International Scientific Conference
CONTEMPORARY ISSUES IN BUSINESS, MANAGEMENT AND ECONOMICS ENGINEERING'2019

9–10 May 2019, Vilnius, Lithuania
Vilnius Gediminas Technical University

eISSN 2538-8711
ISBN 978-609-476-161-4 / eISBN 978-609-476-162-1
Article ID: cibmee.2019.084

https://doi.org/10.3846/cibmee.2019.084

ASSESSING AND MANAGING BUSINESS RISKS FOR ARTIFICIAL INTELLIGENCE BASED BUSINESS PROCESS AUTOMATION

Gergő BARTA*, Gergely GÖRCSI*

Doctoral School of Management and Business Administration, Szent István University,
Páter Károly utca 1, 2100, Gödöllő, Hungary

**E-mails: Barta.Gergo@phd.uni-szie.hu; gorcsi.gergely@nisz.hu*

Abstract. *Purpose* – the number of projects and amount of investment into Artificial Intelligence (AI) based business process automation is increasing. To utilize the power of AI, business organizations shall achieve a certain level of digital maturity that enables of handling the risks arising from AI. AI brings new risk factors to their life that has to be reduced to an acceptable level. If risk mitigation procedures are not in place, then AI might cause a greater headache than a market advantage resulting in expensive implementation with no business benefit.

Research methodology – the objective is to analyze what risk factors can AI bring with itself to the life of corporations by analyzing general IT risk assessment processes and the stages of AI development.

*Findings – We o*bserved that current IT risk assessment methodologies don't detail possible risk scenarios regarding intelligent applications and don't extend their threat catalogs to help organizations consider threats related to AI.

Research limitations – the research work details possible risks for general AI development that might differ across industries, business cases, specific algorithms etc.

Practical implications – the research contributes to organizations to assess possible risks arising from the use of AI.

Originality/Value – since AI based automation is the result of recent research work, analyzing risk management aspects of its use can be considered as a new field for further research.

Keywords: Artificial Intelligence, Machine Learning, IT risk assessment, Risk Management, Business Automation.

JEL Classification: O32

Conference topic: Digitalization of Business Processes: Trends, Challenges, Solutions.

Introduction

AI-based solutions gained extreme popularity in recent years that is due to several advancements in its supporting technology (such as Big Data and hardware capabilities). The exact wording of AI brings many challenges as there are different definitions in an academy and in industry, therefore expressing what AI exactly covers is a hard task and appears to be subjective in many cases. That is supported by a survey performed by MMC venture capital firm, that 40% of startups that are claimed to be AI companies, do not actually use AI (MMC Ventures, 2019). According to a classic definition of Russel and Norvig (2005), AI is "the designing and building of intelligent agents that receive percepts from the environment and take actions that affect that environment." However, this definition today more resembles the definition of "reinforcement learning" which is a subcategory of "Machine Learning" A more general definition was given by Borgulya (1998) that highlights that AI is thinking and acting as a human by "modeling human problem solving". The concept of AI itself exists for a long time and first, it was expressed by John McCarthy that has organized in the summer of 1956, a two-month work program in Dartmouth for the researchers of computer intelligence (McCarthy, Minsky, Rochester, & Shannon, 1955). These early years, AI meant a computer system that could produce human intelligence, but AI was more like a set of conditions that have led to a conclusion by evaluating logical expressions that lie on the principle of traditional/classical computer programming. That was the case still in the 80's and 90's when researchers focused on building expert systems that were to solve domain specific problems. In the late 90's Machine Learning (ML) has got more and more popularity that is the science of building intelligent systems that are capable of making decisions without explicit programming (Dua & Du, 2011). This means that no explicitly created conditions are needed so as to deduce a conclusion from an available dataset, i.e. algorithms are able to learn interrelations among data and produce the so-called intelligent decision (generate the rules automatically). Figure 1 represents the difference graphically between the classical programming and the machine learning paradigm.

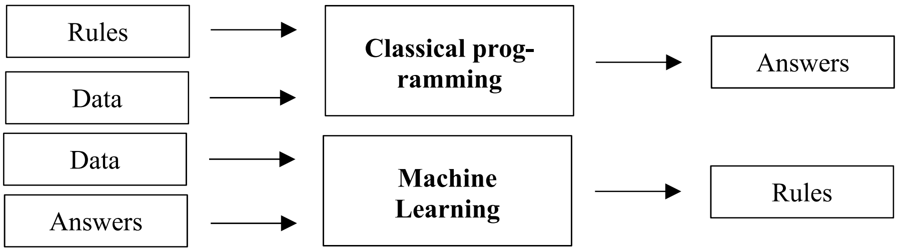

Figure 1. Difference between the classical programming and ML paradigm (source: Chollet, 2018)

ML is, therefore a subcategory of AI that is called AI today by industry, thus we can observe that the academic definition differs. ML is generally divided into three subcategories. First is the supervised learning that aims to predict a target variable based on the available data. Supervised learning is used in many fields such as image recognition, speech recognition, weather prediction, fraud detection etc. In the case of autonomous vehicles, an image recognition system is in use to recognize pedestrians, other vehicles, traffic signs etc. therefore the target variable is the object to be recognized and the available data (the conclusion is deduced from) is the camera pictures i.e. the problem is quasi a prediction problem that a self-driving car shall perfectly determine with no exception. The second category of ML is the unsupervised learning that has an objective of finding interrelations within the data, however with no target variable determined. Generally, classical statistical methods can be an example of it, such as clustering or principal component analysis. Unsupervised learning can be useful to find patterns in the data and then grouping them. Business use can be to determine similarities within customers that can be handled as one group and can be targeted with a similar marketing campaign. The third category is the reinforcement learning that aims to optimize a function, such as minimize a cost function or maximize a profit function, and an agent is collaborating with the environment in order to achieve this goal. A typical example can be a chess program or Google's Alphago that has recently defeated the Go game champion (Russel, 2017).

In this article, the authors understand the supervised learning under intelligent decision making in the first place and analyzed its development processes in order to examine it in terms of risk assessment. The reason is that supervised learning is the subcategory of AI and ML that is widely used in many new applications (virtual assistants, self-driving cars etc.) and introduces a new paradigm as referenced above. Deep Learning (DL), one of the state-of-the-art technologies is used in all three categories, however, it is generally deemed to be a supervised method for classification tasks. Deep Learning lies on the principle of how the human brain works and made it possible to build systems that have a high accuracy of solving classification tasks. Deep Learning is operating in a high dimensional space that makes it impossible to understand how a certain decision was derived, therefore that means a high risk for a business operation to understand how it is exactly working and how specific rules are created to solve a targeted problem.

1. Related work

To the best of the knowledge of the authors, there are not many published research work and related articles in this field as experiences show that research papers either more focus on model development or the concept of intelligent applications in business domains. This can lead to the conclusion that academical research is more excited about how AI can make our life better than addressing business risk from it. Companies are befriending with the idea of finding business cases for AI implementation and working on the proof of concepts, as it was revealed by Andrews et al. (2017) in their research work performed at the end of 2017, then conducting risk assessments to mitigate any arising risk associated with their businesses. However, there was some research conducted that draws into attention that AI is not only the new business oil, but it has disadvantages too, that shall be assessed and handled. Andrew Clark analyzed AI processes in terms of audit considerations and published several articles how internal or external auditors shall adjust their audit plans when it comes to intelligent decision making (Clark, 2016, 2017, 2018). Auditing is an important part of risk assessment, as the primary function of the auditors is to gain reasonable assurance whether internal controls are operating effectively which is an option to reduce business risk to an acceptable level (Barta, 2018b). Thus, developing audit procedures to obtain assurance over controls is an indispensable part of the organizational risk management framework and so as to appropriately assess risk, auditing must be in place and performed continuously. The authors also analyzed the development processes of intelligent systems in respect of audibility and suggested several questions that shall be addressed by auditors when testing AI applications (Barta & Görcsi, 2018). Such questions detailed considerations regarding the source of data to be used, model development and evaluation criteria. Yampolskiy and Spellchecker (2016) performed research work regarding AI failures (that are to be addressed by a risk program) and concluded that "human values are inconsistent and dynamic and so cannot be understood and subsequently programmed into a machine. Suggestions for overcoming this obstacle require changing humanity into something it is not". This implies that AI, most probably, will never achieve its goals and eventually every AI system will fail, therefore the implementation of such systems is questionable. In terms of this article, this is a relevant consideration, as one of the risk-reducing methods is to avoid risk by terminating the business process meaning that not implementing AI will not give the opportunity to possess any risk arising from it. That means excessive risk treatment and in case of such a decision, companies have to give up any benefits AI can bring.

2. Intelligent business automation

The use of intelligent systems has a wide range of business domains to cover that is summarized in Table 1 based on the work of Krishna, Albinson, and Chu (2017).

Table 1. Business domains for AI (source: Krishna et al., 2017)

Business domain	Use case
Business Operation	Automate production and other operational processes Predict quality issues and failures Monitor flow across the supply chain Enable predictive asset maintenance
Finance	Advise on investment decisions Execute automated trades and deals Develop, analyze, and execute contracts Generate automated reports
Sales and Marketing	Develop targeted marketing campaigns Measure the effectiveness of marketing campaigns Monitor social media for consumer insights Calculate discounts based on customer data
Risk Management	Identify, prioritize and monitor risks Spot fraud and conduct investigations Analyze business ecosystems Enforce regulatory compliance
Human Resources	Support workforce planning Source, recruit and hire talent Manage the performance of employees Increase employee engagement and retention
Information Technology	Automate testing of systems Monitor cyber threats Automate system maintenance Support cyber incident response

As the table implies, AI can be implemented to enhance processes through every business function that has a very promising side, namely, employees can focus on tasks that can create high business value and other manual processes can be automated reducing the risk arising from human error. Some may argue that with the increased use of AI, humans will be left with no occupation, however, the authors firmly believe that it will create more opportunities and workplaces, that being a similar case when organizations started using computer automation. In addition, the use of intelligent systems appears to be inevitable in the future when observing relevant trends. In Figures 2 and 3, two statistics can be inspected. Figure 2 shows the statistic that Lee (2017) has published regarding the number of mentions on social media for AI in thousands in each quarter starting from 2016 by different industries. Figure 2 shows the revenues from the AI market worldwide from 2016 to 2025 in million U.S. dollars published by Statista (2019). The values from 2017 are estimated values and shall be handled as such.

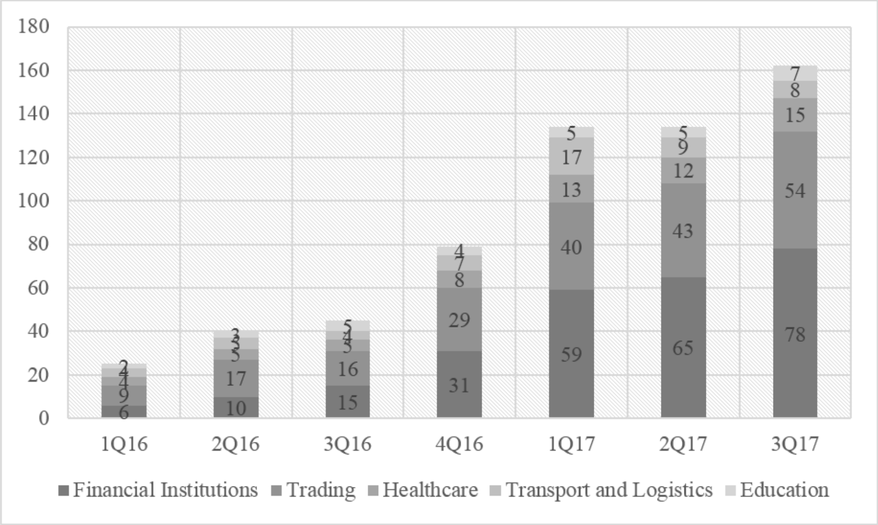

Figure 2. Number of social media mentions in thousands (source: Lee, 2017)

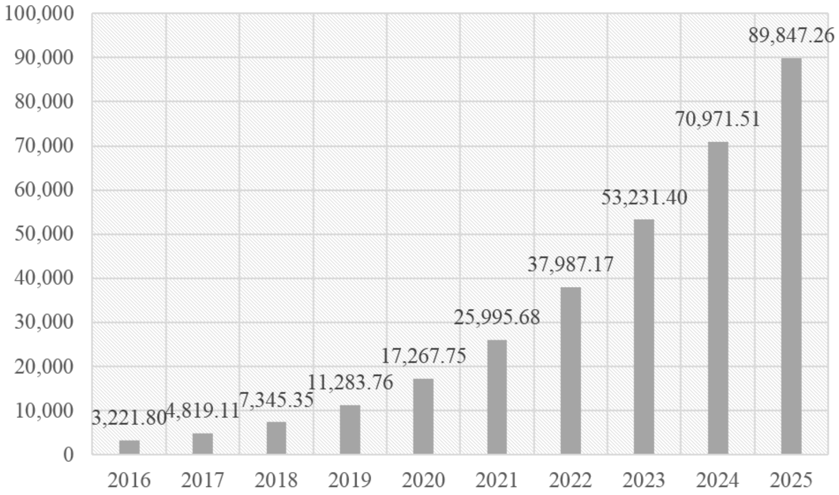

Figure 3. AI market worldwide from 2016 to 2025 in million U.S. dollars (source: Statista, 2019)

3. Information Technology risk assessment

No business can be imagined without any IT systems (applications, databases, operating systems etc.) as of today, thus IT operation is an essential business process for each and every company that has brought several vulnerabilities into the life of organizations, besides its benefits. IT systems are exposed to intentional and unintentional threats and vulnerable against malicious intent if not operated in controlled environments. In order to discover the nature and extent of adequate protection, a risk assessment shall be performed that aims to identify organizational IT assets, (that is the jargon used in the ISO27005 that includes software, hardware, network, organization, site etc., so every asset that can contain, transfer, create or delete information relevant to business operation.) their vulnerabilities and corrective actions that must be taken so as to maintain IT assets confidentiality, integrity, and availability. Therefore, a risk assessment is the process of uncovering existing risks and find a solution on how to mitigate them. IT risk is one of the components of any organizations' overall risk universe (Isaca, 2009) that also contains operational risks, market risks, environmental risks, credit risks etc. There are several standards that describe how to perform a risk assessment such as the ISO/IEC 27005 ("International Organization for Standardization" and "International Electrotechnical Commission") (2011), PCI DSS (Payment Card Industry Security Standards Council) Risk Assessment Guidelines (PCI Security Standards Council, 2012), Guide for Conducting Risk Assessment (NIST, 2012), IRAM2 (Information Risk Assessment Methodology 2) (Information Security Forum, 2014) etc., but the basic approach on how to run a risk program appears to be common in all:

1. Identify business processes – IT risk assessment shall be closely connected to business objectives, thus identifying the relevant business processes is the first step. Business processes must be prioritized based on their criticality to adequately adjust priorities to business areas that must be taking precedence. Business processes shall be detailed in order that each supporting IT system can be covered in the risk assessment activity, such as the procedure of credit scoring or procurement.
2. Identify supporting IT assets – IT assets shall be aligned with business procedures that means the applications, underlying databases, and operating systems, network, hardware, people, locations etc.
3. Assess the business impact of IT assets – Business management shall determine the appropriate business impact of IT assets such as specifying the financial, operational, human etc. loss in case an IT asset is compromised. That often means that IT assets shall be categorized on a scale regarding data confidentiality, integrity, and availability.
4. Maintain a threat catalog – IT assets are exposed to threats. That is the reason why a complete and accurate list shall be maintained that contains relevant threats to the organization and also they shall be aligned to IT assets to assess the impacts if an IT asset is vulnerable against a specific threat. E.g. Data centers might be exposed to earthquakes in several geographic locations, thus "earthquake" as a threat must be included in a threat catalog. In addition, a likelihood shall be aligned to each threat that shows the possible occurrence of the threat.
5. Identify IT assets' vulnerabilities that could be exploited by relevant threats – IT assets are vulnerable, therefore, their vulnerabilities shall be assessed.
6. Calculate inherent risk – the inherent risk is the risk that is calculated from the business criticality of an IT asset, the likelihood of relevant threat occurrence, and the vulnerability of the IT asset.
7. Identify controls that mitigate inherent risk – controls are processes, rules, countermeasures etc. that have the main goal to ensure that business is operating in alignment with management intention without fraud. Controls can be preventative (prevent a negative event to happen e.g. password enforcement to a system to prevent unauthorized access), can be detective (to detect if a negative event has happened e.g. network monitoring) and can be corrective (to correct a process if a negative event has happened e.g. restoring data after a server failure). Controls are risk mitigating procedures and shall be identified in order to obtain a clear picture of actual risk levels.
8. Calculate residual risk – the residual risk is calculated from the inherent risk and relevant controls.
9. Perform corrective action if the risk is higher than the organization's risk appetite – in case there are no controls in place or controls are not adequate or not sufficient, then corrective action might be needed, thus new controls shall be implemented to reduce risk. There are basically four different type of risk treatments such as reducing the risk by implementing controls, transferring a risk to a third party (e.g. buying insurance for critical servers), accepting risk and undertaking the negative impacts if a threat exploits an IT asset vulnerability, or avoiding the risk by totally eliminating the underlying business process (no risky process, no risk) (Isaca, 2009).

Risk assessments can be qualitative or quantitative, or hybrid containing characteristics from both methods. In the case of qualitative risk assessments, the business impacts and corresponding risk indicators are expressed in categorical values such as "high", "medium" or "low" or even more categories can be created depending on the utilized methodology. In quantitative risk assessments, the business impacts and the possible loss in case of negative events are expressed in financial amounts that might be a hard task to estimate (e.g. problem to estimate a reputational loss for a company in case of a server failure).

If we deem the aforementioned general approach as a baseline to assess risk for every IT asset, then we can conclude that intelligent IT systems shall be assessed similarly as traditional IT assets. That's logical. The difference in the details are the threats, vulnerabilities, and controls that must be extended in order to perform a risk assessment on intelligent systems. New methods emerge in IT capabilities, therefore different threats and vulnerabilities appear, new threats and vulnerabilities can be mitigated by new controls. The business objectives, business impacts and calculating the risk ratings remain the same, the change only occurred in the development and used methods in the IT assets. With that in mind, assessing the risk for intelligent systems, firstly the threats and vulnerabilities must be determined, and then the mitigating controls that reduce the risks.

4. Machine Learning development process

In order to reveal relevant threats for intelligent systems, general system development and operational process shall be understood and analyzed where the process differs from traditional software engineering. Based on the work of Raschka (2015), the development of intelligent systems is represented in Figure 4.

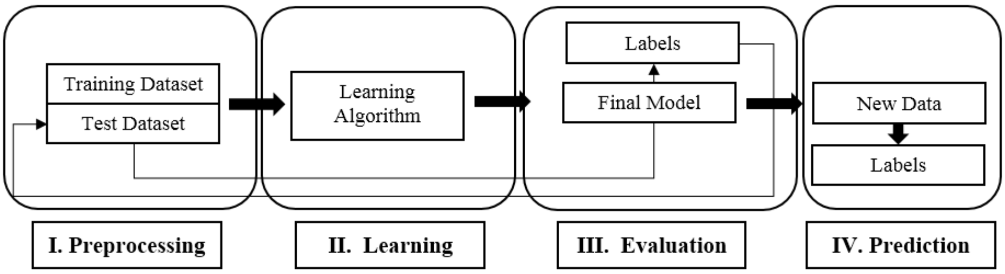

Figure 4. ML development phases (Raschka, 2015)

Raschka (2015) is dividing the process of intelligent system development into 4 phases:

1. The first is the preprocessing when the data manipulation is taking place including the data transformation (feature extraction and scaling) that can differ in case of the selected algorithms (e.g. neural networks need the data to be taken to the same scale, while the decision tree does not), feature selection meaning the process of selecting features that are contributing the most to the final decision or prediction, dimensionality reduction that is a technique to reduce the number of features by compressing them into fewer variable (e.g. principal component analysis), and the sampling that includes the separation of the dataset into training and test set.
2. The second phase is the learning phase when the chosen algorithm is processing the training set and learns the interrelation among the data.
3. The third is the evaluation when the model is tested and analyzed whether it performs according to the expectations e.g. capable of predicting the target. There is a feedback loop in this phase, i.e. if the system does not perform as intended (the accuracy is not in alignment with a predefined value), then the whole process shall be repeated that might result in new data gathering, new data transformation, new algorithm selection etc.
4. The last phase is the prediction when the system is implemented into production and used for its intelligent decision-making purposes.

Understanding the general development process of intelligent system development is the second stage (shall be handled as a business process) and then to appropriately identify relevant threats and vulnerabilities, the procedures must be analyzed.

5. Preprocessing

In the preprocessing phase, the necessary data is selected and transformed. Generally, this phase contains the following items based on Raschka (2015):

1. Feature extraction and scaling
2. Feature selection
3. Dimensionality reduction
4. Sampling
5. Division of the data to the training set and test dataset

Definitely the quality of the data determines the prediction power of the system. If we think that the organization wants to predict whether a customer is going to pay back its loan, then having a dataset about the company's sales will not meet the expectations. Therefore, the first threat that shall be addressed whether the data is relevant for the intended purposes, and if not, operational risk is in place. Let's suppose there is an interrelation between the data and the target. Then what can go wrong? Ribeiro, Singh, and Guestrin (2016) performed research work regarding a binary classifier that could distinguish huskies and wolves on a picture. The algorithm performed relatively well, however, after deeper analysis, it has been revealed that each of the pictures that showed a wolf, there was snow at the back. If a husky was seen on the picture with a snowy background, then the algorithm automatically classified it as a wolf. That leads to two conclusions. First, the system has to generalize well on unseen data in order that it could be implemented into production environment meaning that the input dataset must be complete and accurate that might be the hardest task to achieve as for many tasks, especially for complex business problems, the universe of possible inputs determining the outcome is nearly infinite. Second of all, as Andrew Ng (2018) highlights, the input data used in production shall be similar to the training data on what the algorithm learned the interrelations. Ng (2018) demonstrates the problem with a cat recognition application. If there is an algorithm that learns from data that were obtained through the internet but performs prediction on data that was uploaded by users recorded by their mobile phone, the algorithm is not going to have a high accuracy as the training data quality differs from production data. The same problem occurs if the data is outdated: e.g., collected from the past, but predicting for the future with changes business conditions, or data has missing values that have to be corrected or eliminated (Hastie, Tibshirani, & Friedman, 2009). Lastly, semantical problems can also occur that is due to inappropriate human interpretation that might be one of the hardest issues that can emerge, as human error always counts to be critical in any business field.

Several algorithms are performing well if the input data are transformed. Deriving the lessons learned from the publication of Rashid (2016), neural networks are performing better if the data are standardized or normalized. That is a similar case to clustering algorithms, the results are better if the data are on the same scale as variables with higher values can dominate the calculation, thus distorts the outcomes (e.g. if in a clustering problem the age variable and the wage variable are compared, the wage variable will dominate, as it usually represents a higher scale than age). This distortion can also be observed when there are highly correlated variables among the data meaning if one data point changes, then other changes, respectively, therefore it makes it harder to discover the explaining variables. However, algorithms, such as neural networks, can neglect this problem as it automatically reduces corresponding weights (gives less importance) for correlated variables. The issue can be solved by dimensionality reduction techniques such as principal component analysis (Sajtos & Mitev, 2007). In addition, ever since the General Data Protection Regulation came to force, special attention is needed to appropriately transfer data to make them not contain any personal information, one of the techniques can be anonymization to solve this issue (Barta, 2018a).

The last item on the list in the preprocessing phase is dividing the data into training and testing datasets. It can be read in several research work e.g. in the work of Kása (2011) that even the whole dataset is divided into three categories, the last one is the validation set. The training set is the data on what the training is performed, the validation set is the data on what the hyperparameter tuning is taking place i.e. configuring the algorithm to reach the desired accuracy, and the testing set is the dataset on what the system is evaluated. There should be an appropriate balance to be found. If the dataset is not divided and learning is performed on the whole dataset, filtering out the noise is impossible, and the system may learn interrelations that only occurred in specific cases. At this very moment, there is no optimal parameter determined regarding this issue, there are only recommendations such as Ng (2018), generally, the following distribution can be observed for the training set, validation set, and testing set: 70%, 10%, 20%.

After the examination of the aforementioned cases, it can be concluded that for the preprocessing phase the intelligent system is vulnerable against misinterpreted and erroneous data and data shall be cleaned and quality must be provided, before any programming activity is actually taking place, since it has an impact of the whole prediction power of the system. IT risk assessment shall be carefully planned to address this risk as consequently whichever intelligent algorithm is going to be further used, none of them will tolerate incorrect data, and wrong conclusions might be derived if used.

6. Learning

In the learning phase, the training data is loaded to the algorithm and the algorithm is ready to process it. Based on the interrelation found in the processed data the system is deducing a conclusion. The conclusion is determined by the internal rules that the system has declared based on the training data, but these rules not always to be interpreted by humans. As mentioned earlier e.g. neural network is a type of algorithm that is operating in a high dimensional space that basically cannot be understood why the conclusion is what it is. That phenomenon is called the black-box effect. The user can be ascertained that the algorithm works e.g. from a picture, it is able to identify a specific object, but the how is questionable. That can lead to several problems. In the research work of Wang and Kosinski (2018) a classification task was to determine if a face of a human being of their sexual orientation. The algorithm has achieved 91% accuracy in the case of men, and 83% in the case of women, while repeating the same task with humans, the results were 61% and 54%, respectively. This means that the algorithm has found interrelations that even a human cannot inspect, even the human results were only close to a guess like when flipping a coin and then you guess it would be heads or tails. This means that in case of the use of black-box algorithms, one can never be sure the nature and extent of interrelations what the data contains. The algorithm can discriminate people, violate regulations in order to achieve optimized performance or result in other unethical decisions. A similar case can be studied in the research work of Wilson, Hoffman, and Morgenstern (2019) where the authors noted that the skin color can have an impact on whether a self-driving car will hit the pedestrian or not.

Therefore, when performing a risk assessment on systems utilizing black-box algorithm a threat shall be addressed whether the algorithm is not causing reputational damage to the company by making unethical, unlawful and unacceptable decisions that can only be assured if the system has gone through a comprehensive testing analyzing each and every considerable test scenario and the whys are inspected. Typical black-box algorithms are neural networks (deep learning), support vector machine and random forests (Chen, Hsu, & Shen, 2005; Zhang & Zulkernie, 2016).

7. Evaluation

In the evaluation phase, the prediction power of the systems is assessed among several predefined KPIs that often contain the accuracy, cross-validation results and the analysis of confusion matrices. It is observed whether the algorithm was able to learn from the data or otherwise making random decisions. In addition, the data can be full of random noise meaning that the algorithm might learn this noise and therefore it will not generalize well, i.e. memorized the training set, but underperforms on the test set or in the worst case, in production. The first event is called underfitting, the latter one is overfitting. Underfitting means that the algorithm has a low accuracy on the training set (high bias) and test set, respectively, and based on the work of Ng (2018) either more data shall be collected, or a more complex algorithm shall be used. Overfitting indicates that a high accuracy was achieved on the training set, but low accuracy on the test set (high variance) that can be corrected by either regularizing the data (assigning penalty values to high weights) or using less complex algorithms, such as applying the rules from Occam's razor (Pitlik, 2014). Experiences show that usually having a low bias and low variance at the same is quite hard when the bias is decreasing than variance is increasing and vice versa (Ng, 2018).

8. Prediction

The prediction phase means that the system is ready to be implemented into the production environment. Developers and business users also tested the application, and management is satisfied with the results and produced accuracy. There are general IT risks and specific risks too, to be addressed. Appropriate user access management is considered important at this phase, as developers cannot perform the implementation. The reason is that once business units are done with the testing and the system is approved, no opportunities shall be provided to modify system configuration, data, program code etc. as it might negatively affect the application and its prediction power. This prevents unauthorized access. However, further development might be needed, but then the whole testing process must be repeated from the very beginning. Clark (2018) highlights that the system may communicate with other applications, thus interfaces and the whole IT environment must be tested. Another thing to consider is that system development did most probably not stop at this point, as the data are collected and fed into the algorithm to further improve its accuracy in time, meaning that continuous monitoring and testing is inevitable, thus the risk assessment shall also address risk such as having and educating appropriate staff that are dedicated to performing this job on a daily basis, and as mentioned before, the quality of data, evaluation of the system etc. In addition, there is another phase in traditional system development that is the post-implementation review that involves business analysis that must be performed to verify that the investment was worth, i.e. there is more benefit and profit in the long run than cost, established KPIs are met, and that the system is operating as management intended.

9. Mitigating controls

The result of a comprehensive IT risk assessment is the Risk Treatment Plan. This document contains the affected IT assets and risks that might be treated by a management approved strategy. If the organization does not accept, avoid or transfer the risk, then it shall mitigate by implementing controls over the development processes. Risk shall have a risk owner that is in charge of keeping track of the risk monitoring process and has the authority to implement risk-reducing countermeasures. Without an owner, the risk will not be reduced as there is no responsible personnel appointed. In addition, the severity of each risk shall be assessed, which is usually the risk rating of the residual risk, in order to prioritize tasks and risks to be handled and so then projects can be built to mitigate them. Definitely, if every countermeasure is implemented to eliminate risk, the IT risk assessment still should be continuous as new threats and vulnerabilities arise, and therefore they need to be assessed and managed.

Conclusions

The main objective of the article was to understand the general main risks that shall be considered when performing IT Risk assessments as an internal business process for business organizations. The authors performed literature reviews to obtain an overview of general IT risk assessments and the development phases of intelligent system applications that have led to the conclusion that the general risk assessment processes can be applied, however, the threats, vulnerabilities and mitigating procedures shall be appropriately adjusted as described above. This means at first place that data quality controls must be in place in order to ensure that intelligent algorithms are fed with adequately cleaned data that can contribute to the prediction power of the system including eliminating any semantical pitfall. Care should be taken when using black-box algorithms since the defined rules are not to be interpreted, thus in contrast to traditional statistical methods, the nature and extent of interrelations and dependencies cannot be discovered the meaning that the concluding logic is more like to be believed, than trusted. It might cause the consequence that in several application black-box algorithms must be excluded: e.g., in case of customer communication, as the user of the system cannot explain to business partners why it decided the way it did. Furthermore, intelligent algorithms shall be continuously maintained as the operating environment might change therefore newly obtained data have to be processed and the system needs continuous testing to ensure integrity. The authors recommend current risk assessment methodologies and threat catalogs to include intelligent system characteristics in order to hat companies can receive a complete list of risks to be handled and prepared for the changes that the intelligent system operation and development era bring. Further research is recommended to be conducted to understand and interpret black-box algorithms as in the case of big data, these methods appear to result in high accuracy. It is also recommended to perform research work on whether current risk assessment methods are truly appropriate to identify and analyze possible risks what intelligent systems may bring, or they might be too tailored for traditional IT systems meaning that new risk assessment methods shall be developed.

Disclosure statement

We do not have any competing financial, professional, or personal interests from other parties.

References

Andrews, W., Sau, M., Dekate, C., Mullen, A., Brant, K., Revang, M., & Plummer, D. (2017). *Predicts 2018: Artificial Intelligence*. Retrieved from https://www.gartner.com/document/3827163?ref=solrAll&amp;refval=193910164&amp;qid=780b332f7d9a fba6f17865ea8b939339

Barta, G. (2018a). Challenges in the compliance with the General Data Protection Regulation: Anonymization of personal information and related information security concerns. In: P. Ulman & P. Wołoszyn. *Knowledge – Economy – Society. Business, Finance and Technology as Protection and Support for Society* (pp. 115-121.). Poland: Foundation of the Cracow University of Economics.

Barta, G. (2018b). The Increasing Role of IT Auditors in Financial Audit: Risks and Intelligent Answers. *Business Management and Education, 16*(1), 81-93. https://doi.org/10.3846/bme.2018.2142

Barta, G., & Görcsi, G. (2018). Artificial Intelligence and Audit: Why is it necessary to audit the intelligent decision support? In: P. Földi; A. Borbély; Z. Kápolnai; M. B. Zsarnóczky; C. Bálint; E. Fodor-Borsos; I. Gerencsér; A. K. Gódor; F. Gubacsi; A. Nyírő; A. Szeberényi (Eds.) *Közgazdász Doktoranduszok és Kutatók IV. Téli Konferenciája*. (pp. 225-234.). Hungary: Doktoranduszok Országos Szövetsége.

Borgulya I. (1998). *Neurális hálók és fuzzy-rendszerek*. Budapest – Pécs: Dialóg Campus Szakkönyvek.

Chen, W. H., Hsu, S. H., & Shen, H. P. (2005). Application of SCM and ANN for intrusion detection. *Computers and Operations Research, 32*(1), 2617-2634. https://doi.org/10.1016/j.cor.2004.03.019

Chollet, F. (2018). *Deep Learning with Python.* New York: Manning Publications Co.

Clark, A. (2016). *Focusing IT Audit on Machine Learning Algorithms*. Retrieved from https://misti.com/internal-audit-insights/focusing-it-audit-on-machine-learning-algorithms

Clark, A. (2017). *Machine Learning Audit in the 'Big Data Age'*. Retrieved from https://www.cioinsight.com/it-management/innovation/machine-learning-audits-in-the-big-data-age.html?lipi=urn%3Ali%3Apage%3Ad_flagship3_profile _view_base%3BZBLUSmhrSLqwbpJa%2F%2BH7Wg%3D%3D

Clark, A. (2018). *The Machine Learning Audit—CRISP-DM Framework.* Retrieved from https://www.isaca.org/Journal/archives/2 018/Volume-1/Pages/the-machine-learning-audit-crisp-dm-framework.aspx?lipi=urn

Dua S., & Du X. (2011). *Data Mining and Machine Learning in Cybersecurity*. USA: Taylor and Francis Group. https://doi.org/10.1201/b10867

Hastie, T., Tibshirani, R. , & Friedman, J. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Second Edition*. New York: Springer Science. https://doi.org/10.1007/978-0-387-84858-7

Isaca (2009). *The Risk IT Framework*. USA: Isaca.

ISO/IEC. (2011). *27005 Information technology – Security techniques – Information security risk management*. Switzerland: International standard.

Information Security Forum (2014): *IRAM2. The next generation of assessing information risk.* Information Security Forum Limited.

Kása, R. (2011). *Neurális Fuzzy rendszerek alkalmazása a társadalomtudományi kutatásban az innovációs potenciál mérésére.* Doktori disszertáció. Retrieved from http://193.6.1.94:9080/JaDoX_Portlets/documents/document_6323_section_1701.pdf

Krishna, D., Albinson, N., & Chu, Y. (2017). *Managing algorithmic risks. Safeguarding the use of complex algorithms and machine learning.* Retrieved from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-risk-algorithmic-machine-learning-risk-management.pdf

Lee, A. (2017). *Market Trends: Three Conversational AI Trends That Will Distinguish Next-Generation Digital Commerce.* https://www.gartner.com/document/3833373?ref=solrAll&refval=194792633&qid=9af84aa06551e5b53fc096c4903b34a5

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. E. (1955*). Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Dartmouth: Tech. Rep., 3p.

MMC Ventures (2019). *The state of AI 2019: Divergence*. Retrieved from https://www.mmcventures.com/wp-content/uploads/2019/02/The-State-of-AI-2019-Divergence.pdf

National Institute of Standards and Technology. (2012). *Guide for Conducting Risk Assessments*. USA: NIST Special Publication 800-30.

Ng, A. (2018). *Machine Learning Yearning. Technical Strategy for AI Engineers, In the Era of Deep Learning (Draft ver.).* USA: deeplearning.ai.

PCI Security Standards Council. (2012). *Information Supplement: PCI DSS Risk Assessment Gudielines*. PCI Data Security Standard (PCI DSS).

Pitlik, L. (2014). *Occam hermeneutikája*. Magyar Internetes Agrárinformatikai Újság. Retrieved from http://miau.gau.hu/miau2009/index.php3?x=e0&string=occam

Raschka S. (2015). *Python Machine Learning*. Birmingham: Packt Publishing.

Rashid, T. (2016). *Make Your Own Neural Network* (1st ed.). CreateSpace Independent Publishing Platform.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *Why Should I Trust You?* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16. https://doi.org/10.1145/2939672.2939778

Russel, J. (2017). *Google's AlphaGo AI wins three-match series against the world's best Go player.* Retrieved from https://techcrunch.com/2017/05/24/alphago-beats-planets-best-human-go-player-ke-jie/

Russel, S., & Norvig, P. (2005). *Artificial Intelligence: A Modern Approach* (3rd ed.). India: Pearson Education.

Sajtos, L. & Mitev, A. (2007). SPSS kutatási és adatelemzési kézikönyv. Budapest: Alinea Kiadó.

Statista (2019). *Revenues from the artificial intelligence (AI) market worldwide from 2016 to 2025 (in million U.S. dollars)*. Retrieved from https://www.statista.com/statistics/607716/worldwide-artificial-intelligence-market-revenues/

Wang, Y. & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology, 114*(2), 246-257. https://doi.org/10.1037/pspa0000098

Wilson, B., Hoffman, J., & Morgenstern, J. (2019). *Predictive Inequity in Object Detection*. Retrieved from https://arxiv.org/pdf/1902.11097.pdf

Yampolskiy, R. V., & Spellchecker, M. S. (2016). *Artificial Intelligence Safety and Cybersecurity: a Time of AI Failures*. Retrieved from https://arxiv.org/ftp/arxiv/papers/1610/1610.07997.pdf

Zhang, J., & Zulkernie, M. (2016). *A hybrid network intrusion detection technique using random forests.* Proceedings of the First International Conference on Availability, Reliability and Security (ARES'06), pp. 262-269.