthropomorphism [65]. Explainable AI and model interpretability may be used in a synonymous manner. However, we think that *explainability* may fall under the causality domain and *interpretability* may belong to the mechanistic explanation of the algorithmic and model internals [66].

Recent deep learning algorithms provide high predictive performance but limited ways to provide reasoning on how an algorithm produces such level of high performance that exceeds human abilities [67]. Even though there have been studies addressing this problem and proposing solutions [68, 69, 70], a common consensus on performing interpretation of ML and especially DL models has not been reached. In fact, even the definition of interpretability itself is not established, neither mathematically nor axiomatically in the literature [66]. Furthermore, recent studies question the robustness and security of these interpretation methods (e.g. to adversarial attacks) [71].

From HARKing perspective, one can relatively easily reverse engineer results to fit in a desired interpretation [69, 71, 72]. To avoid such practices, interpretable algorithms should not be reversible, nor should they only provide interpretation depending upon algorithmic priors. In this regard, approaches aiming at more theoretical explanations of *why* deep learning works, from learning theory to statistical physics [73, 74, 75], may be classified as *true* XAI research. These approaches, rather than focusing only on interpretation of the mechanistic approaches after the results are known, aim at finding an *ab-initio* technique, i.e., from the first-principles, to design a deep learning system without HARKing. Similarly, use of causal inference has recently been shown to be promising in understanding underlying mechanisms of deep learning systems [76] and if descriptive, causal modals can answer prediction, intervention and counterfactual questions [77].

In terms of transparency, an interesting question is whether we are, as humans, required to know all the details about the AI capabilities of the equipment and sensors that surround us. This can be argued both ways; for example, we know virtually nothing about the abilities of human drivers that use the same highway as we do. But similar to what happened with established technology in automotive (like ABS and automatic transmission), we should be able to know the workings, accuracy stats, advantages and disadvantages of emerging AI technologies. This concept overlaps with abovementioned mechanistic interpretability issue and perception of human-like attributions.

6.4 Reproducible AI

AI research is known and as a result appreciated for its significant contributions to open science (e.g. preprint archives), open source (e.g. code repositories, sharing of trained models etc.), open data and reproducible research paradigms. Yet, as a sub-field of computer science, it still shares a similar reproducibility crisis [78, 79, 80, 81, 82]. As Donoho et al. suggested, a computational research paper is merely an advertisement unless it is presented with an underlying code and data [78]. We believe one of the reasons of this reproducibility crisis is HARKing.

One essential contribution to this crisis in ML and especially in DL research is the lack of understanding of distinction between *repeatability* and *reproducibility* [83]. We consider repeatability as the ability to recreate the results of a study/paper and reproducibility as the ability to reach the same conclusions despite the variations in the irrelevant components of the experiments [84]. Obviously, the role of hypothesizing driven by sound scientific methodology is essential in differentiating the two. As discussed in Section 2, competitive nature of the field and elevated pressure of achieving research and business outputs in a fast manner, lead to hurried claims of reproducibility (often confused with repeatability) just like the hurried claims of SotA. Once this is coupled with the avoidance of reporting negative results or similar selective reporting (see Section 3), reproducibility crisis becomes inevitable.

It is important to acknowledge the initiatives for encouraging and increasing reproducibility in ML/DL research. For instance, in NIPS 2019, a *reproducibility checklist* and a code submission policy is introduced, in which the code is expected to accompany the accepted papers. In *AAAI Conference on Artificial Intelligence* in 2019, a workshop on reproducible AI has been held. Similarly, a workshop on reproducibility in ML was held in *International Conference on Learning Representations (ICLR)* in 2019. Nevertheless, open questions remain such as "How can we measure reproducibility?", "What does it mean for a paper to have successful or unsuccessful replications?" or "What can the ML community learn from other fields?".

6.5 Accountable AI

Accountability of algorithmic decision-making systems (e.g. credit scoring) has been under discussion as well as under implementation for decades especially from the regulatory and legal perspective. However, the rapid pace of AI developments and real-world applications of them, introduced circumstances in which high-stakes decisions with significant consequences for people and broader society are made by ML algorithms. One such potential impact is an *accident* which can be, in this context, defined as an unintended and harmful behavior that emerges from poor design of real-world AI systems. Amodei et al. provides several concrete examples of such possible problems in AI safety including negative side effects (e.g. due to poorly designed objective functions), sensitivity to distributional shifts (the environment shifting away from the training environment) and reward hacking (the system gaming its objective function) [85].

Naturally, AI accountability is intertwined with explainability, reproducibility, fairness and human-centrism of design of these systems. Policies for demanding explanations of algorithmic decisions may help preventing negative consequences or may unintentionally hinder innovation while providing little meaningful protection, depending on their implementation and execution. For instance, European Union General Data Protection Regulation (GDPR) [86] introduced a potential accountability mechanism by *right to explanation* since May 2018, but the concrete consequences are still yet to be observed. Regarding the role of reproducibility in accountability of AI systems, the fatal accident recently caused by an autonomous car (belonging to Uber) is a suitable example. The preliminary report released by the United States National Transport Safety Board stated that the self-driving system software misclassified the pedestrian and the system was not designed to alert the human operator under such emergency conditions [87]. For the fair design of AI systems from the accountability perspective, the Gender Shades study [88] serves as an interesting example. In the study, biases present in commercial automated facial analysis algorithms are presented [88] and consequently, a recent study elaborated on the concept of *actionable auditing* by investigating the impact of publicly naming biased performance results of commercial AI products [89]. Certain opportunities for hybrid models in which humans and machines interact (for explaining failures [90] or intervening operations [91]) towards better AI accountability are also proposed in recent studies.

From the industry perspective, considering large companies and corporations entering an "AI race" in order to be the first to successfully employ AI in their domains, it is not surprising for accountability to take lower priority over invention and market leadership. But from the scientific methodology perspective, taking accountability of ML/DL models into account in the early stages of the research process, such as hypothesis forming, is imperative.

6.6 Privacy-aware AI

Current implementations of ML algorithms require access to data, which essentially opens up potential security and privacy risks. Therefore, privacy-aware or privacy-preserving AI notion and several studies along this paradigm has been conducted, leading to influential concepts including *federated learning* and *differential privacy* [92, 93]. With the use of *homomorphic encryption*, deep learning model inference on encrypted data was shown to be possible with a little trade-off from accuracy as well [94, 95]. In addition, Shokri et al. introduced and elaborated on the concept called *membership inference attack*, i.e., given a black-box machine learning model and a data record, determining whether this record was used as part of the model’s training dataset or not [96]. All these advancements are crucial to declare that several metrics are needed to assess and compare ML models and privacy preserving capability is one of them. For a good scientific conduct, our hypotheses on both the methods and impacts of our research should consider these concepts.

7. Conclusion

Hypothesizing after the results are known has been observed in several fields of research throughout the history and recently deep learning research exhibits several instances of it as well. In this work, we tried to give examples of HARKing in machine learning and especially in deep learning research. We elaborated on the reasons and consequences of this troubling trend by discussing overemphasis on single-metric model comparisons and benchmarks (Section 2), tendency to refrain from reporting negative results (Section 3), failure of generalization (Section 4) and automatic machine learning (Section 5). Finally, HARKing and importance of formulating an a priori hypothesis is reviewed from the perspective of ethical, human-centric, explainable, reproducible, accountable and privacy-preserving AI notions (Section 6).

We would like to emphasize the importance of discussions for achieving concrete reforms in the mentioned issues. Cultural change and legitimate interventions (such as the proposal in Section 3) in deep learning research should be encouraged by addressing these issues as much as we can in a constructive manner. As the aimed progress is a collaborative effort, researchers, practitioners, reviewers, editors, policy-makers, decision-makers, funding agencies, corporations and governmental entities need to act collectively. We believe that prevention of HARKing will help in engineering ethical, accountable, transparent, unbiased and scientifically superior deep learning solutions for the common good of the society we will be living in eventually. We also hope and believe that this work will stir discussions and debates and will contribute towards that goal.

References

[1] Norbert L Kerr. Harking: Hypothesizing after the results are known. Personality and Social Psychology Review, 2(3):196–217, 1998.
[2] Carl G. Hempel. Philosophy of Natural Science. O'Reilly Media, Incorporated, 1966.
[3] Kevin R Murphy and Herman Aguinis. Harking: How badly can cherry-picking and question trolling produce bias in published results? Journal of Business and Psychology, pages 1–17, 2017.
[4] Christopher Hitchcock and Elliott Sober. Prediction versus accommodation and the risk of overfitting. The British journal for the philosophy of science, 55(1):1–34, 2004.
[5] Mark Rubin. When does harking hurt? identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. Review of General Psychology, 21(4):308–320, 2017.
[6] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. Nature, 521(7553):436, 2015.
[7] Zachary C Lipton and Jacob Steinhardt. Troubling trends in machine learning scholarship. arXiv preprint arXiv:1807.03341, 2018.
[8] European Commission. Building trust in human-centric artificial intelligence. April 2019.
[9] Regina Nuzzo. Fooling ourselves. Nature, 526(7572):182, 2015.
[10] David H Bailey, Jonathan Borwein, Marcos Lopez de Prado, and Qiji Jim Zhu. Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance. Notices of the American Mathematical Society, 61(5):458–471, 2014.
[11] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. arXiv preprint arXiv:1609.04836, 2016.
[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning: data mining, inference, and prediction, springer series in statistics, 2009.
[13] Gavin C Cawley and Nicola LC Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research, 11(Jul):2079–2107, 2010.
[14] Nils Reimers and Iryna Gurevych. Reporting score distributions makes a difference: Performance study of lstm-networks for sequence tagging. arXiv preprint arXiv:1707.09861, 2017.
[15] Hanan C Selvin and Alan Stuart. Data-dredging procedures in survey analysis. The American Statistician, 20(3):20–23, 1966.
[16] Megan L Head, Luke Holman, Rob Lanfear, Andrew T Kahn, and Michael D Jennions. The extent and consequences of p- hacking in science. PLoS biology, 13(3):e1002106, 2015.
[17] Anton Olsson-Collentine, Marcel ALM van Assen, and Chris HJ Hartgerink. The prevalence of marginally significant results in psychology over time. Psychological science, page 0956797619830326, 2019.

[18] Theodore D Sterling. Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. Journal of the American statistical association, 54(285):30–34, 1959.
[19] Fujian Song, A Eastwood, Simon Gilbody, Lelia Duley, and A Sutton. Publication and related biases: a review. Health technology assessment, 4(10), 2000.
[20] Phillipa J Easterbrook, Ramana Gopalan, JA Berlin, and David R Matthews. Publication bias in clinical research. The Lancet, 337(8746):867–872, 1991.
[21] Ronald Moscati, Dietrich Jehle, David Ellis, Albert Fiorello, and Michael Landi. Positive-outcome bias: comparison of emergency medicine and general medicine literatures. Academic emergency medicine, 1(3):267–271, 1994.
[22] Theodore D Sterling, Wilf L Rosenbaum, and James J Weinkam. Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. The American Statistician, 49(1):108–112, 1995.
[23] Natalie Matosin, Elisabeth Frank, Martin Engel, Jeremy S Lum, and Kelly A Newell. Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture, 2014.
[24] Cvpr statistics. http://jponttuset.cat/are-gans-the-new-deep/. Accessed: 2019-04-07.

[25] Acceptance rates and submission numbers for main machine learning conferences. https://github.com/lixin4ever/Conference-Acceptance-Rate. Accessed: 2019-04-07.

[26] Christophe G Giraud-Carrier and Margaret H Dunham. On the importance of sharing negative results. SIGKDD explorations, 12(2):3–4, 2010.

[27] Carin M Olson, Drummond Rennie, Deborah Cook, Kay Dickersin, Annette Flanagin, Joseph W Hogan, Qi Zhu, Jennifer Reiling, and Brian Pace. Publication bias in editorial decision making. Jama, 287(21):2825–2828, 2002.

[28] Annie Franco, Neil Malhotra, and Gabor Simonovits. Publication bias in the social sciences: Unlocking the file drawer. Science, 345(6203):1502–1505, 2014.

[29] Silas Boye Nissen, Tali Magidson, Kevin Gross, and Carl T Bergstrom. Publication bias and the canonization of false facts. Elife, 5:e21451, 2016.

[30] Joseph J Locascio. Results blind science publishing. Basic and applied social psychology, 39(5):239–246, 2017.

[31] Michael R Hyman. Can "results blind manuscript evaluation" assuage "publication bias"? Basic and applied social psychology, 39(5):247–251, 2017.

[32] Haley M Woznyj, Kelcie Grenier, Roxanne Ross, George C Banks, and Steven G Rogelberg. Results-blind review: a masked crusader for science. European Journal of Work and Organizational Psychology, 27(5):561–576, 2018.

[33] A Torralba and AA Efros. Unbiased look at dataset bias. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, pages 1521–1528. IEEE Computer Society, 2011.

[34] Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do cifar-10 classifiers generalize to cifar-10? arXiv preprint arXiv:1806.00451, 2018.

[35] Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do imagenet classifiers generalize to imagenet? arXiv preprint arXiv:1902.10811, 2019.

[36] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition, 2008.

[37] Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. arXiv preprint arXiv:1804.10306, 2018.

[38] Ding-Xuan Zhou. Universality of deep convolutional neural networks. arXiv preprint arXiv:1805.10769, 2018.

[39] Ronen Eldan and Ohad Shamir. The power of depth for feedforward neural networks. In Conference on learning theory, pages 907–940, 2016.

[40] Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl Dickstein. On the expressive power of deep neural networks. In Proceedings of the 34th International Conference on Machine Learning-Volume 70, pages 2847–2854. JMLR. org, 2017.

[41] Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578, 2016.

[42] Esteban Real, Sherry Moore, Andrew Selle, Saurabh Saxena, Yutaka Leon Suematsu, Jie Tan, Quoc Le, and Alex Kurakin. Large-scale evolution of image classifiers. arXiv preprint arXiv:1703.01041, 2017.

[43] Masanori Suganuma, Shinichi Shirakawa, and Tomoharu Nagao. A genetic programming approach to designing convolutional neural network architectures. In Proceedings of the Genetic and Evolutionary Computation Conference, pages 497–504. ACM, 2017.

[44] Hieu Pham, Melody Y Guan, Barret Zoph, Quoc V Le, and Jeff Dean. Efficient neural architecture search via parameter sharing. arXiv preprint arXiv:1802.03268, 2018.

[45] Chenxi Liu, Barret Zoph, Maxim Neumann, Jonathon Shlens,Wei Hua, Li-Jia Li, Li Fei-Fei, Alan Yuille, Jonathan Huang, and Kevin Murphy. Progressive neural architecture search. In Proceedings of the European Conference on Computer Vision (ECCV), pages 19–34, 2018.

[46] Esteban Real, Alok Aggarwal, Yanping Huang, and Quoc V Le. Regularized evolution for image classifier architecture search. arXiv preprint arXiv:1802.01548, 2018.

[47] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le. Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 8697–8710, 2018.

[48] Aaron Klein, Stefan Falkner, Simon Bartels, Philipp Hennig, and Frank Hutter. Fast bayesian optimization of machine learning hyperparameters on large datasets. arXiv preprint arXiv:1605.07079, 2016.

[49] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. arXiv preprint arXiv:1710.05941, 2017.

[50] Irwan Bello, Barret Zoph, Vijay Vasudevan, and Quoc V Le. Neural optimizer search with reinforcement learning. arXiv preprint arXiv:1709.07417, 2017.

[51] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501, 2018.

[52] Azalia Mirhoseini, Hieu Pham, Quoc V Le, Benoit Steiner, Rasmus Larsen, Yuefeng Zhou, Naveen Kumar, Mohammad Norouzi, Samy Bengio, and Jeff Dean. Device placement optimization with reinforcement learning. arXiv preprint arXiv:1706.04972, 2017.

[53] Christian Sciuto, Kaicheng Yu, Martin Jaggi, Claudiu Musat, and Mathieu Salzmann. Evaluating the search phase of neural architecture search. arXiv preprint arXiv:1902.08142, 2019.

[54] Liam Li and Ameet Talwalkar. Random search and reproducibility for neural architecture search. arXiv preprint arXiv:1902.07638, 2019.

[55] Deepfake - an introduction. https://www.scip.ch/en/?labs.20181004. Accessed: 2019-03-04.

[56] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human pose estimation using radio signals. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 7356–7365, 2018.

[57] Giuseppe Contissa, Francesca Lagioia, and Giovanni Sartor. The ethical knob: ethically-customisable automated vehicles and the law. Artificial Intelligence and Law, 25(3):365–378, Sep 2017.

[58] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. Technical report, Technical report, OpenAi, 2018.

[59] Matt J. Kusner, Joshua R. Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. In Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA, pages 4069–4079, 2017.

[60] David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. Fairness through causal awareness: Learning causal latent-variable models for biased data. In Proceedings of the Conference on Fairness, Accountability, and Transparency, pages 349–358. ACM, 2019.

[61] European Commission. Ethics guidelines for trustworthy ai. April 2019.

[62] Michael L Barnett, Dhruv Boddupalli, Shantanu Nundy, and David W Bates. Comparative accuracy of diagnosis by collective

intelligence of multiple physicians vs individual physicians. JAMA network open, 2(3):e190096–e190096, 2019.
[63] Casey Ross and Ike Swetlitz. Ibm's watson supercomputer recommended 'unsafe and incorrect cancer treatments, internal documents show. Stat News https://www. statnews. com/2018/07/25/ibm-watson-recommended-unsafeincorrect-treatments, 2018.
[64] Eric J Topol. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1):44, 2019.
[65] David Gunning. Darpa's explainable artificial intelligence (xai) program. In Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI '19, pages ii–ii, New York, NY, USA, 2019. ACM.
[66] Zachary C. Lipton. The mythos of model interpretability. Queue, 16(3):30:31–30:57, June 2018.
[67] Taehyun Ha, Sangwon Lee, and Sangyeon Kim. Designing explainability of an artificial intelligence system. In Proceedings of the Technology, Mind, and Society, TechMindSociety '18, pages 14:1–14:1, New York, NY, USA, 2018. ACM.
[68] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PloS one, 10(7):e0130140, 2015.
[69] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pages 1135–1144, New York, NY, USA, 2016. ACM.
[70] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, pages 618–626, 2017.
[71] Amirata Ghorbani, Abubakar Abid, and James Zou. Interpretation of neural networks is fragile. arXiv preprint arXiv:1710.10547, 2017.
[72] Patrick Hall and Navdeep Gill. Introduction to Machine Learning Interpretability. Oxford, England: Prentice-Hall, 2018.
[73] N. Tishby and N. Zaslavsky. Deep learning and the information bottleneck principle. In 2015 IEEE Information Theory Workshop (ITW), pages 1–5, April 2015.
[74] Mehmet Süzen, Cornelius Weber, and Joan J. Cerdà. Spectral ergodicity in deep learning architectures via surrogate random matrices. CoRR, abs/1704.08303, 2017.
[75] Charles H Martin and Michael W Mahoney. Implicit self-regularization in deep neural networks: Evidence from random matrix theory and implications for learning. arXiv preprint arXiv:1810.01075, 2018.
[76] Tanmayee Narendra, Anush Sankaran, Deepak Vijaykeerthy, and Senthil Mani. Explaining deep learning models using causal inference. CoRR, abs/1811.04376, 2018.
[77] Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. CoRR, abs/1610.02391, 2016.
[78] David L Donoho, Arian Maleki, Inam Ur Rahman, Morteza Shahram, and Victoria Stodden. Reproducible research in computational harmonic analysis. Computing in Science & Engineering, 11(1):8–18, 2009.
[79] Marcus R Munafò, Brian A Nosek, Dorothy VM Bishop, Katherine S Button, Christopher D Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J Ware, and John PA Ioannidis. A manifesto for reproducible science. Nature Human Behaviour, 1(1):0021, 2017.
[80] Monya Baker. 1,500 scientists lift the lid on reproducibility. Nature News, 533(7604):452, 2016.
[81] Odd Erik Gundersen, Yolanda Gil, and David W Aha. On reproducible ai: Towards reproducible research, open science, and digital scholarship in ai publications. AI Magazine, 39(3), 2018.
[82] Matthew Hutson. Artificial intelligence faces reproducibility crisis, 2018.
[83] Hans E Plesser. Reproducibility vs. replicability: a brief history of a confused terminology. Frontiers in neuroinformatics, 11:76, 2018.
[84] Bronwyn Woods. Expanding search in the space of empirical ml. arXiv preprint arXiv:1812.01495, 2018.
[85] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety. arXiv preprint arXiv:1606.06565, 2016.
[86] Paul Voigt and Axel Von dem Bussche. The eu general data protection regulation (gdpr). A Practical Guide, 1st Ed., Cham: Springer International Publishing, 2017.
[87] Preliminary report highway hwy18mh010, by the united states transportation security board. https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf. Accessed: 2019-03-04.
[88] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability and Transparency, pages 77–91, 2018.
[89] Inioluwa Deborah Raji and Joy Buolamwini. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products. In AAAI/ACM Conf. on AI Ethics and Society, 2019.
[90] Besmira Nushi, Ece Kamar, and Eric Horvitz. Towards accountable ai: Hybrid human-machine analyses for characterizing system failure. In Sixth AAAI Conference on Human Computation and Crowdsourcing, 2018.
[91] Christoph Dann, Lihong Li,WeiWei, and Emma Brunskill. Policy certificates: Towards accountable reinforcement learning. arXiv preprint arXiv:1811.03056, 2018.
[92] H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, et al. Communication-efficient learning of deep networks from decentralized data. arXiv preprint arXiv:1602.05629, 2016.
[93] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pages 308–318. ACM, 2016.
[94] Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song. Homomorphic encryption for arithmetic of approximate numbers. In International Conference on the Theory and Application of Cryptology and Information Security, pages 409–437. Springer, 2017.
[95] Fabian Boemer, Yixing Lao, and Casimir Wierzynski. ngraph-he: A graph compiler for deep learning on homomorphically encrypted data. arXiv preprint arXiv:1810.10121, 2018.
[96] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In Security and Privacy (SP), 2017 IEEE Symposium on, pages 3–18. IEEE, 2017.

AI-based Facial Recognition in Emotional Recognitions

AMIR DIRIN

Haaga-Helia University of Applied Sciences, Helsinki, Finland, amir.dirin@haaga-helia.fi

JYRKI SUOMALA

Laurea University of Applied Science, Espoo, Finland

ARI ALAMÄKI

Haaga-Helia University of Applied Sciences, Helsinki, Finland

Facial recognition is an approach to recognize a human face with the help of computer vision. The popularity of smart gadgets and advancement on the cameras capabilities have caused the concept of facial recognition to become a hot topic among academician and practitioners. Besides the tradition facial recognition in the surveillance system, commercial facial recognition system to measures emotions have nowadays become popular. These systems are often AI-based and use facial recognitions algorithms along with biometrics to map face features from an image or through a livestream to identify the motions. The aim of this paper is to study the credibility of these systems to detect emotion accurately. Humans have complex personalities and the personality often express in our facial expressions which is not necessary reflected to the emotion. For example, personal disorders such as narcissistic personal or histrionic personality disorder have different facial expressions than persons who have not been diagnosis with any disorders. The facial expressions of those persons are not representations of emotions that will be detected through the diagnostic systems. Therefore, the complement technologies and solutions are needed to make the measurement more accurate.

1. Introduction

Facial recognition based emotional measurements devices have become very popular specifically as a supplementary for usability and user experience measurement. In addition, many companies have promoting their facial recognition solution to promote sales and improve customer relationships [1]. The advancement of these devices mainly based on the significant improvements on related technologies such as HD based camera and facial recognition algorithms such as Fisherfaces [2], Local Binary Patterns Histograms (LBPH) [3], Deep Neural Network (DNW) [4], Rectified Linear Units Layer (ReLU) [4], and Convolutional Neural Network (CNN) [5]. These algorithms are widely used by industries for their facial and emotional recognitions.

The facial recognition’s application getting very popular and increasing in the industries as well as in consumer level. The facial recognitions’ based solutions have become very available even among children for example, Snapchat, which is based on computer vision, google search engine use the widely the pattern recognitions, or Facebook, which detect the face on the picture [6]. These examples combine the artificial intelligent approaches and computer visions [7] to teach the algorithm to make more accurate measurements. Many products such as iMotions, FaceReader, Deepface, pursue to measure the emotion through the facial recognitions.

The emotion’s measurement can be achieved by three main approaches, subjective, behavioral, and physiological approaches [8]. Behavioral measurements cover many approaches for measuring user behavior, for example, Facial Action Coding System (FACS) [9] and [10], which measures facial poses. Physiological measurements allow to measure emotions change, for example, autonomic nervous system [11] or detecting galvanic skin response via a sensor.

The purpose of this paper is to investigate the efficiency of the latest facial recognitions based emotional detection. This study is based on literature review in which we argue and demonstrate that human personality impact on the facial expression. The result of this study helps the practitioner to learn about the reliability and for academician a further research topic on AI based facial recognitions.

2. Related Research

2.1 Facial recognition

The term Computer Vision (CV) refers as a field of research that aim to develop proper techniques to enable computer to see and process the content of images and videos. The Computer vision is a part of Artificial Intelligent (AI). The objects detection in an image is the main task of the CV algorithm as what and where the objects are seen. Additionally, the CV algorithms must identify the properties of those identified objects, for example whether it is a face, building, or a door. In most cases we store these identified images and compare new objects. Furthermore, the CV enable use to have multiple metric on the selected objects. In the following, the major facial recognition as briefly explained. Therefore, the CV has been utilized in various sectors, such as in safety, health, security, entertainment, cars, robotics, and in sport.

Facial recognition is a subset of CV technique used by computer algorithms to identify or verify an object or a face through images. Facial recognition is a part of computer vision which have been around for many decays. The importance of facial recognition has become evident with the popularity of social media and social networking. Figure 1 presents the process of the facial recognition steps. Figure 1 presents the process of the facial recognition steps.

The process of the facial recognitions are as follow:

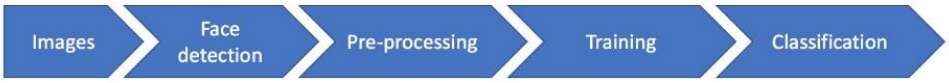


Fig. 1. Facial Recognition process

The applications of the facial recognitions are enormous such as in security system, marketing, and as an identification and clearance system. In addition to all these facial recognitions are widely applied in facial expression assessments. The facial expression and emotion measurements has already studied in 1993 [12] by Ekman. The facial experience measure is used to study the nonverbal expression and behavior of the person. By the measure we aim to identify the invoke feelings and emotions through observing the changes on the face. Ekman’s [13] Facial Action Coding System (FACS) measures the basic emotions, e.g., happy, anger, fear, and surprise. Through these measurements we evaluate the emotion reaction that invoke basic emotion that user interact with the product.

Facial Recognitions (FR) are done in two ways one verifications and the second is identification. In verification, the system compares the given object with the existing stored objects. In identification, the system identifies the object and gives a rank of the matches. In both cases, the biggest and most complex step is teaching the machine to recognize faces. The FR technology implementation consist of several stages: image acquisition, image processing, characteristic identifications, e.g., eye sockets, nose shape, template creation, and template matching [14]. Facial recognition algorithm often measures the distance between the eyes, width of the nose, depth of the eye socket, cheekbones, and chin. Many pictures are needed in the training data and the machine will have to learn how to differentiate faces. Different algorithms can be trained for that, some of them use a statistical approach or search for patterns and some others use a neural network. In the following the major facial recognition algorithms.

2.2 Emotion and Feeling

Darwin identified that emotions are product of evolution. Emotions have evolved through adaption with our surroundings. Theories of emotions in psychiatric and neuroscience research have proven that humans are equipped with basic sets of emotions [15] [16]. Each emotion is associated with psychological and physiological behaviors. Whereas traditional approaches to human higher –order cognitive processes ignore emotions, emerging decision neuroscience evidence suggests that rational decision making depends on emotional processing [17]. For example, fear is the automatic and subconscious result of unpleasant emotion and our neocortex interpret these emotional signals as conscious feelings [18]. Furthermore, Ekman & Friesen [19] studied the method to recognize the facial expressions masks which is reflected by basic emotions. Their findings indicate that the basic emotion and the facial expression are expressed in the same way universally. Habibi and Damasio [20] defined feelings as mental experiences that are connected to an activity in a certain brain region that maps body states. In alignment with this definition, Damasio [18] considered feelings as a mental representation of the physiological changes that accompany emotions.

2.3 Personal Disorder and facial expressions

Lynch et al [21] have studied the emotional sensitivity with Borderline Personality Disorder (BPD). Lynch et al demonstrated that for BPD participants the facial expressions changes from neutral to maximum very fast compare with healthy participants. Furthermore, BPD participants are more sensitive than healthy participants in identifying emotional expressions. Pelc et al [22] has conducted a study with Attention-Deficit Hyperactivity Disorder (ADHD) participant on the impact of facial expression reactions on basic emotions. Their findings indicate that there were correlation between interpersonal problems and emotional facial expression decoding impairment on anger expressions. Pelc et al [22] findings that nonverbal decoding abilities have implications during the therapy sessions for ADHD. Thomas et al. have demonstrated that children with anxiety disorder showed and exaggerated fearful faces. Marissen, Deen, & Franken [23] have showed that the person with Narcissistic Personality Disorder (NPD) perform worse on facial emotion recognition task.

The elicitation and measurement of emotions can be achieved by three approaches: subjective, behavioral, and physiological approaches [8]. We briefly describe these approaches in the following paragraphs.

Behavioral measurements cover versatile approaches that are used to measure user behavior. Two examples are the facial action coding system (FACS) [9], which measures facial poses (e.g., when we are happy, we tend to smile), and the specific affect coding system (SPAFF) [24], which measures emotions during interactions, for example, between couples.

Physiological measurements allow for identifying how the body behaves when emotions change, for example, via the autonomic nervous system [11]. An example of a physiological measurement is detecting galvanic skin response via a sensor, which may be indicative of emotions such as happiness, surprise, disgust, anger, or fear [25].

Researchers often employ subjective measurements to measure subjective behavior using instruments such as questionnaires, rating scales, and experimental sampling. Scholars have also developed systematic subjective behavior measurement approaches, including the Positive and Negative Affect Schedule [26]. In such measures, users are asked how they currently feel (e.g., nervous, scared, inspired). Other methods, like the Stress Appraisal Measure [27] , measure the user's stress level. Finally, with the help of experience sampling methods [28], it may be possible to capture people's emotions.

3. Towards a complemented approach on emotional detection

In the facial recognition testing environment, there are many emotions such as enjoyement, curiosity, interest, hope, anger, anxitym shame, confusion, frustration or even boredom are frequent to experience. Calder & Young [29] have demonstrated that faces contain social signals and identity the functional and neural levels. Attempt to detect the emotions and feeling has a long history. For many decays in psychotherapy sessions the therapists have followed the behavior and body language of the patient in addition to the verbal communications. Non-verbal communications complement the facial expressions. Kulkarni & Bagal [30] have revealed that accurate to interpret the facial expressions is critically important to consider the non-human primates that relies on non-verbal signals and communications.

The personality disorder impacts the facial expression and facial expression interpretations. For example, Surguladze et al [31] demonstrated that in major depression there are impaired facial expression. Their study has demonstrated that these group of people have different facial stimuli in sadness and in happiness and neutral expressions in compassion with healthy people. In addition to personality disorder, the culture also influences the facial expressions of a person. Furthermore, the social context and social status reflect our facial expressions as Turner and Stets [32] revealed.

The reviewed literatures indicate that mere facial recognition algorithms or physiological measurements are not sufficient to come up with an accurate emotion detection. To achieve optimal and reliable results we have to identify more complement solutions and approaches to the existing methods and tools. We are investigating to extend the existing facial recognition technology with additional technologies that help the system to detect the personality disorders. The identifications of the personality disorder may help the facial recognition and biometric system more accurately measure the emotions. The details of the new technological solution is considered as out of the scope of this paper and will be publish as soon as the efficiency of the new technology is proofed.

4. Discussion

Although facial recognition has already many advantages, such as recognizing users' emotional states in the scale of happy, neutral or unhappy, it has several limitations, reliability problems, users may feel privacy concerns.

Firstly, the emotional cues of personality disorders differ from other population making the reliability of emotional recognition problematic. Many famous actors have suffered depression although they have looked like happy and wealthy in social situation. For example, people were really surprised when very famous comedian actor Jim Carrey revealed tobe depressed even in tens of years (Mental Health Daily, 2014). This is an example about difficulty to recognize emotional states even by persons who closely follow those people years after years. Thus, we cannot recognize all emotional cues from the faces of people in a single spot, but we should analyze their behavior in different contexts as well. Additionally, if we are able to connect other biometric data to the facial information, we could improve the reliability of our analyze.

Secondly, there are still several technological limitations. The facial recognition cameras recognize the faces of people if they walk toward camera with face visible frontally, but not if they just pass the camera. For example, Allgovision technology allows only plus or minus 20 degrees tilt in both x- and y-direction. It is probable similar tolerance with the algorithms of other technology developers. The facial recognition algorithms have been developed to recognize the different targets, such as eyes, nose, mouth and their distances, moves and micro expressions on the faces. If they are scanned from the different angles the machine learning algorithms cannot receive all data needed for reliable processing. However, the facial expression recognition technologies develop fast. The facial recognition is nowadays very accurate when we are dealing with the high-quality two-dimensional images, and we are not looking for complicated or detailed results. The reports of NIST [33] states that the most accurate facial recognition algorithms will find matching entries among of 12 million individual images with error rates below 0.2%. The report also reveals that the facial recognition technologies have significantly developed from year 2013 to year 2018, even more than before. Nevertheless, those excellent results are only valid with the high-quality photos where faces are visible frontally, they are two-dimensional images and objects do not move. The report of NIST [33] did not speak anything about the accuracy of emotional detection. However, they state that although most facial algorithms cannot recognize twins from the images, there is nowadays at least one patented algorithm that is able to recognize twins.

Like many facial recognition technologies, Azure Face API [34] can recognize the gender and age of people, it finds similar faces from the catalogs or group unknown similar faces to the same group as well as it identifies a person if it already has his or her face image. Thus, the facial recognition can identify the person identity when they enter to a building or other location that is using the facial recognition. If the owner of the building or location includes the emotion recognition to the facial recognition, they are also collection information about the emotional state of a person during a timeframe. Privacy concern is an unsolved issue in developing facial recognition to the retail stores, malls and other locations where people are visiting. For example, Buzzfeed.com states the thousands of U.S. retail stores are purchasing and implementing facial recognition to the retail stores due to the security reasons. There is a short step from the identifying the persons to the scanning of their emotions. The goal of the service providers and retailers is that customers are happier when they go out than entering to the stores. For example, there is initiatives in China to identify customers when they enter to the stores and connect them their purchase, preference and network data [35]. Although, customers in other countries are more open to the privacy issues than others, it is still unclear how the facial recognition could analyze the emotions related to the purchase behavior of different customer segments. Detecting emotional purchase behavior of customers from the previous purchase history data or from the social media data is probable more reliable than purely facial recognition of emotions.

5. Conclusions

The emotional detection devices have become popular widely in the marketing sector. This study attempts to elaborate that these devices cannot be accurate enough especially for those who have a personality disorder. The emotional detection through facial expressions is a complex process which require extensive study. A simple facial gesture does not indicate or correlates with emotions.

As a future study we aim to investigate further and develop a proper solution that help us to complement the existing devices.

References

[1] S. Jyrki, “The Neuroscience Research Methods in Management. In: Moutinho L., Sokele M. (eds) Innovative Research Methodologies in Management,” 2018.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs . Fisherfaces: Recognition Using Class Specific Linear Projection,” vol. 19, no. 7, pp. 711–720, 1997.

[3] S. Moore and R. Bowden, “Local binary patterns for multi-view facial expression recognition,” Comput. Vis. Image Underst., 2011.

[4] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” Proc. 27th Int. Conf. Mach.

Learn., no. 3, pp. 807–814, 2010.

[5] L. Xu, J. S. Ren, C. Liu, and J. Jia, “Deep Convolutional Neural Network for Image Deconvolution,” Adv. Neural Inf. Process. Syst., 2014.

[6] J. P. Mello Jr., “Facial Recognition Beyond Facebook.,” PCWorld, vol. 29, no. 12, pp. 13–14, 2011.

[7] F. Abdat, C. Maaoui, and A. Pruski, “Human-computer interaction using emotion recognition from facial expression,” in Proceedings - UKSim 5th European Modelling Symposium on Computer Modelling and Simulation, EMS 2011, 2011.

[8] K. R. Scherer, “What are emotions? And how can they be measured?,” Soc. Sci. Inf., vol. 44, no. 4, pp. 695–729, 2005.

[9] P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. 1978.

[10] M. Allen, “Facial Action Coding System,” in The SAGE Encyclopedia of Communication Research Methods, 2017.

[11] S. D. Kreibig, “Autonomic nervous system activity in emotion: A review,” Biol. Psychol., vol. 84, no. 3, pp. 394–421, 2010.

[12] P. Ekman, “Facial expression and emotion,” Am. Psychol., 1993.

[13] P. Ekman, Emotions revealed: recognizing faces and feelings to improve communication and emotional life. 2003.

[14] I. Naseem, R. Togneri, and M. Bennamoun, “Linear regression for face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., 2010

[15] P. Ekman, “Basic emotions,” Cognition, vol. 98, no. 1992. pp. 45–60, 1999.

[16] K. Oatley and P. N. Johnson-Laird, “Towards a Cognitive Theory of Emotions,” Cogn. Emot., 1987.

[17] A. Bechara and A. R. Damasio, “The somatic marker hypothesis: A neural theory of economic decision,” Games Econ. Behav., 2005.

[18] A. Damasio, “Fundamental feelings,” Nature. 2001.

[19] P. Ekman and W. V Friesen, Unmasking the face: A guide to recognizing emotions from facial clues. 2003.

[20] A. Habibi and A. Damasio, “Music, feelings, and the human brain.,” Psychomusicology Music. Mind, Brain, 2014.

[21] T. R. Lynch, M. Z. Rosenthal, D. S. Kosson, J. S. Cheavens, C. W. Lejuez, and R. J. R. Blair, “Heightened sensitivity to facial expressions of emotion in borderline personality disorder,” Emotion, 2006.

[22] K. Pelc, C. Kornreich, M. L. Foisy, and B. Dan, “Recognition of Emotional Facial Expressions in Attention-Deficit Hyperactivity Disorder,” Pediatr. Neurol., 2006.

[23] M. A. E. Marissen, M. L. Deen, and I. H. A. Franken, “Disturbed emotion recognition in patients with narcissistic personality disorder,” Psychiatry Res., 2012.

[24] J. a Coan and J. M. Gottman, “The Specific Affect Coding System (SPAFF).,” Handb. Emot. elicitation assessment., pp. 267–285, 2007.

[25] W. Guanghua, L. Guangyuan, and H. Min, “The analysis of emotion recognition from GSR based on PSO,” 2010. [Online]. Available: http://ieeexplore.ieee.org/document/566325.

[26] D. Watson, L. a. Clark, and a. Tellegen, “Positive and negative affect schedule (PANAS),” J. Pers. Soc. Psychol., vol. 54, pp. 1063–1070, 1988.

[27] E. Peacock and P. Wong, “The stress appraisal measure (SAM): a multidimensional approach to cognitive appraisal,” Stress Med., vol. 6, no. 3, pp. 227–236, 1990.

[28] T. S. Conner, H. Tennen, W. Fleeson, and L. F. Barrett, “Experience Sampling Methods: A Modern Idiographic Approach to Personality Research.,” Soc. Personal. Psychol. Compass, vol. 3, no. 3, pp. 292–313, 2009.

[29] A. J. Calder and A. W. Young, “Understanding the recognition of facial identity and facial expression,” Nature Reviews Neuroscience. 2005.

[30] K. R. Kulkarni and S. B. Bagal, “Facial expression recognition,” in Proceedings - IEEE International Conference on Information Processing, ICIP 2015, 2016.

[31] S. A. Surguladze, C. Senior, A. W. Young, G. Brébion, M. J. Travis, and M. L. Phillips, “Recognition Accuracy and Response Bias to Happy and Sad Facial Expressions in Patients with Major Depression,” Neuropsychology, 2004.

[32] J. H. Turner and J. E. Stets, The sociology of emotions. 2005.

[33] P. Grother, M. Ngan, and K. Hanaoka, “Ongoing face recognition vendor test (FRVT) part 2:,” 2018.

[34] Microsoft, “Face API – Facial Recognition Software - Microsoft Azure,” Microsoft Azure, 2019..

[35] E. J. Peacock Ma and A. N. D Pault P Wong, “The stress appraisal measure (sam): a multidimensional approach to cognitive appraisal,” stress med., 1990.

Using Google Cloud AI/ML API's With News Media Websites

BRUNO AMARO ALMEIDA
Futurice, Annankatu 34 B, Helsinki, Finland, hello@brunoamaro.com

1. Introduction

Today, a wide range of people and organizations worldwide (e.g. universities, companies, governments, etc.) actively use the public cloud. While there are dozens of public cloud vendors, three companies are currently the undisputed global market leaders: Amazon Web Services, Google Cloud and Microsoft Azure.

AI and Machine Learning are key areas of investment, growth and differentiation for many organizations and that is no exception for the three major public cloud vendors (AWS, GCP and Azure). In this context, pre-trained AI/ML API's in combination with other Serverless services is one area that has been on the rise and with fast adoption. While all of these vendors provide very interesting functionalities in their AI/ML API's offering, Google Cloud has been standing out among them.

News media websites are - and have been in the past years - in the epicenter of multiple society debates (e.g. fake news, public elections and geo-political influence, monetization and clickbait, etc.). Using simple software engineering techniques is fairly easy to periodically collect and extract metadata information from different news media websites (fake news or credible sources), and therefore allow us to gather interesting insights and draw some conclusions.

However, with the AI/ML API's available today this could be taken one step further. By leveraging those ready-made capabilities (e.g. Image Classification, Translation, Sentiment Analysis), it is possible to enrich this metadata and gain valuable information and insights from the powerful pre-trained Google Cloud AI/ML in a matter of milliseconds and without any kind the engineering or data science heavy lifting.

What can Google Cloud AI/ML APIs tell us about news media websites?

2. Public Cloud Landscape

In the present day, it is impossible to talk about Public Cloud – in particular, infrastructure and platform as a service – without referring the names Amazon Web Services, Google Cloud and Microsoft Azure. Together, these three companies dominate the market and are considered by Gartner (Fig. 1.) as the Leaders of the pack.

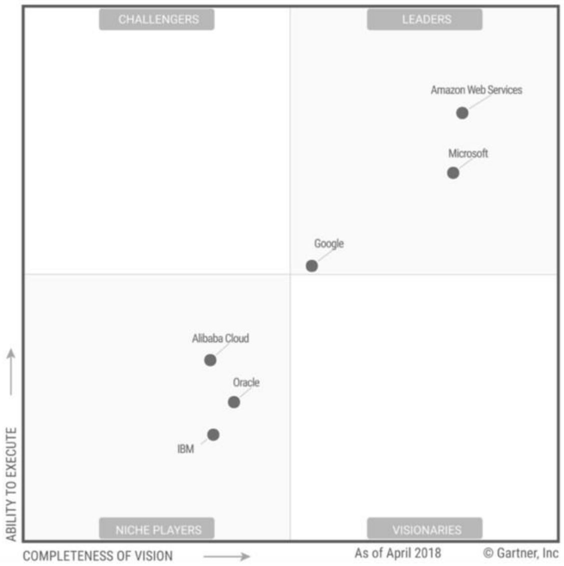


Fig. 1. Gartner Magic Quadrant for Cloud (May 2018)

It is equally important to realize that 2018 was the first year were Google Cloud was considered as a Leader (previously Visionary) [1]. The market has clearly been consolidating among those three vendors and, to that extend, Garter went as far as drop 8 of the 14 vendors that were on the 2017 Magic Quadrant [1].

3. AI/ML as part of the Public Cloud Offering

AI and Machine Learning are key areas of investment, growth and differentiation for the three biggest public cloud players (AWS, GCP and Azure). While their service offering varies wildly in those areas, they can always be aggregated into three distinct groups: Data Engineering, AI/ML Platform and AI/ML API's.

The *Data Engineering* service offering helps to ingest, prepare, transform and analyze data. While this is not AI (or Machine Learning) per se, these services are the backbone of what you are enabled to do in terms of AI/ML.

The *AI/ML Platform* service offering helps to build, train and deploy Machine Learning models. Often, they rely on one or multiple known open source frameworks. One of those, common across the three vendors is Tensorflow.

The *AI/ML API's* service offering allows you to take advantage of different pre-trained models provided to anyone out-of-the-box at the distance of a simple API call.

In this context, pre-trained *AI/ML API's* in combination with other *Serverless* services is one area that has been on the rise and with fast adoption. Without any prior AI knowledge, it allows us to leverage ready-made capabilities such as: Text to Speech, Image & Video Classification, Translation, Speech Recognition, Sentiment Analysis, etc.

While the three public cloud vendors (AWS, GCP and Azure) offer these AI/ML API's services, Google has been a bit ahead of the curve by spearheading and bringing disruption and innovation. What I find particularly appealing in their AI/ML API's offering is the maturity of the service (clearly leveraging the internal knowledge from many years as a technology powerhouse) combined with a wider range of languages (not so English-centric).

4. News Media Websites (Fake vs Credible)

News media websites are - and have been in the past years - in the epicenter of multiple society debates (e.g. fake news, public elections and geo-political influence, monetization and clickbait, etc.). The term "fake news" took the World by storm. It was at the center stage of the 2016 U.S. presidential election and moreover with the growing role of Social Media in the World geo-politics ever since.

The impact of these fake stories compared with real stories from credible and accredit news outlets can be tremendous. During the 2016 U.S. presidential elections, BuzzFeed looked at the top 20 fake news stories, compared them to the top 20 election stories from 19 major media outlets and discovered that the fake ones got more engagement on Facebook [2].

The motivations behind these fake stories vary, but one of the major drivers seems to be financial gain – generated from the massive amounts of internet traffic (ads) the sites can get. An investigation by Wired magazine *"Inside The Macedonian Fake-News Complex*" [3] portraits well some of the real-world stories behind it.

5. Website Metadata Extraction Methods

Extracting Metadata Information from Websites is not a new topic in Software Engineering. With the wide range of tools and techniques available today, it is fairly easy to collect and extract information from any website.

There are many well-known methods to extract metadata from websites. Inclusive, one can find entire business services dedicated to metadata extraction by very sophisticated methods. However, by using some of the built-in capabilities of a Linux system combined with some open source tools it is possible to get quick results. To a given website, the data will be retrieved in three simple ways: taking a screenshot of the webpage; saving the text output and last by saving all the loaded images.

5.1 Saving the Text Output

Lynx is a test-based web browser widely popular in Linux systems. It can used with the flag --dump to extract all the text of a particular website.

Example:

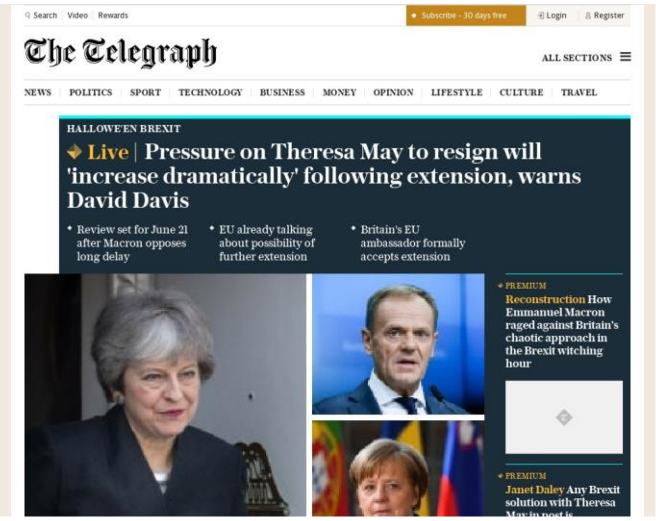
```
lynx --dump $TARGET_SITE > $RESULTS/out.txt
```

5.2 Website Screenshot

A combination between *xvfb-run* and *wkhtmltoimage* Linux utilities allow us to capture a screenshot of a given website (Fig. 2). *xvfb-run* is an in-memory display server for Linux. It enables running graphical applications without a display while also having the ability to take screenshots. *Xvfb-run* can be used to run *wkhtmltoimage*, which is a tool that renders HTML into PDF and various image formats.

Example:

```
xvfb-run --server-args="-screen 0, 1280x1200x24" wkhtmltoimage --quality 100 --crop-h 800 $TARGET_SITE
$RESULTS/out.png
```



Live | Pressure on Theresa May to resign will 'increase dramatically' following extension, warns David Davis

- Review set for June 21 after Macron opposes long delay
- EU already talking about possibility of further extension
- Britain's EU ambassador formally accepts extension

Reconstruction How Emmanuel Macron raged against Britain's chaotic approach in the Brexit witching hour

Janet Daley Any Brexit solution with Theresa May in post is

Fig. 2 Example output from xvfb-run combined with wkhtmltoimage

5.3 Saving all the loaded images

ImageScraper is a Linux tool that as the name suggest, scrapes a given website and downloads all the images available (Fig. 3).

Example:

```
image-scraper $TARGET_SITE
```

Fig. 3 Example output from ImageScraper

6. Enrich website metadata with AI/ML API

With the three types of website metadata (screenshot, text, child images) already extracted, it is possible to leverage the Google AI/ML API's to gain more insights.

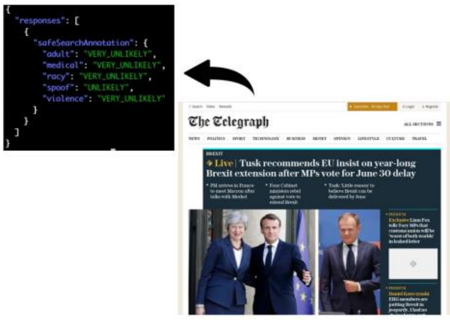
The list of available API's and their features is quite interesting and rich: Text to Speech, Image & Video Classification, Translation, Speech Recognition, Sentiment Analysis, Conversational Bots, etc.

To this use case, two of them really stand out: Google Cloud Vision AI and Natural Language API's.

6.1 Vision AI

Vision AI provides image analysis and classification functionalities. It uses pre-trained AI models by Google to detect labels, recognize individual objects, faces, and words.

Given that it was extracted a screenshot of the website (a mixed and complex image), the Vision API can be used with different detection features such as: labels, text (OCR), safe search annotations and web entities (Fig. 4).



```
{
  "responses": [
    {
      "safeSearchAnnotation": {
        "adult": "VERY_UNLIKELY",
        "medical": "VERY_UNLIKELY",
        "racy": "VERY_UNLIKELY",
        "spoof": "UNLIKELY",
        "violence": "VERY_UNLIKELY"
      }
    }
  ]
}
```

The Telegraph
Live | Tusk recommends EU insist on year-long Brexit extension after MPs vote for June 30 delay

Fig. 4 Example of Google Vision API (Safe Search) output

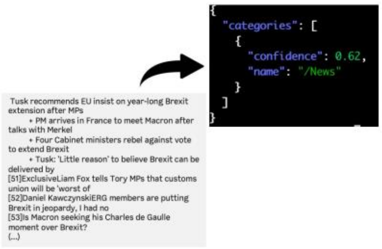
An example of the commands can be found below:

```
gcloud ml vision detect-labels $RESULTS/out.png > $RESULTS_RAW/labels.json
gcloud ml vision detect-text $RESULTS/out.png > $RESULTS_RAW/text.json
gcloud ml vision detect-safe-search $RESULTS/out.png > $RESULTS_RAW/safe-search.json
gcloud ml vision detect-web $RESULTS/out.png > $RESULTS_RAW/web.json
```

Also, since the child images that belong to the website were extracted, the Vision AI API can be used to detect the labels that are associated with each one of them.

6.2 Natural Language API

The Natural Language API uses pre-trained models by Google to reveal the structure and meaning of a text. Applied to this use case, it provides text classification (Fig. 5) and sentiment analysis of the extracted website text output.



Tusk recommends EU insist on year-long Brexit extension after MPs
- PM arrives in France to meet Macron after talks with Merkel
- Four Cabinet ministers rebel against vote to extend Brexit
- Tusk: 'Little reason' to believe Brexit can be delivered by

[51]ExclusiveLiam Fox tells Tory MPs that customs union will be 'worst of
[52]Daniel KawczynskiERG members are putting Brexit in jeopardy, I had no
[53]Is Macron seeking his Charles de Gaulle moment over Brexit?
(...)

```
{
  "categories": [
    {
      "confidence": 0.62,
      "name": "/News"
    }
  ]
}
```

Fig. 5 Example of Google Natural Language API (Text Classification) output

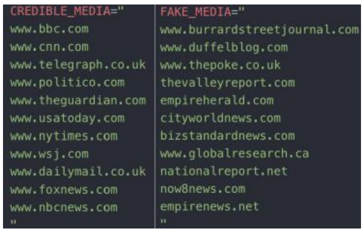
Analyzing text sentiment is interesting and not so trivial to interpret the results at first glance. The response contains the overall text sentiment of the text provided split in two fields: score and magnitude.

Score is a numerical value that ranges between -1.0 (negative) and 1.0 (positive) and corresponds to the overall emotional leaning of the text.

Magnitude on the other hands, indicates the strength of the emotion (both positive and negative) within the text. It ranges between 0.0 and +inf. and, unlike score is not normalized. Each expression of emotion contributes to the magnitude, so longer text blocks may have greater magnitudes.

7. Analysis & Results

Selecting credible and fake news media websites to this analysis was not a trivial task. Fake media websites are especially hard because the domain names tend to be short lived. One good source was Opensources.co [5], a research group that curates an opensource database of websites and provides fine-grain classification for them, ranging from conspiracy theory to hate news. Other good source was Politifact [6], a group that does fact-checking of news stories that go viral on the internet. For credible news media websites, the selection criteria were the most popular and accredited news websites from U.K. and U.S. In the end, we have two lists of 11 websites each (Fig. 6).



```
CREDIBLE_MEDIA="
www.bbc.com
www.cnn.com
www.telegraph.co.uk
www.politico.com
www.theguardian.com
www.usatoday.com
www.nytimes.com
www.wsj.com
www.dailymail.co.uk
www.foxnews.com
www.nbcnews.com
"
```

```
FAKE_MEDIA="
www.burrardstreetjournal.com
www.duffelblog.com
www.thepoke.co.uk
thevalleyreport.com
empireherald.com
cityworldnews.com
bizstandardnews.com
www.globalresearch.ca
nationalreport.net
now8news.com
empirenews.net
"
```

Fig. 6 List of Credible and Fake Media Websites used for the analysis

For each of those websites, metadata will be extracted - screenshot (Fig.7), child images and text output - and enriched with the Vision AI and Natural Language API's.

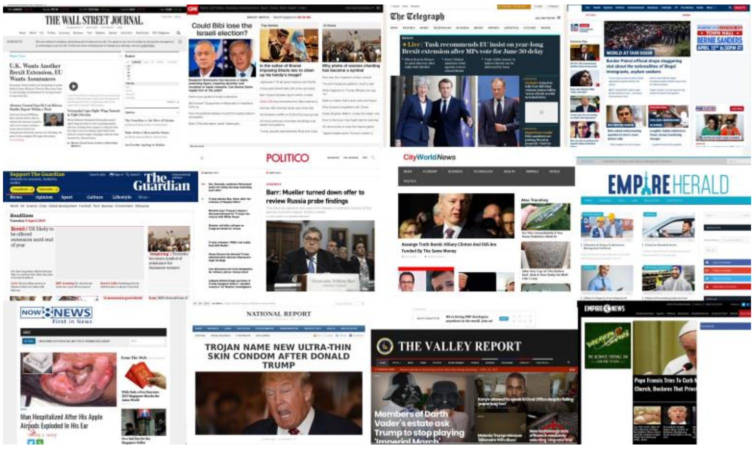

THE WALL STREET JOURNAL.
U.K. Wants Another Brexit Extension, EU Wants Assurances
Could Bibi lose the Israeli election?
The Telegraph
Live: Tusk recommends EU insist on year-long Brexit extension after MPs vote for June 30 delay
BERNIE SANDERS
The Guardian
POLITICO
Barr: Mueller turned down offer to review Russia probe findings
CityWorldNews
EMPIRE HERALD
NOW NEWS
Man Hospitalized After His Apple Airpods Exploded In His Ear
NATIONAL REPORT
TROJAN NAME NEW ULTRA-THIN SKIN CONDOM AFTER DONALD TRUMP
THE VALLEY REPORT
Members of Darth Vader's estate ask Trump to stop playing 'Imperial March'
EMPIRE NEWS
Pope Francis Tries To Curb Church, Declares That Priests

Fig. 7 Multiple Screenshot captures from credible and fake media websites

To analyze the results, the outcome (enriched metadata) can be aggregated by fake and credible news media websites.

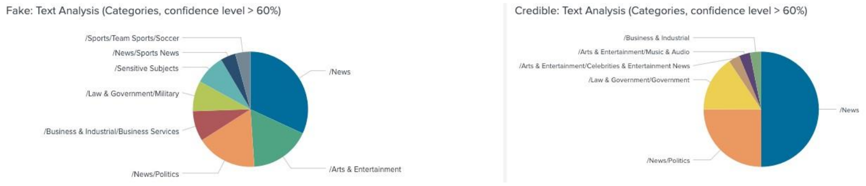


Fig. 8 Categories (Text Analysis) with Confidence Level > 60% for both Fake and Credible Media Websites

The categories returned by the Natural Language API (Text analysis) gave some interesting insights (Fig. 8). For the credible websites, it returned as expected a majority of texts labeled as *'News'* or *'News/Politics'* – this was a week dominated by Brexit in the news cycle. For fake websites, the results were very mixed. While a big portion was also labeled as *'News'* or *'News/Politics',* there were a few other categories that didn't appeared with the credible websites: '*Sensitive Subjects'* and '*Sports*'.

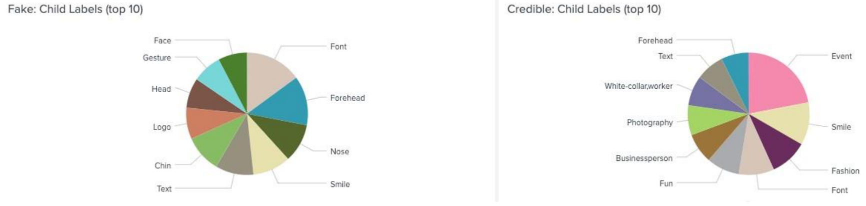


Fig. 9 Top-10 Labels (from child website images) for both Fake and Credible Media Websites

By running all the extracted child images with the Vision AI API, it was detected a big number of labels (Fig. 9). The fake news websites seem to have a big number of images related to people, therefore the labels are mostly about upper body parts (Face, Forehead, Nose, etc). Credible news labels are in contrast more mixed and with a big portion about Business, White Collar workers and Events.

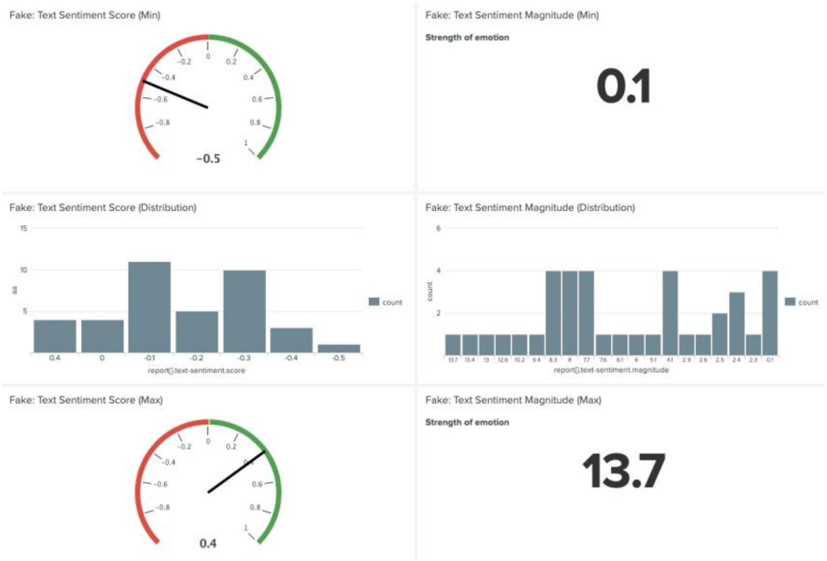


Fig. 10 Text Sentiment Score & Magnitude for Fake Media websites

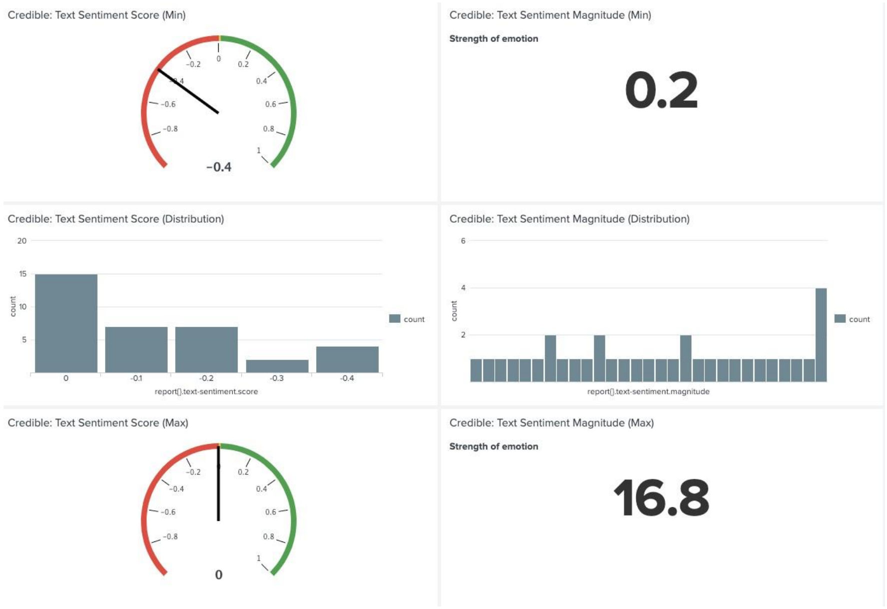


Fig. 11 Text Sentiment Score & Magnitude for Credible Media websites

The results from the Natural Language API regarding the sentiment of the text were at first glance very similar (Fig. 10 and 11). Yet, there are some differences that are worth pointing out. Fake media sentiment score was mostly negative between -0.1 and -0.5, with a lot of high magnitude scores. Credible media sentiment also had a lot of negative scores, yet a big majority was classified as neutral (i.e. value 0). The distribution of the magnitude values was very spread among high values.

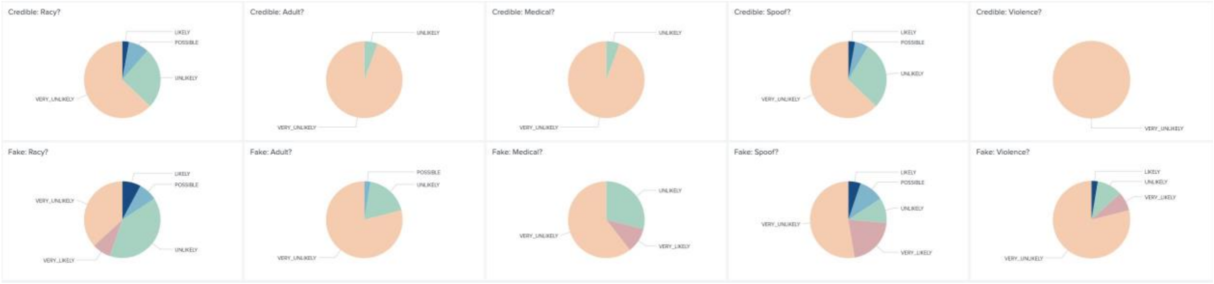


Fig. 12 Safe Search Annotation (from Screenshot images) for both Fake and Credible Media Websites

The results from the Vision AI API regarding Safe Search Annotations were perhaps the most interesting and revealing (Fig. 12). These insights came from analyzing the screenshot of each website. There were clearly some areas where the fake media websites really stand out: Racy (Possible, Likely, Very Likely), Medical (Very Likely), Spoof (Possible, Likely, Very Likely) and Violence (Very Likely). On those same areas, credible media websites did in general quite well, with the exception of Racy (Likely, Possible) and Spoof (Likely, Possible). While none of these areas were majorities (percentage wise), they do indicate that there are big differences (visually) between credible and fake news media websites.

8. Conclusion & Next Steps

This was a project sparked by curiosity. From an engineering perspective, I wanted to see what type of insights could be gained by applying Google pre-trained ML models to different types of metadata. Also, from a data scientist point of view, would those insights be valuable in solving a real-world complex problem such as credible vs fake media?

It can't be said that these methods would be able to distinguish a single website between fake or credible. However, by aggregating the insights gathered from multiple fake and credible websites over the course of a few days it was possible to clearly distinguish them.

Regarding next steps, there are definitely some interesting possibilities to explore further. Typically, credible news cycles have a dominant story in the headlines for a few days (this week Brexit). Therefore, extending the data collection period from days to some months could lead to deeper understanding and some interesting findings. Other possibility would be to use more advanced methods to extract the initial metadata from the websites. Having a good quality and filtered data is quite important before using the AI/ML API's.

Lastly, an interesting research avenue to purse would be to apply the same concept to other public vendors AI/ML APIs such as Amazon Web Services and/or Microsoft Azure. It would allow additional information to be discovered plus, it would create a very interesting comparison and result validation.

References

[1] Gartner, “Magic Quadrant for Cloud Infrastructure as a Service, Worldwide”, 2018, https://www.gartner.com/doc/reprints?id=1-2G2O5FC&ct=150519&st=sb

[2] ABC News, "When Fake News Stories Make Real News Headlines", 2016, http://abcnews.go.com/Technology/fake-news-stories-make-real-news-headlines/story?id=43845383

[3] Wired, “Inside The Macedonian Fake-News Complex”, 2017, https://www.wired.com/2017/02/veles-macedonia-fake-news/

[4] Google, “Google Cloud Natural Language API – Sentiment Analysis” https://cloud.google.com/natural-language/docs/basics#sentiment-analysis-values

[5] OpenSources.co, “Curated resource for assessing online information sources”, http://www.opensources.co/

[6] Polifacts, “PolitiFact's guide to fake news websites and what they peddle”, 2017, https://www.politifact.com/punditfact/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/

Digital Amnesia and Personal Dependency in Smart Devices: A Challenge for AI

AMIR DIRIN
Haaga-Helia University of Applied Sciences, Ratapihantie 13, Helsinki, Finland amir.dirin@haaga-helia.fi

ARI ALAMÄKI
Haaga-Helia University of Applied Sciences, Ratapihantie 13, Helsinki, Finland

JYRKI SUOMALA
Laurea University of Applied Science, Vanha maantie 9, Espoo, Finland,

The terms "digitalization" and "Artificial Intelligent" have become a buzzword in the contemporary life. We read, see, and hear these words almost in all media in a daily basis. Digitalization have affected our life significantly, such as faster accessing to information, always connected to peer, and social networking. This article however, focus to the negative impact of the digitalization in our life. The downside of the digitalization has been impacted to all sectors from children up to elders. For example, children even under 3 years old staring at the mobile phones or tablet screen for many hours during a day. Similarly, their parents themselves are occupied with other applications, e.g., messaging, Facebook, and whatsApp. The authors believe that the forthcoming AI based applications, e.g., social media, entertainments, and gaming application will increase the dependency even more. The aim of this paper is to elaborate the dependency consequences to an individual from the psychological and sociologically perspective. This paper is a based on the literature review and the main contribution is to raise awareness about the individual impact of the digitalization.

Keywords: Digitalization, Digital Amnesia, Artificial Intelligent

1. Introduction

The technological revolutions have brought to humanity lots of comfort and tranquility. The human life expectancies have significantly increased in comparison with the previous century. Among all these benefits, we have witness many downward impact to our life the undoubtable example is the pollutions that for example cars brought to us, environmental damages. The digitalization, however have brought us different kind of the individual behavior changes.

The popularity and the penetrations of the smart gadgets to our life have been unconsciously substantial. We have been driven to the situation where not remembering love ones´ phone number or our schedule for the next day. These types of behavior is the results of having a personal assistance call smartphones with us all the time. These types of dependency to smart devices a has already resulted in a behavioral implications [1] and [2] in addition to physiological impacts such as digital amnesia [3] and digital fiction [4]. The mutation in our lifestyle is already visible and obvious specifically among children who play with a gadget in daily bases often for long hours. The main concerning issue is that the new generation are losing the human touch [5], sympathy, and empathy [6] that is developed often during childhood and always have been promoted as positive humanistic features. The development trends indicate that the application dependency will be increasing significantly especially with the help of AI and neural psychology by engaging users emotionally and anticipate user needs more accurately.

The personal dependency is not restricted to the smart devices as a personal assistance. The vast amount of information that this device produces through mobile applications are yet another challenges. The resulted information are not often useful but harmful [7] especially for children. We deal with unwanted and mislead information in regular bases from social media [8] such as Facebook, WhatsApp's, telegram, twitter, and Instagram. This information has become another source of human dependency [9] to the device that lead to behavioral changes [10] that human previously has not experienced.

In the following sections, we elaborate the related researches on the field and then construct bases that the dependency to smart devices has occurred unlike to Personal Computers (PC), and Tablet. Finally, in discussions we argue the significance of AI based digitalization in individual level from psychology, sociology, and physiological impacts.

2. Research Question and research methodology

The main research that this study purse to elaborate is as follow.

Does the digitalization resulted digital amnesia?

By answering the following sub-questions, we pursue to answer the main question.

Does the digitalization has resulted personal dependency to an external mean? Through this question, we seek to reveal the changes the way people experience reading material dependency.

Does the digitalization make us more secure and trustworthy to our surroundings? With this question, we aim to reach whether the digital generations have more secure environment secure life, secure feeling.

Does the digitalization has resulted a social disconnect? This question helps us to explore in an individual level people withdraw from the society and rather to isolated.

This is an introductory study on the impact of digitalization on an individual level. Despite of long coexistence of smart devices in our life there have not been in-depth study on the physiological impacts. This is a literature review based study, which we pursue to define the problems domains in digitalization from an individual level. The results of this study helps to define existing academics gaps for further researchers on the field.

3. Literature Review

The term Artificial Intelligent (AI) has become a buzzword across industries, politics, and academics nowadays. AI has already being coexists as potential solutions for over four decays. The aim of AI is to make the computers to think and behave the way that human thinking and behave [11]. For example making a decision when is needed, or learn new ways to handle a specific task or situation. The first computer-learning program is already written in 1952 and following that, the development pattern recognition algorithm happened in 1967 using nearest neighbor algorithm and have been evolving ever since. The aim of AI is to perceive, reason and act accordingly based on the pre-defined algorithm. The subsets of AI are expanding by the advancement of new technologies, e.g., processing powers, memories, and programing capabilities. The widely use AI subsets are machine learning, computer vision, robotic, expert system, and neural network. Machine learning is an AI solution that allow machine to learn to perform a task without specifically being programed in a specific context. To be able to perform task the machine needs to analysis relevant data and train itself accordingly.

3.1 People dependency and Personal Computers (PC)

These types of benefits and impediment also applies in digitalization and smart devices. The first programmable computer was introduced already in 1936 by invention of transistor in 1947 by Bell Telephone company [12] personal computers become more powerful than before. These developments continue until 1953 where the IBM international Business Machine came out with first personal computer. The popularity of personal computer reached to its peak when IBM and Apple introduced the personal computer in 1974-1977. The main users of these devices mainly were the professionals on the fields at the early phase and then expert users such as programmers and end up with ordinary people who have not even education on the field. Personal dependency was not the main concern, as people tend to use these devices in specific and dedicated tasks. But, the usefulness of these devices in people life were questioned by many researchers, e.g., Yoon [13] asked the productivity of computers in the troubled with computers: usefulness, usability, and productivity. Despite some research such as Subrahmanyam [14] the impact of the computer use on children’s and adolescent’s development. However, rarely if ever the concept of addiction or dependency were the main concern.

3.2 People dependency and Tablet

Tables have been around since 1970, e.g., Dynabook and evolved ever since and reached to its peak in 2020 with Apple iPAD which surprised the market by Steve Jobs [15]. iPAD fast capture the attentions of customer specially youth with enormous entertainment capabilities encourage other manufacture to stablish a new product tablet development line.

3.3 People dependency and smart gadgets and internet

With the popularity of smartphones and the accessibility, and affordability of the internet on those devices. People utilize these devices in various purposes. Unlike the other two major smart devices, people are keener to their gadget to the extend in which they develop an emotional bond [16] which may lead to the addition [17]. Internet Addiction (IA) become a major concern since it is on the rise, which brings many social and psychological challenges. Despite the fact, that there is still some ambiguity in internet addiction definitions, for example, some researchers believe that the excessive use of the internet is considered as behavioral addiction while others suspect that addiction is not applicable for social networking and chat application in mobile and internet. In this article, the term Internet Addiction (IA) refers to as an excessive internet usage.

4. Discussions

The smart gadgets such as smart phones, smart table, smart TV, smart classroom, smart society, and many more smarts appliances have surrounded us. These devices without a doubt has affected positively our life. The service and contribution of the smart devices have improved the quality of our life, impact on health system and medical offering such as personalized medicine [18].

However, there were many researchers, who were skeptical about the new trend in digitalization. Rintala and Suolanen [19] expressed their concern on fast development of the technology which impacts the experts competence development. Technologies have entered to our life faster than people expected and have time to digest.

Does the digitalization has resulted personal dependency to external means such as devices or virtual environments?

Personal dependency is a natural phenomenon that is considered as an essential human behavioral for survival, for example, the dependency of the child to parents. However, when comes to the external stimuli the excessive use and dependency is called addiction. Cerniglia et al [20] defined internet addiction as a non-chemical, behavioral addiction that involves human-computer interaction. Lee et al. [21] show that smartphones dependency and anxiety have been increasing in South Korea significantly. Their study demonstrates that smartphones dependency has a correlation with anxiety.

The smart phones dependency have also impact to the behavioral changes for example Harun et al. [22] demonstrated that smartphones dependency has influence on the purchasing behavior. Every type of psychologically and sociologically addiction is harmful.

Although the prior research shows that significant part of individuals are addicted and dependent on their mobile devices, there is an opposition trend. In tourism, there is a growing trend “no smartphones allowed” where tourists leave their smartphones to the tourism organizers who safeguard it and they receive basic phones without cameras and Internet- access [23]. This helps tourists to concentrate themselves and focus on beautiful views, emotional experiences and novel situations in the tourism location without continuous need to interact with social media applications. Thus, users cannot only rest from the everyday duties, but they also receive a break from the smartphone usage.

A study conducted in USA [24] with (n=1605) adults between 18-54 years reveals that 21% of the participants wake up in the middle of the night to read the updates which 39% identify themselves as Facebook addict.

The results of IA is on the rise, for example, divorce, impact on the task performance at workplace, loneliness, concentration problems, and physical problems such as obesity, eyes.

Does the digitalization make us more secure and trustworthy to our surroundings?

Despite this wealth of positive contributions, we also experienced significant among of negative experience such fraud, criminal acts, cyberbullying, spying, pornography, gambling, and cyber racism have been a painful experience especially in our society. The internet criminal [25] activities nowadays are a huge business and they target specifically to vulnerable people due to the lack of competence and awareness on internet danger. Furthermore, it has become much harder to have personal privacy in the digital world [26].

Does the digitalization has resulted a social disconnect?

There are huge tendency among people to communicate and socialize through social media, e.g., Facebook[27] and digital mean than face-to-face meetings. These has leaded to disconnection from the human touch and further isolations King et al [28] demonstrates that virtual environment have become a safe zoon for psychiatric disorder patient.

Salehan and Negahban [29] show that the size of mobile social network correlates to the usage activity of social media application predicting the higher level of mobile usage intensity. Thus, users wanted to be connected to their social network, and the larger social network was, more they spent time with the smartphones. However, they also found that the usage of mobile social networking applications significantly correlates to the mobile addiction. This example points it out that users have closer connection to their social networks in using mobile devices. Nevertheless, it does not indicate anything about the quality of social interaction. It only shows how much users are using smartphones for connecting to the friends. Additionally, it shows relationship of addiction and smartphone usage activity. Ahluwalia [30] reports the survey of an insurance company who surveyed 2000 social media users. They found that 73 per cent of them find others' holiday posts annoying, These studies reveal that although users want to keep their social media updated and they want to share their experiences with their friends and followers, most the them think it as irritating activity. This kind of interaction in the social networks do not necessary generate positive emotions but makes others jealous, annoying or frustrated. In psychological terms, users are sharing each other holiday posts that deliver audio-visual information that causes unintentionally negative feelings in their counterparts. Instead of sharing great emotional experiences, they might share negative cues to the social networks.

We should discuss the role of subjective norm in the context of digital amnesia and addiction. For example, Arpaci [31] found that subjective norms effect significantly to the attitude to use of mobile cloud services of students. The subjective norm is a variable that affects to the usage behavior of smartphone users. It is a social pressure to use digital devices and its application. It probable also affects to the social media usage and sharing of holiday posts as friends and other followers behave similarly.

Does the digitalization result digital amnesia?

Literature review reveals that among all the digital devices that we have had during last three decays, smart phones and the associated application are the only devices that people are attached emotionally. Korucu and usta [32] demonstrates that the attachment and dependency is the result of applications and features such as social media, calendar, and individual applications such as banking capabilities. These dependencies often occur overtime. These features and services help users have instance access to the needed information at any time and any places. The instance access has resulted that the information transfer from short-term memory in brain to smart devices. Therefore, information regarding next meeting or the phone numbers have been much convenient to have on the phone. In addition, people tend to spend time with smart phones on chatting through social media, spending time in Facebook, or interacting with peers shared video in Instagram, which lead to lead to the social dependency [32].

Greenwood and Quin [3] have demonstrated that digital amnesia implicate the businesses such as tourism industries as people need to recall their experience based on the digital reimagining of the visit. Furthermore, Başaran [33] demonstrate that the digitalization has brought new culture and language for communication such as *ruok "are you ok?"* Additionally, the cultural and communication transformations have resulted a strong challenge on collaboration practice by youth.

It is obvious the digital digitalization has brought some emotional bonding and dependency specifically through smart phones. The main question remains intact how artificial intelligent may overcome the language, cultural and emotional dependency. Is AI provide a mean that the raise awareness about the dependency and use of personal use of memory beside smart devices. The latest trend on the existing applications such as Facebook, Instagram, and WhatApp indicate that the dependency to the devices and application will be increased. Furthermore, the neural psychology principle has been and will be employed to ensure the dependency.

5. Conclusions

This study shows that the prior research on digital amnesia is very scarce. This study extended the review to the downside value factors, namely the negative consequences of smartphone and their application usage. The individual who experiences negative feelings as the consequences of someone's smartphone usage might be also his or her friends. However, it is a line drawn in water when experience is negative, neutral or positive as experiences are situational and contextual phenomena.

In addition to digital amnesia, we concentrated the concepts of personal dependency, addiction and negative social outcomes in this study. Personal dependency causes addiction that becomes negative dependency if it is harmful for the managing everyday personal duties. Examples of harmful consequences are social disconnection, vulnerability of security or unintentional irritating social cues in social media.The AI is still in its infancy in terms of affective computing and emotional recognition. A challenge of AI is in its ability to behave intuitively in recognizing psychological patterns [34]. Smartphone is already excellent personal assistant in many areas but we still are far from "the augmented human" who are used to live in peaceful symbiosis with the digital world. Maybe this will never happen due to the continually increasing production of addictive audio-visual content and human's primitive motives and behavior models. This sets however several future research themes how to develop AI and machine learning to recognize negative consequences of the usage of digital solutions and their content. The affective computing, where AI has an ability to adopt the principles of intuitive psychology in recognizing emotions and in creating reliable analyzes from various biometrics and behavioral models call significantly more research.

References

[1] J. Harwood, J. J. Dooley, A. J. Scott, and R. Joiner, "Constantly connected - The effects of smart- devices on mental health," Comput. Human Behav., 2014.

[2] C. V. Priporas, N. Stylos, and A. K. Fotiadis, "Generation Z consumers' expectations of interactions in smart retailing: A future agenda," Comput. Human Behav., 2017.

[3] C. Greenwood and M. Quinn, "Digital amnesia and the future tourist," J. Tour. Futur., 2017.

[4] A. Bell, A. Ensslin, and H. K. Rustad, Analyzing digital fiction. 2013.

[5] J. D. Elhai, J. C. Levine, R. D. Dvorak, and B. J. Hall, "Fear of missing out, need for touch, anxiety and depression are related to problematic smartphone use," Comput. Human Behav., 2016.

[6] E. T. Vieira and M. Krcmar, "The Influences of Video Gaming on US Children's Moral Reasoning About Violence," J. Child. Media, 2011.

[7] Royal Society for Public Health (RSPH), "Social media and young people's mental health and wellbeing," 2017.

[8] E. W. T. Ngai, S. S. C. Tao, and K. K. L. Moon, "Social media research: Theories, constructs, and conceptual frameworks," Int. J. Inf. Manage., 2015.

[9] X. Han, W. Han, J. Qu, B. Li, and Q. Zhu, "What happens online stays online? —— Social media dependency, online support behavior and offline effects for LGBT," Comput. Human Behav., 2019.

[10] T. T. Goh, Z. Xin, and D. Jin, "Habit formation in social media consumption: a case of political engagement," Behav. Inf. Technol., 2019.

[11] A. Dirin and M. Linnaalmi, "Valtaako tekoäly tunteet?," ESignlas, 2019. [Online]. Available: https://esignals.haaga-helia.fi/2019/03/19/valtaako-tekoaly-tunteet/.

[12] R. A. Allan, History of the Personal Computer: The People and the Technology. 2001.

[13] Y. J. Yoon and T. K. Landauer, "The Trouble with Computers: Usefulness, Usability, and Productivity," South. Econ. J., 2006.

[14] K. Subrahmanyam, P. Greenfield, R. Kraut, and E. Gross, "The impact of computer use on children's and adolescents' development," J. Appl. Dev. Psychol., 2001.

[15] D. A. Norman and J. Nielsen, "Gestural Interfaces: A Step Backward In Usability," interactions, 2010.

[16] C. A. Hoffner, S. Lee, and S. J. Park, "'I miss my mobile phone!': Self-expansion via mobile phone and responses to phone loss," New Media Soc., 2016.

[17] A. Dirin, "Internet Addiction, new challenges for educational institutes," eSignals, Helsinki, 2017.

[18] L. Hood, R. Balling, and C. Auffray, "Revolutionizing medicine in the 21st century through systems approaches," Biotechnol. J., 2012.

[19] N. Rintala and S. Suolanen, "The Implications of Digitalization for Job Descriptions, Competencies and the Quality of Working Life," Nord. Rev., 2017.

[20] L. Cerniglia, F. Zoratto, S. Cimino, G. Laviola, M. Ammaniti, and W. Adriani, "Internet Addiction in adolescence: Neurobiological, psychosocial and clinical issues," Neuroscience and Biobehavioral Reviews, vol. 76. pp. 174–184, 2017.

[21] K. E. Lee et al., "Dependency on smartphone use and its association with anxiety in Korea," Public Health Rep., 2016.

[22] A. Harun, L. T. Soon, A. W. Mohd Kassim, and R. S. Sulong, "Smartphone dependency and its impact on purchase behavior," Asian Soc. Sci., 2015.

[23] M. Abadi, "Business Insider Nordic," Business Insider Nordic, 7AD. [Online]. Available: https://nordic.businessinsider.com/millennials-travel-cell-phone-off-the-grid-2018-3.

[24] N. Abhijit, "Facebook Addiction," 2012. [Online]. Available: http://www.buzzle.com/articles/facebook-addiction.html.

[25] FBI, "2014 Internet Crime Report," Fed. Bur. Investig. Internet Crime Complain. Cent., 2014.

[26] Bélanger and Crossler, "Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems," MIS Q., 2017.

[27] G. Seidman, "Self-presentation and belonging on Facebook: How personality influences social media use and motivations," Pers. Individ. Dif., 2013.

[28] A. L. S. King, A. M. Valença, A. C. O. Silva, T. Baczynski, M. R. Carvalho, and A. E. Nardi, "Nomophobia: Dependency on virtual environments or social phobia?," Comput. Human Behav., 2013.

[29] M. Salehan and A. Negahban, "Social networking on smartphones: When mobile phones become addictive," Comput. Human Behav., vol. 29, no. 6, pp. 2632–2639, 2013.

[30] R. Ahluwalia, "Two thirds of people find holiday selfies and 'hot dog legs on the beach' Instagram posts annoying." [Online]. Available: https://www.independent.co.uk/travel/news-and- advice/holiday-selfies-annoying-instagram-facebook-photos-hot-dog-legs-vacations-a7816386.html.

[31] I. Arpaci, "Understanding and predicting students' intention to use mobile cloud storageservices," Comput. Human Behav., 2016.

[32] A. T. Korucu and E. Usta, "The Analysis of New Generation Mobile Device Dependencies of Students in Faculty of

Education," Particip. Educ. Res., 2016.

[33] G. Başaran İnce, "Digital Culture, New Media and The Transformation of Collective Memory," Galatasaray Üniversitesi İleti-ş-im Derg., 2017.

[34] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," Behav. Brain Sci., 2017.

Privacy Concern, Data Quality and Trustworthiness of AI-Analytics

ARI ALAMÄKI
Haaga-Helia University of Applied Sciences, Ratapihantie 13, 00520 Helsinki, Finland, ari.alamaki@haaga-helia.fi

MARKO MÄKI
Haaga-Helia University of Applied Sciences, Helsinki, Finland

R. M. CHANDIMA RATNAYAKE
University of Stavanger, N-4036 Stavanger, Norway

The present study investigates the role of trustworthiness of data analytics from the data quality and privacy concern perspectives. In addition to the privacy concern of users, we investigated conceptually the requirements and impacts of data quality to the business processes. The goal of the conceptual analyze was to gain more knowledge about the factors affecting to the data quality, its accuracy and business impacts. The privacy concern is a part of data quality. The behavior of users is closely related to the data that they insert to the software systems. The research approach is the case study, that allowed to develop a new understanding of the relationship of privacy concern, data quality and trustworthiness of machine learning. The case study used the abductive qualitative research method, as the study aims to build a new conceptual understanding trustworthiness of AI-based data analytics. Using the iterative research process allowed for developing a deeper understanding while contributing to the conceptual models. The contribution of this paper is to show that data quality affects the trustworthiness of results. The privacy concern is a factor that influences indirectly to the trustworthiness. For the managerial implication, this paper suggests to put special emphasizes to the very first phases of data collection processes where human factors or sensor technological shortages might corrupt the data quality. To sum up, the present study underlines the importance of data quality, reliability and validity in different data categories. Data trustworthiness and data quality evaluation should be included to all marketing and business operations where data is utilized.

1. Introduction

The trustworthiness of data analytics, privacy concern and data quality are interrelated concepts. Companies need the consent from users to use their personal information. This enables large, more accurate and detailed databases about the users' online behaviour. Furthermore, if users do not have privacy concern and they trust to the digital service provider, they probable do not fake their personal information in registering, filling and using the data collection menus of digital services. Privacy concern refers in this study to users' emotional uncertainty to provide consent or correct information to the digital service provider in using her or his personal information. The prior research points [1] (Fletcher, 2003) it out that there is a growing concern among users about having to reveal personal information. In addition, many users are not satisfied with the way in which service providers collect and use information [1] (Fletcher 2003). Furthermore, this research shows that privacy concern is very important for the users of digital services.

The quality of data is also essential for the trustworthiness of analytics generated by AI. Data quality refers the features of data that affects its consistency, integrity, accuracy and completeness. Thus, in developing artificial intelligence based solutions, it is important to identify the quality of data that the AI-systems processes. If users provide fake data as they do not trust the service provider or they do not allow to use their real data in a legal way, the data analyses concerning users' online behaviour might become trustworthiness. Similarly, the conclusions are inaccurate or even misleading if the quality of original data is poor.

The present study investigates the role of trustworthiness of data analytics from the data quality and privacy concern perspectives. In addition to the privacy concern of users, we investigated conceptually the requirements and impacts of data quality to the business processes. The goal of the conceptual analyze was to gain more knowledge about the factors affecting to the data quality, its accuracy and business impacts. The privacy concern is a part of data quality. The behavior of users is closely related to the data that they insert to the software systems.

The research approach is the case study, [2] that allowed to develop a new understanding of the relationship of privacy concern, data quality and trustworthiness of machine learning. The case study used the abductive qualitative research method, [3] as the study aims to build a new conceptual understanding trustworthiness of AI-based data analytics. Using the iterative research process allowed for developing a deeper understanding while contributing to the conceptual models. The abductive research method enabled us to build explanations about the phenomena by combining empirical findings of the privacy concern survey to conceptual study and literature of data quality. In this research process, we simultaneously processed the prior literature and the analysis of the survey and conceptual study [2].

The aim of this paper is to focus on the relationship of privacy concern, data quality and trustworthiness of data analytics. There is little prior research on this relationship, and research on data privacy concern in connection with AI-based analytics is scant. The research questions of this study were as follows: To study empirically how users think about the privacy concern, to study conceptually data quality from the business perspective and how these findings contribute to the trustworthiness of machine learning capabilities.

2. Prior research on privacy concern

The privacy concern is the natural part of users' online behavior. Privacy concern is related to the perceived risk that triggers the feeling of uncertainty. Thus, perceived risk estimation is a significant determinant of privacy concern [4]. The users may feel uncertainty while sending personal information to the digital services as the Internet-based information systems have ability to monitor, track and save their online behavior. The brand or reputation of service provider affect to the privacy concern. The research shows that users felt less concerned about privacy issues if they interacted with the service providers that they were able to trust [5,6]. Similarly, if users felt that service provider's data collection processes is fair, the users allowed easier to use their personal information [7].

The previous research [5,6,8] show that users differ from each other in terms of their privacy concern. Their individual factors affect to the level of privacy concern, but also the service provider based factors trigger privacy concern and uncertainty. According to the prior research, females have higher privacy concern than males, and healthy adults have also higher privacy concern than the ailing elderly [8]. The opportunity to control personal information on the digital sites decreased privacy concern [9]. Sjöberg [10] has found that users evaluate negative risks, such as online shopping risks, differently. The usage of digital services may generate several risks that can cause financial, functional physical or social consequences [11].

Any information is not similar from the users' perspective and trustworthiness. Users differ from each other in their privacy concern and information sensitivity. In addition to individual factors of users, the reputation, brand and other features of service providers affect to the degree of privacy concern. The recognizing the factors that influence to the privacy concern assist companies to design digital services and their customer support functions to meet the expectation of users. This impact directly to the users' willingness to provide consent to utilize their personal information in data analytics. Additionally, trust to the service provider improve the reliability of data analytics as users can trust that the service provider use their personal information anonymously and legal ways. Privacy concern may even create as a major obstacle for service providers the growth and develop their business [12]. It is also a significant source for incorrect analyse as machine learning cannot recognize false information that users who do not trust the service providers insert to the websites and social media applications.

Users differ from each other in terms of information sensitivity. Information type affects to the privacy concern as users evaluate sensitivity of information [4,9,12]. Information sensitivity is closely related to the information type, such as gender, age, politics, religion, contact information, social networks, purchase behavior, attitude or socio-economic information. Users allow easier to collect information that is public and general information, such as age and gender [13]. Additionally, they allow easier to collect information that does not include personal identification information [9]. Similarly, the information that they provide to the service provider also affects their privacy concern. All information is not similar, and users evaluate the sensitivity of information that they allow to use. Our study is align with the prior research that reveals that users differ from each other concerning their privacy concern.

3. Users' privacy concern in using digital services

The companies collect user information from different digital touchpoints when they are using their digital services in searching, purchasing and using products and services [14]. Technologically digital service primarily use cookies in identifying the individual online behavior. There are also other means to identify individual users in the Internet, such as the IP-address and hardware MAC-address that help to identify the device. Additionally, users have logged in to many digital services that reveal their user profile and behavior from the personal data and users' own posts. For example, Google's and Facebook's services know more details about users than the most users can expect. Several mobile applications are constantly communicating to the external web-servers for writing various usage information to their databases, sometimes without formal permission [15]. The location features of mobile services send users' location information to the service providers. Despite to those invisible backend-roaming processes, users also share quite openly their personal information in the Internet. It is important to notice that although users have provided consent to use their information, but it does not guarantee that information that they insert is correct. Thus, it is important to research the users' privacy concern concerning the usage of digital services.

We investigated user’s privacy concern related to the digital services. From the viewpoint of trustworthiness of analytics, service providers should receive consent from users to use their authentic data in analyzing user behavior. The reliable analytic requires trust samples of user behavior. The information that companies collect through digital channels relates, for example, to demographics, personal characteristics, contact information, purchase history, financial transactions or emotional issues. The companies collect data from various digital touchpoints when users are searching, reading, communicating, purchasing and using services digitally and physically.

Data for understanding privacy concern of users was collected among university students (N=299) in Finland, representing potential users of artificial intelligence-powered e-commerce, social media and functional systems. The sample is female-dominant: 67 percent (n=201) of the respondents are female and 33 percent (n=98) are male. The questionnaire was sent to participants in the email that included the web-link. We measured privacy concern using four questionnaire items adopted from Martin et al. (2017). The respondents were asked to rate value using a five-point Likert scale ranging from totally agree (5) to totally disagree (1).

We found that 60 % of the respondents were worried about data privacy threats, and similarly they state that privacy is very important to them. Thus, the survey shows that significant part of users are concerned about their privacy in the Internet, whereas only 2% disagree. Additionally, 84 % of respondents perceived important that their privacy will remain untouched by on-line companies. Over 80 % of users perceived important that they know why the websites collect data from them. The findings indicate that privacy is very important to the respondents.

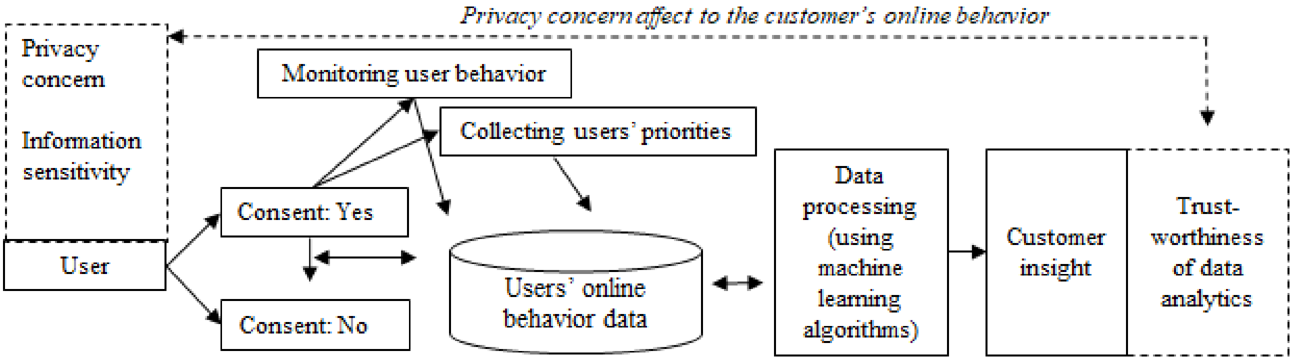

Privacy concern affect to the customer's online behavior
Privacy concern
Information sensitivity
User
Monitoring user behavior
Collecting users' priorities
Consent: Yes
Consent: No
Users' online behavior data
Data processing (using machine learning algorithms)
Customer insight
Trust-worthiness of data analytics

Fig. 1. The relationship of privacy concern and trustworthiness of AI-analytics.

The conceptual model (Fig. 1) illustrates how privacy concern of users and the sensitivity of information are related to the users’ online behavior. The user may provide consent to use her or his personal information but it does not have causal connection to the trustworthiness of data analytics. The digital services are able to collect data in two ways; monitoring user behaviour by using e.g. cookies and collecting and saving users’ direct comments, posts, writings or voice messages that they insert through the user interfaces of digital services.

4. Data quality and machine learning from business process perspectives

Onshore and offshore industrial asset integrity assessment and control with the support of data analytics [i.e. together with machine learning (ML)] has been a significant challenge due to the data management difficulties arise by ‘data quality’. For instance, Ratnayake and Kusumawardhani [16] have revealed that piping wall thickness reduction measurement data reliability is within the range of (82-90)% and (80-92)% for welds and bend respectively. Hence, operational/life cycle data requires thorough cleansing and preparation to be used as input to any analytics supported intelligent system. In this context, it is considered that data has a quality if they “fit for [its] intended uses in operations, decision making and planning and data is deemed of high quality if it correctly represents the real-world construct to which it refers” [17]. In an era of automated self-service analytics and intelligent systems, data quality has assumed even more significance as most of the users often have no prior knowledge or skills to differentiate between bad and good data. On the contrary, the piles of complex raw data are rapidly equipped with advanced analytics software tools supported by ML techniques (i.e. supervised or unsupervised) for extracting patterns to reflect competitive and actionable intelligence. However, modern IT systems are not yet fully capable of dealing with ‘data quality’, which directly has an impact on data extraction from multiple sources, data preparation, and data cleansing. This has been further exacerbated by heterogeneous data sources, high volumes of data, and a myriad of unstructured data types. The data quality has several dimensions: consistency, integrity, accuracy and completeness.

It is possible to assess data quality in relation to the level of compliance of a data set with a circumstantial normality in which the normality can be set by operational conditions' related and/or statistically (or empirically) derived rules. The data quality is contextual, in the sense that rules reflect the logic of particular industrial plant's design and fabrication resume, level of aging [i.e. "ageing is not about how old your equipment is; it is about its condition, and how that is changing over time" [18], geographical location (i.e. product and process conditions differs based on the production field [19], and regulatory concerns (i.e. Health, safety, environmental and societal conditions). For instance, a property (e.g. pumps, turbines, piping, structures, etc. in an offshore production and process facility) of the similar structural/ mechanical characteristics could have different validation rules depending on the operational environment or conditions [e.g. depending on the maturity of the production field [20] resulting different data quality requirements [i.e. reduced piping wall thickness as opposed to original design intent may not increase risk of a potential failure as the production well pressure goes down over the time; corrosion/erosion rates might be quite different at the end of the life, etc. [20] to make final assessments. Hence, the systems are in the need of exposing dirty, inaccurate or incomplete data when assessments, evaluations and recommendation about production components/clients have been made via data analytics.

In this context, an outlier is a critical operational discovery, or it can be an unknown/poorly-handled data. The worst case arises when the real-time decisions have been made by poor data with data analytics via ML. In this kind of situation, it may not be able to identify and handle poor data, which causes eventually, accidentally, or even intentionally to be fed them into the process. Hence, it is vital to integrate adaptive rule-based systems (i.e. to maintain circumstantial normality) to cater problematic situations resulting in poor-data quality entries in a way that the system will be able to recognize the level of quality and proactively notify the end users. This requires integrating risk-based assessments to evaluate the level of risk of serving data to the end users or to serve data whilst rising an alert/flagging about the level of risk of following the current recommendations. The aforementioned enables to mitigate data quality issues and improving the trust in data and data analytics, waste of resources and/or poor decisions. It is inherent fact that 'the things that do not measure, would not be able to manage' [21, 22], which does apply to the data quality. Hence, it is vital that the metadata underpinning an industrial data governance initiative to be assessed in relation to a set of metrics for data quality. Such assessment or measurement enables to benchmark current performance and to plan for future improvement. Figure 2 illustrates the metrics of data quality that has influence on assessment and control tasks, which deploy data analytics via ML.

The improved compliance is assessed in relation to the transparency of the risk potential fines, capital charges or reputational damage. The level of capability to satisfy regulatory requirements are assessed by knowledge of data sources, applicability and timeliness. The faster results are assessed in relation to the efficiency for accessing the data set to enable faster and better decision-making. Level of waste is assessed how the enhanced quality data can streamline operations across the overall target areas focusing on decreasing the risk of discrepancies and costly compromises, mitigating the occurrence of regulatory penalties, and minimizing the cost of unreliable data. It is possible to avoid using fake data having defined metrics and data quality assessment focus.

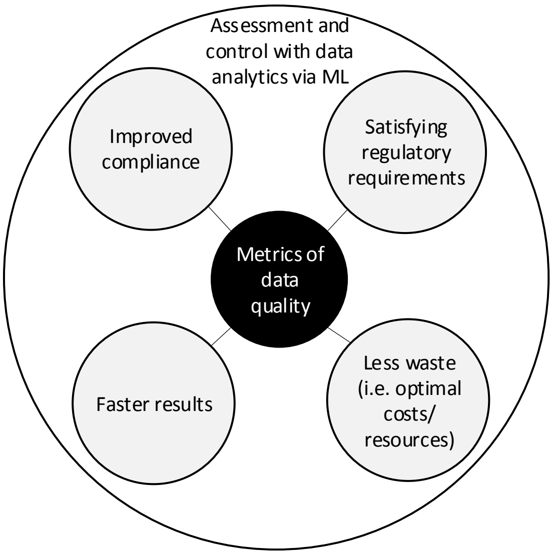

Assessment and control with data analytics via ML
Improved compliance
Satisfying regulatory requirements
Metrics of data quality
Faster results
Less waste (i.e. optimal costs/ resources)

Fig. 2. Metrics of data quality assessment: data analytics via ML.

.

5. Discussion

This paper shows that data quality has several dimensions and factors that influence its trustworthiness. The trustworthiness is related to the accuracy, validity and business value of data. The privacy concern affects to the trustworthiness of data as users can manipulate information that they provide. The challenge for the most today's artificial intelligence system is that its machine learning is not able to recognize biased or corrupted data from the high quality data if data fulfills other requirements. In other words, the machine learning processes data but it is not able to evaluate the process how data is created. Data is often created, shared and managed in the business networks [23,24]. The trustworthiness of data is not the same thing than complete, integral and consistent data. Thus, data can fulfill "technically" the requirements although its content is biased or corrupted. Additionally, knowledge workers are often incapable to differentiate bad data from the high quality data if they do not know how data has been collected and pre-processed.

We summarized the findings to the Table 1. It presents different data how they are related to the user and business perspectives, reliability, validity, accuracy and machine learning. The fake data is an example of situation where users have provided wrong information or sensors' calibration have been broken and they are sending too low value. The knowledge workers or machine learning cannot recognize fake data from the high quality data. For example, the one in three users of a new digital service have told that they are younger and more educated than they are. The internet of things application measures temperature falsely as the exhaust of engine is heating its sensor. The incomplete data is easier to recognize as it has "technical" shortages, such as some values are missing. The compensatory data, such as testing or simulation data looks like the original high quality data but it is artificial. The outdated data have been collected from the real-life situation but its business value is out of date. Sometimes, data might become out of date within seconds like in the IoT-systems that control real-time processes of devices. The findings of this study is align with the research of data governance [25].

The contribution of this paper is to show that data quality affects the trustworthiness of results. The privacy concern is a factor that influences indirectly to the trustworthiness. For the managerial implication, this paper suggests to put special emphasizes to the very first phases of data collection processes where human factors or sensor technological shortages might corrupt the data quality. These human and technological factors merit further research.

TABLE I
The relationship of privacy concern, data quality, trustworthiness of data and output of machine learning.

	High quality data	Fake data	Incomplete data	Compensatory data	Outdated data
Definition	Data represents sample, its sample size is sufficient for generalization and it does not have biases, etc.	Data looks like reliable and its sample size is sufficient but responders have given false information	Data represents the sample, but it has some faults or its sample size is too small for generalization	Data has not been collected from the real customers but it simulates the sample and it has been validated	Data represents sample and its sample size is sufficient but the actual situation that it measures is out of date.
User perspective	No privacy concern: users have trusted service provider or information is not sensitive	Privacy concern: users have not trusted service provider or information is sensitive	Privacy concern: users have not fully trusted service provider or information is sensitive	There is no users, information is highly sensitive or it cannot be used	No privacy concern: users have not updated information or they have rejected the service.
Business perspective	Data is consistent, integral and complete providing accurate results.	Data is consistent, integral and complete "technically" but its content is biased or corrupted.	Data is not consistent, integral or complete providing inaccurate results.	Data is consistent and integral providing accurate results with medium or high uncertainty	Data is consistent, integral and complete providing accurate results only from the history
Reliability	High	Low	Medium / Low	High / Medium	Low
Validity	Objective	Fake	Partly objective	Objective	Objective as history data
Accuracy	Reliable insight	False insight	Gives some hints or trends	Reliable insight	Trusted / Untrusted
Ability to train machine learning	ML is able to learn the patterns of real-life phenomena	ML is not able to learn the patterns of real-life phenomena, only fictional	ML has difficulties to create meaningful patterns and requires interaction of human experts	ML is able to learn the patterns of real phenomena, but results should be validated by the human expert	ML is partly able to learn the patterns of real-time phenomena if updated data is later available, and human expert validates results

References

[1] Fletcher, K. (2003). Consumer power and privacy: the changing nature of CRM. International Journal of Advertising, 22(2), 249-272.

[2] Eisenhardt, K.M., Graebner, M.: Theory building from cases: opportunities and challenges. Academy of Management Journal, 50(1), 25-32 (2007)

[3] Dubois, A., Gadde, L.E.: Systematic combining: An abductive approach to case research. Journal of Business Research, 55(7), 553-560 (2002)

[4] Dinev, T., Xu, H., Smith, J. H., & Hart, P. (2013). Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts. European Journal of Information Systems, 22(3), 295-316.

[5] Chellappa, R. K., & Sin, R. G. (2005). Personalization versus privacy: An empirical examination of the online consumer's dilemma. Information technology and management, 6(2-3), 181-202.

[6] Bleier, A., & Eisenbeiss, M. (2015). The importance of trust for personalized online advertising. Journal of Retailing, 91(3), 390-409.

[7] Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. Organization science, 10(1), 104-115.

[8] Wilkowska, W., & Ziefle, M. (2012). Privacy and data security in E-health: Requirements from the user's perspective. Health informatics journal, 18(3), 191-201.

[9] Phelps, J., Nowak, G., & Ferrell, E. (2000). Privacy concerns and consumer willingness to provide personal information. Journal of Public Policy & Marketing, 19(1), 27-41.

[10] Sjöberg, L. (2000). Factors in Risk Perception. Risk Analysis: An International Journal, 20(1), 1-12.

[11] Laroche, M., Bergeron, J. and Goutaland, C. (2003) "How intangibility affects perceived risk: the moderating role of knowledge and involvement", Journal of Services Marketing, Vol. 17 Iss: 2, pp.122 – 140

[12] Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. Information systems research, 15(4), 336-355.

[13] Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In International workshop on privacy enhancing technologies(pp. 36-58). Springer, Berlin, Heidelberg.

[14] Hallikainen, H., Alamäki, A., & Laukkanen, T. (2018). Individual preferences of digital touchpoints: A latent class analysis. Journal of Retailing and Consumer Services. Advanced online publication.

[15] Yle, (2019) "Nokia 7 Plus handsets sent data to Chinese servers, broadcaster reports" YLE News https://yle.fi/uutiset/osasto/news/nokia_7_plus_handsets_sent_data_to_chinese_servers_broadcaster_reports/10701132

[16] Ratnayake, R.M.C., and Kusumawardhani, M., (2013), "Reliability analysis of condition monitoring data on aging plants: A case study from topside static mechanical systems", Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), DOI: 10.1109/IEEM.2013.6962641

[17] Wikipedia (2019), "Data quality", https://en.wikipedia.org/wiki/Data_quality (accessed on 02.04.2019).

[18] Horrocks P., Mansfield D., Parker K., Thomson J., Atkinson T., and Worsley J., (2010) "Managing Ageing Plant: A Summary Guide" http://www.hse.gov.uk/research/rrpdf/rr823-summary-guide.pdf (accessed on 02.04.2019).

[19] Ratnayake, R.M.C. (2012), " Challenges in Inspection Planning for Maintenance of Static Mechanical Equipment on Ageing Oil and Gas Production Plants: The State of the Art", Proceedings of the ASME 31st International Conference on Ocean, Offshore and Arctic Engineering, Paper No. OMAE2012-83248, pp. 91-103; doi:10.1115/OMAE2012-83248

[20] Ratnayake, R.M.C., (2013), " Utilization of Piping Inspection Data for Continuous Improvement: A Methodology to Visualize Coverage and Finding Rates", ASME 32nd International Conference on Ocean, Offshore and Arctic Engineering (OMAE2013), Paper No. OMAE2013-10025, pp. V003T03A001; doi:10.1115/OMAE2013-10025

[21] Behn, B. (2005) 'On the philosophical and practical: resistance to measurement', Public Management Report, November, Vol. 3, No. 3, pp.1–2.

[22] Ratnayake, R.M.C., and Markeset, T. (2010), "Implementing company policies in plant level asset operations: measuring organisational alignment ", European Journal of Industrial Engineering , Vol. 4, No. 3, pp. 355-371.

[23] Alamäki, A, Rantala, T., Valkokari, K. and Palomäki, K. (2018). Business Roles in Creating Value from Data in Collaborative Networks". In Camarinha-Matos, L.M, Afsarmanesh, H. & Rezgui, Y. (Eds.) Collaborative networks of cognitive systems. The proceedings of the 19th IFIP/SOCOLNET Working Conference on Virtual Enterprises, Pro-Ve 2018, Cardiff, UK, September 17-19, 595-606.

[24] Valkokari, K., Rantala, T., Alamäki, A. and Palomäki, K. (2018) Business Impacts of Technology Disruption – A Design Science Approach to Cognitive Systems' Adoption within Collaborative Networks". In Camarinha-Matos, L.M, Afsarmanesh, H. & Rezgui, Y. (Eds.) Collaborative networks of cognitive systems. The proceedings of the 19th IFIP/SOCOLNET Working Conference on Virtual Enterprises, Pro-Ve 2018, Cardiff, UK, September 17-19, 325-336

[25] Aunimo, L., Alamäki, A. and Ketamo, H. (2019). Big Data Governance in Agile and Data-Driven Software Development: A Market Entry Case in the Educational Game Industry. In Strydom, S.K. & Strydom, M. (Eds.) Big Data Governance and Perspectives in Knowledge Management, 335 pages, IGI Global, pp. 179-199