MAB MAANDBLAD VOOR ACCOUNTANCY EN BEDRIJFSECONOMIE

Maandblad voor Accountancy en Bedrijfseconomie 94(5/6) (2020): 219–230
DOI 10.5117/mab.94.47158

Research Article

The application of Artificial Intelligence in banks in the context of the three lines of defence model

Alette Tammenga

Received 8 October 2019 | **Accepted** 20 January 2020 | **Published** 30 June 2020

Abstract

The use of Artificial Intelligence (AI) and Machine Learning (ML) techniques within banks is rising, especially for risk management purposes. The question arises whether the commonly used three lines of defence model is still fit for purpose given these new techniques, or if changes to the model are necessary. If AI and ML models are developed with involvement of second line functions, or for pure risk management purposes, independent oversight should be performed by a separate function. Other prerequisites to apply AI and ML in a controlled way are sound governance, a risk framework, an oversight function and policies and processes surrounding the use of AI and ML.

Relevance to practice

The use of Artificial Intelligence and Machine Learning in the banking industry is increasing. What do these techniques entail? What are their main applications and what are the risks concerned? Is the three lines of defence model still fit for purpose when using these techniques? These are the topics that will be addressed in this article.

Keywords

Artificial intelligence, banks, machine learning, risk management, three lines of defence, governance

1. Introduction

Technology and data are playing an increasingly important role in the banking industry. While Artificial Intelligence (AI) was initially mostly used in client servicing domains of the bank, more and more applications for risk management purposes can be observed.

A common model to use within banks is the three lines of defence (3LoD) model. This model consists of a first line in the business, being responsible for managing risks, a second line risk management function in an oversight role and a third line function: internal audit. Given the expanding use of AI and machine learning (ML) within banks, the question arises whether this 3LoD model is still fit for purpose given these new developments, or if changes to the model are necessary.

This article aims to answer the question: "How can the application of Artificial Intelligence and Machine learning techniques within banks be placed in the context of the Three lines of defence model?"

This article will first address the basic concepts of AI and ML and the 3LoD model. It will then give an overview of the applications observed throughout banks and the risks and challenges of using AI and ML. After that, AI and ML are placed in the context of the 3LoD model, addressing the prerequisites to apply AI and ML in a controlled way. The article finishes with a regulatory view, the emergence of potential new market wide risks, conclusions and recommendations.

Copyright *Alette Tammenga*. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY-NC-ND 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Amsterdam University Press

2. Artificial Intelligence and Machine Learning: basic concepts

As a start, it is important to clarify the concepts of Artificial Intelligence (AI) and Machine Learning (ML), which are often interchanged. Several definitions can be found. AI is mostly viewed as intelligence demonstrated by machines, with intelligence being defined with reference to what we view intelligence as in humans (Turing 1952 cf Shieber 2004 in Aziz and Dowling 2019). Or another definition: AI refers to machines that are capable of performing tasks that, if performed by a human, would be said to require intelligence (Scherer 2016).

AI uses instances of *Machine Learning* as components of the larger system. These ML instances need to be organized within a structure defined by domain knowledge, and they need to be fed data that helps them complete their allotted prediction tasks (Taddy 2018). As such, ML delivers the capability to detect meaningful patterns in data, and has become a common tool for almost any task faced with the requirement of extracting meaningful information from data sets (Leo et al. 2019). ML may also be defined as a method of designing a sequence of actions to solve a problem, known as algorithms which optimise automatically through experience and with limited or no human intervention (FSB 2017). ML is limited to predicting a future that looks like the past, they are a tool for pattern recognition (Taddy 2018). According to Mullainathan and Spiess (2017), the appeal of ML is that it manages to uncover generalizable patterns. In fact, the success of ML at intelligence tasks is largely due to its ability to discover complex structure that was not specified in advance. It manages to fit complex and very flexible functional forms to the data without simply overfitting; it finds functions that work well out-of-sample (Mullainathan and Spiess 2017). So ML is a core technique of AI, learning from data, but AI often involves additional techniques and requirements (Aziz and Dowling 2019). So as Taddy (2018) states, AI is a broader concept, meaning that an AI system is able to solve complex problems that have been previously reserved for humans. It does this by breaking these problems into a bunch of simple prediction tasks, each of which can be attacked by a 'dumb' ML algorithm.

As Reddy (2018) states, ML comprises a broad range of analytical tools, which can be categorized into '*supervised*' and *'unsupervised'* learning tools. Supervised learning is an approach to ML where the historical input data is tagged with its corresponding business outcomes and the ML solution is expected to identify and learn the patterns in the input data associated with a business outcome and self-develop an algorithm based on this learning to predict a business outcome for a future instance. So supervised ML involves building a statistical model for predicting or estimating an output based on one or more inputs (e.g., predicting GDP growth based on several variables). The supervised learning approach usually operates with a classification aim (e.g. will a loan default yes or no) or based on regression, in which a quantified value is predicted (e.g. what is the probability of loan default) (Reddy 2018).

In unsupervised learning, a dataset is analysed without a dependent variable to estimate or predict. Rather, the data is analysed to show patterns and structures in a dataset (Van Liebergen 2017). So the historical input data is fed into the ML solution without any tagging of the business outcomes and the solution is expected to decipher or self-develop an algorithm for prediction based on its own interpretations of the patterns in the data without any guidance or indicators. The unsupervised learning approach usually performs via Clustering (e.g. of customers in segments for credit risk) or Association (e.g. impact of increased draw-down on credit lines prior to default) (Reddy 2018).

So the main difference between supervised and unsupervised ML is the tagging of historical data with business outcomes in supervised learning, where this is not done in unsupervised learning. '*Reinforcement learning*' falls in between supervised and unsupervised learning. In this case, the algorithm is fed an unlabelled set of data, chooses an action for each data point, and receives feedback (perhaps from a human) that helps the algorithm learn. For instance, reinforcement learning can be used in robotics, game theory, and self-driving cars (FSB 2017).

In discussions about AI, the concept of *deep learning* or *neural networks* is also mentioned often. In deep learning, multiple layers of algorithms are stacked to mimic neurons in the layered learning process of the human brain. Each of the algorithms is equipped to lift a certain feature from the data. This so-called representation or abstraction is then fed to the following algorithm, which again lifts out another aspect of the data. The stacking of representation-learning algorithms allows deep-learning approaches to be fed with all kinds of data, including low-quality, unstructured data; the ability of the algorithms to create relevant abstractions of the data allows the system as a whole to perform a relevant analysis. Crucially, these layers of features are not designed by human engineers, but learned from the data using a general-purpose learning procedure. They are also called 'hidden layers' (Van Liebergen 2017). Deep learning can be both supervised and unsupervised forms of learning, depending on the purpose for which it is applied. Deep learning techniques are complex, they are often perceived as a black box. It is not always clear how inputs have been recombined to create a predicted output (Aziz and Dowling 2019). This has obvious implications for use in risk management, the presence of a black box in decision making has its own challenges and can be a risk in itself.

Also, the concepts of *predictive* versus *prescriptive* AI are relevant. Predictive AI is about understanding and predicting the future, so about using statistical models and forecast techniques to understand the future to predict what could happen. Prescriptive AI uses optimization and simulation algorithms to advice on possible outcomes and to instigate what action to take.

Other concepts within AI are speech recognition and Natural Language Processing (NLP). This is the ability to understand and generate human speech the way humans do by, for instance extracting meaning from text or generating text that is readable, stylistically natural and grammatically correct (Deloitte 2018).

One could wonder in which way AI and ML are different from more traditional statistical modelling techniques. Statistical modelling gives insight in correlation, derives patterns in the data using mathematics. It is a formalization of relationships between variables in the form of mathematical equations.The main difference compared to AI/ML is that the ML model trains itself using algorithms, it can learn from data without relying on rule based programming (Srivastava 2015). ML requires almost no human intervention because it is about enabling a computer to learn on its own from a large set of data without any set instructions from a programmer. It explores the various observations and creates definite algorithms that are self-sufficient enough to learn from data as well as make predictions (Mittal 2018).

Figure 1. Applications of AI in practice: examples in banks.

Topic	Application	In practice
Stress testing	Improving stress testing models by using AI and ML	Limit the number of variables used in a scenario analysis
Model validation	Automated validation of models	Less human involvement in model validation processes
Market Risk	Monitoring traders	Surveillance of conduct breaches by traders
Capitalisation	Optimizing regulatory capital	Machine learning tools can increase efficiency and speed of capital optimization
Compliance	Transaction monitoring to detect money laundering	Detecting patterns of suspicious transactions
Credit approval	Automated credit approval	Analyse and interpret patterns that lead to credit approval to improve the credit approval process
Compliance	Fraud detection	Detecting anomalies or patterns in large volumes of transaction data
Market Risk	Portfolio management	Monitoring volatility from a portfolio management perspective

3. The three lines of defence model

In the 3LoD Defence model (IIA 2013):

1. management control is the first line of defence in risk management: they own and manage risks;
2. the various risk control and compliance oversight functions established by management are the second line of defence: they oversee risks;
3. an independent audit function is the third: they provide independent assurance.

Regardless of how the 3LoD model is implemented, senior management and governing bodies should clearly communicate the expectation that information should be shared and activities coordinated among each of the groups responsible for managing the organization's risks and controls (IIA 2013).

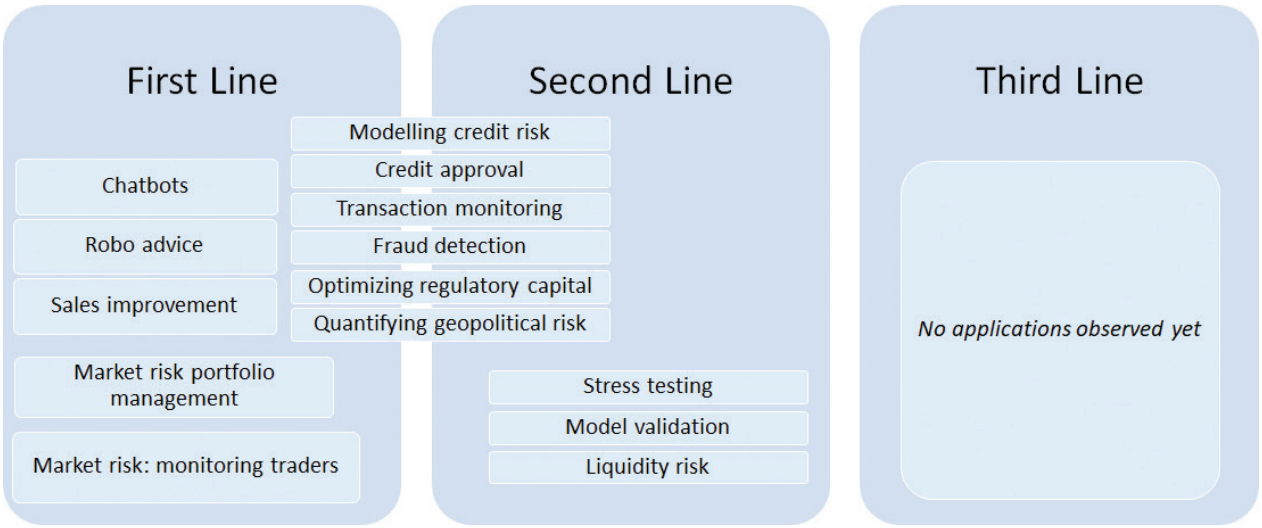
The 3LoD model has also received criticism. The core concern according to Davies and Zhivitskaya (2018) is that the existence of three separate groups who are supposed to ensure proper conduct towards risks has led to a false sense of security. If several people are in charge, no one really is. Different criticism addresses that the 3LoD model could downplay the importance of strong risk management in the business areas themselves: "not enough emphasis is placed on the first line of defence which is management" or that it could lead to an excessively bureaucratic, costly, and demotivating approach to risk management. The Financial Stability Institute (2017) also mentions weaknesses in the 3LoD model. The responsibility for risk in the first line conflicts with their primary task which is generating sufficient revenues and profit, which requires risk-taking. So there are misaligned incentives here. In other cases, second line functions may not be sufficiently independent, or lack sufficient skills and expertise to effectively challenge practices and controls in the first line (Arndorfer and Minto 2015). Lim et al. (2017) state that whilst the 3LoD model has formally spread the responsibility for risk management across different organisational lines, a real impact on the hierarchy within the organisation is not observed enough yet: often, traders are perceived as more valuable to the organisation than risk and compliance personnel (Lim

et al. 2017). Supporters of 3LoD argue that, while these criticisms may have been valid in the past, the system has been made stronger since the Global Financial Crisis (Davies and Zhivitskaya 2018). When placing the use of AI and ML into the context of the 3LoD model, the criticism should be kept in mind.

4. Applications of AI and ML within banks

To get a better insight in the risks associated with using AI and ML, this section addresses some use cases of AI and ML within banks throughout all of the 3LoD functions. These are depicted in figure 2 as well.

Figure 2. Applications of AI and ML in the 3LoD in banks.



4.1 Applications in the first line

AI and ML techniques are frequently used in servicing clients. Applications such as chatbots for e.g. customer support or robo advice (digital platforms that provide automated, algorithm-driven financial planning services with little to no human supervision) have increased in the past years. A big 4 audit firm has developed a voice analytics platform that uses deep learning and various ML algorithms to monitor and analyse voice interactions, and identify high risk interactions through Natural Language processing. The interactions are then mapped to potential negative outcomes such as complaints or conduct issues and the platform then provides details as to why they have occurred (Deloitte 2018). Automated financial advice based on AI and ML techniques is also observed in an increasing number of financial institutions, but is more prevalent for securities than for banking products (González-Páramo 2017). Also, some banks use AI and ML to improve how they sell to clients. Both external market data and internal data on clients is used to develop risk advisory robots that offer advanced insights into client needs. The techniques being explored aim to help banks predict client behaviour, identify market opportunities, extract information from news and websites, and alert sales based on market triggers (Sherif 2019).

In the field of market risk, the use cases of ML from a risk management perspective appear to be limited and are mainly observed in first line functions. Here, the focus is on e.g. market volatility or market risk from a portfolio or investment risk management perspective. Also, ML is increasingly being applied within financial institutions for the surveillance of conduct breaches by traders working for the institution. Examples of such breaches include rogue trading, benchmark rigging, and insider trading – trading violations that can lead to significant financial and reputational costs for financial institutions (Van Liebergen 2017). In terms of the 3LoD, these applications occur purely in the first line of defence. From a bank risk management perspective, the papers appear limited.

4.2 Applications by first and second line

Modelling credit risk has been standard practice for several years already. In banks, such models are developed within a modelling department that is often part of a risk management function, with the involvement of business users. The model is used by the business in the first line. The general approach to credit risk assessment has been to apply a classification technique on past customer data, including delinquent customers, to analyse and evaluate the relation between the characteristics of a customer and their potential failure. This could be used to determine classifiers that can be applied in the categorization of new applicants or existing customers as good or bad (Leo et al. 2019). Enhancing the existing models with ML applications increases the quality of the models and therefore, the accurate predictions of e.g. default. The aim is to better identify the early signs of credit deterioration at a client or the signs for an eventual default based on time series

data of defaults. When the accuracy of creditworthiness prediction increases, the loan portfolio could grow and become more profitable. ML techniques can be effectively used for Regression based forecasting as well. Primarily, forecasting models for Probability of Default (PD), Loss Given Default (LGD) and Credit Conversion Factor (CCF) can show greater levels of accuracies in forecasting the quantum of risk with greater degree of precision and accuracy (Reddy 2018). Predominant methods to develop models for PD are classification and survival analysis, with the latter involving the estimation of whether the customer would default and when the default could occur. Classifier algorithms were found to perform significantly more accurately than standard logistic regression in credit scoring. Also, advanced methods were found to perform extremely well on credit scoring data sets such as artificial neural networks (Leo et al. 2019). For consumer credit risk, the outperformance of ML techniques compared to traditional techniques was shown based on research of Khandani et al. (2010): they developed a ML model for consumer credit default and delinquency which turned out to be surprisingly accurate in forecasting credit events 3–12 months in advance. When tested on actual lending data, the model lead to cost savings in total losses of up to 25% (Khandani et al. 2010). In SME lending, Figini et al. (2017) show that a multivariate outlier detection ML technique improved credit risk estimation using data from UniCredit Bank (Figini et al. 2017). Clustering techniques in ML can also benefit the required segmentation of retail clients into pool of loans exhibiting homogenous characteristics (Reddy 2018).

In the field of credit risk, ML is used not only for predicting payment problems or default but also in the credit approval process in the first line. ML could help analyse and interpret a pattern associated with approvals and develop an algorithm to predict it more consistently (Reddy 2018).

Within the Operational risk domain, a field where ML is frequently used is Transaction monitoring as part of anti-money laundering. This is performed in the first line, with the second line Compliance function involved. ML techniques are able to detect patterns surrounding suspicious transactions based on historical data. Clustering algorithms identify customers with similar behavioural patterns and can help to find groups of people working together to commit money laundering. Also, fraud detection can be improved by using ML techniques. Models are estimated based on samples of fraudulent and legitimate transactions in supervised detection methods while in unsupervised detection methods outliers or unusual transactions are identified as potential cases of fraud. Both seek to predict the probability of fraud in a given transaction (Leo et al. 2019).

Optimization of bank's regulatory capital with ML is another use case. AI and ML tools build on the foundations of computing capabilities, big data, and mathematical concepts of optimization to increase the efficiency, accuracy, and speed of capital optimization (FSB 2017). Deutsche Bank has created an AI/ML tool to quantify geopolitical risk and predict its effect on financial markets by mining global financial news creating a picture of a country's political risk profile (Kaya 2019).

4.3 Application in the second and third line

Liquidity risk has limited use cases (Leo et al. 2019). One of the largest asset managers has recently shelved a promising AI Liquidity risk model because they have not been able to explain the models' output to senior management (Kilburn 2018). In a study of Tavana et al. (2018), the authors proposed an assessment method of liquidity risk factors based on ML. They focused on the concept of solvency as definition of the liquidity risk, focusing on loan-based liquidity risk prediction issues. "A case study based on real bank data was presented to show the efficiency, accuracy, rapidity and flexibility of data mining methods when modeling ambiguous occurrences related to bank liquidity risk measurement. The ML implementations were capable of distinguishing the most critical risk factors and measuring the risk by a functional approximation and a distributional estimation. Both models were assessed through their specific training and learning processes and said to be returning very consistent results." (Tavana et al. 2018).

Application of AI and ML for Model risk management purposes is expected to increase. A few use cases have been observed for model validation, where unsupervised learning algorithms help model validators in the ongoing monitoring of internal and regulatory stress-testing models, as they can help determine whether those models are performing within acceptable tolerances or drifting from their original purpose (FSB 2017). Model validation is in practice often performed by a separate function within the second line.

Similarly, AI and ML techniques can also be applied to stress testing. The increased use of stress testing following the financial crisis has posed challenges for banks as they work to analyse large amounts of data for regulatory stress tests. In one use case, AI and ML tools were used for modelling capital markets business for bank stress testing, aiming to limit the number of variables used in scenario analysis for 'Loss Given Default" and "Probability of Default" models. By using unsupervised learning methods to review large amounts of data, the tools can document any bias associated with selection of variables, thereby leading to better models with greater transparency (FSB 2017). The research into the area of stress testing and tail risk capture appears limited (Leo et al. 2019). Comparable to model validation, stress testing is often performed by a separate function in the second line.

According to Leo et al. (2019), much of the other areas of non-financial risk management, country risk management, compliance risk management — aside from money laundering related uses — and conduct risk cases haven't been explored adequately.

No Applications of AI and ML have been observed in the third line yet.

4.4 Benefits of using AI and ML

Obviously, a number of benefits arise from the use of AI and ML. The techniques may enhance machine-based processing of various operations in financial institutions, thus increasing revenues and reducing costs (FSB 2017). Kaya (2019) shows that AI has had a significant positive impact on European banks' return on assets (ROA): "AI patents positively impact ROA at statistically significant levels and explain 7% of the variation in bank profitability".

It is expected that the time for data analysis and risk management will decrease, making risk management more efficient and less costly. AI and ML can be used for risk management through earlier and more accurate estimation of risks. For example, to the extent that AI and ML enable decision-making based on past correlations among prices of various assets, financial institutions could better manage these risks. Despite being critiqued for operating like a black box, the ability of ML techniques to analyse volumes of data without being constrained by assumptions of distribution and deliver much value in exploratory analysis, classification and predictive analytics, is significant (Leo et al. 2019). Also, meeting regulatory requirements could become more efficient by automating repetitive reporting tasks and by the increased ability to organize, retrieve and cluster non-conventional data such as documents (Aziz and Dowling 2019). But there are also risks and challenges to address, which will be discussed in the next section.

5. Risks and challenges when using AI and ML

As depicted in figure 3, there are quite a few risks that need to be addressed when using AI and ML techniques.

5.1 Modelling and data issues

As Aziz and Dowling (2019) mention, the availability of suitable data is very important. Banks are struggling to organize the internal data that they have. The data is usually scattered across different systems and departments throughout the bank. Also, internal or external regulations could prevent the sharing of the data and informal knowledge within a bank is often not present in datasets at all.

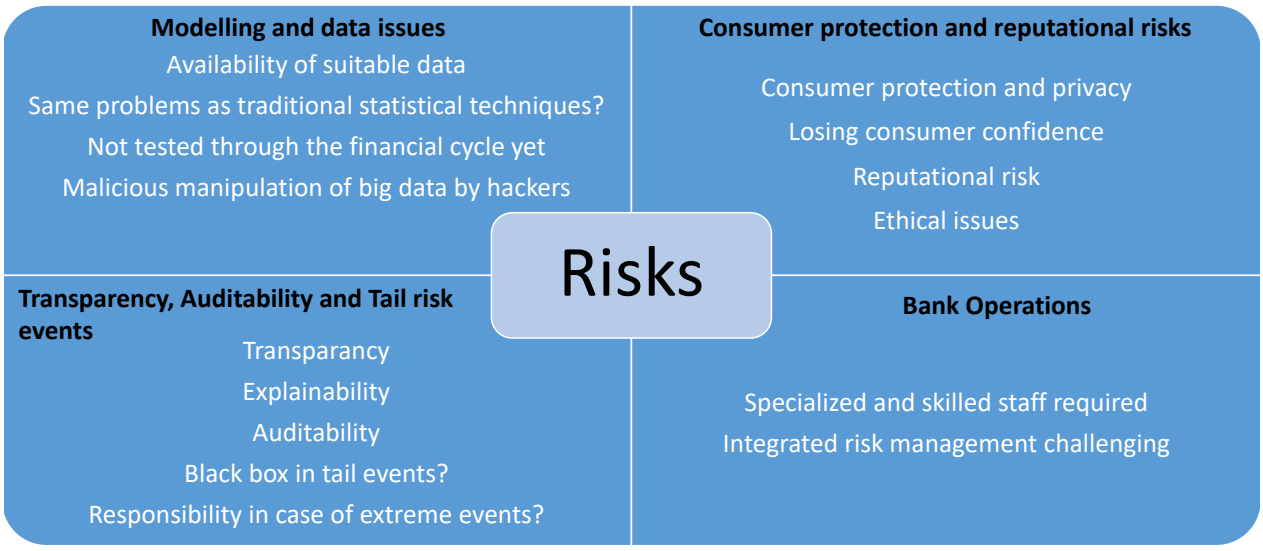
As ML bases much of the modelling upon learning from available data, it could be prone to the same problems and biases that affect traditional statistical methods. As machine-learning methods are compared to traditional statistical techniques, it would be beneficial to evaluate and understand how problems inherent to traditional statistical research methods fare when treated by ML techniques (Leo et al. 2019). An AI ML model could fail if it is not properly trained for all eventualities or in case of poor training data (Van der Burgt 2019).

The lack of information about the performance of these models in a variety of financial cycles, has been noted by authorities as well. AI and ML based tools might miss new types of risks and events because they could potentially 'over train' on past events. The recent deployment of AI and ML strategies means that they remain untested at addressing risk under shifting financial conditions (FSB 2017).

DNB (Van der Burgt 2019) points out that in the financial sector, due to cultural and legal differences, very specific data environments exist, that are often only representative for domestic markets. "This may provide a challenge for the development of data-hungry AI systems, especially for relatively small markets as that of the Netherlands".

According to DNB (Van der Burgt 2019), historical data could quickly become less representative because of continuous changes to the financial regulatory framework. This makes the data not usable for training AI-enabled systems.

Figure 3. Risks and challenges when using AI and ML.



5.2 Consumer protection and reputational risks

Then there is the issue of consumer protection. All processing of personal data has to be authorized by the consumer and be subject to privacy and security standards (González-Páramo 2017). Two parts of the General Data Protection Regulation (GDPR) are directly relevant to ML: the right to non-discrimination and the right to explanation. GDPR article 22 places restrictions on automated individual decision making that 'significantly affect' users. This also includes profiling, meaning algorithms that make decisions based on user-level predictors. So if the outcome of the decision significantly (or in a legal way) affects the user, it is prohibited to decide based solely on automated processing, including profiling (apart from a few exceptions mentioned). Also, users can ask for an explanation of an algorithmic decision that significantly affects them (Goodman 2017). According to Kaya (2019), the intervention of human programmers might be required in order to be fully compliant with these GDPR rules, which is considered a setback for the expected efficiency gains of AI.

A risk that is also present here is losing consumer confidence and reputational risk arising from AI and ML decisions that might negatively affect customers. Efforts to improve the interpretability of AI and ML may be important conditions not only for risk management, but also for greater trust from the general public as well as regulators and supervisors in critical financial services (FSB 2017). DNB (Van der Burgt 2019) also points towards the serious reputation effects that incidents with AI could have.

There are also ethical issues when using AI and ML. AI could adopt societal biases. "Even if all data is tightly secured and AI is kept limited to its intended use, there is no guarantee that the intended use is harm free to consumers. Predictive algorithms often assume there is a hidden truth to learn, which could be the consumer's gender, income, location, sexual orientation, political preference or willingness to pay. However, sometimes the to-be-learned 'truth' evolves and is subject to external influence. In that sense, the algorithm may intend to discover the truth but end up defining the truth. This could be harmful, as algorithm developers may use the algorithms to serve their own interest, and their interests – say earning profits, seeking political power, or leading cultural change – could conflict with the interest of consumers" (Jin 2018). Discrimination based on race, gender or sexuality is usually hardcoded in e.g. AI and ML techniques concerning credit risk and lending decisions. In deep learning, it is harder to guard that the model is not inadvertently making decisions that go against the hardcoded lines, by means of indirect proxies (Aziz and Dowling 2019). Consumers might be unfairly excluded from access to credit as a result of outdated or inaccurate data or due to incorrect or illegal inferences made by algorithms (González-Páramo 2017). AI could adopt societal biases.

According to Kaya (2019), there is also the risk of potentially malicious manipulation of big data by hackers. AI could be corrupted by malicious intent. If hackers flood systems with fictitious data (e.g. fake social media accounts and fake news), they might influence AI decision making. This makes continuous monitoring by programmers necessary.

5.3 Transparency, Auditability and Tail risk events

There is the issue of transparency. As mentioned above, deep learning techniques might pose a risk in itself, as the 'black box' system hinders effective risk oversight. These techniques are often quite opaque, leading to difficulties in terms of transparency, explainability and auditability towards management of the bank as well as its auditors. It can also cause regulatory compliance issues around demonstrating model validity to auditors and regulators (Aziz and Dowling 2019).

More complex AI algorithms lead to an inability of humans to visualize and understand the patterns. AI algorithms update themselves over time, and are by their nature unable to communicate its reasoning (Kaya 2019). This could become even more challenging when taking regulation into account which is aimed at the internal control structure surrounding financial reporting (Sarbanes Oxley) and requirements regarding effective risk data aggregation and risk reporting (BCBS 239). Sarbanes Oxley requires effective controls to be in place for financial reporting, so as to make every step in the process of reporting annual statements and other disclosures auditable. BSBS239 goes a step further in requiring clear, documented and tested data lineage for all risk data that is aggregated within a bank. If the reasoning of an AI algorithm cannot be communicated, being compliant with these regulations can become challenging. A solution to this might be the involvement of human programmers and overseers, also this might cancel out efficiency gains (Kaya 2019).

Also, 'black box' techniques could create complications in tail risk events. According to the Financial Stability Board (2017), "Black boxes' in decision-making could create complicated issues, especially during tail events. In particular, it may be difficult for human users at financial institutions – and for regulators – to grasp how decisions, such as those for trading and investment, have been formulated. Moreover, the communication mechanism used by such tools may be incomprehensible to humans, thus posing monitoring challenges for the human operators of such solutions. If in doubt, users of such AI and ML tools may simultaneously pull their 'kill switches', that is manually turn off systems. After such incidents, users may only turn systems on again if other users do so in a coordinated fashion across the market. This could thus add to existing risks of system-wide stress and the need for appropriate circuit-breakers. In addition, if AI and ML based decisions cause losses to financial intermediaries across the financial system, there may be a lack of clarity around responsibility" (FSB 2017).

5.4 Bank operations

Specialized and skilled staff is required to implement new techniques such as AI and ML. It might be challenging to attract sufficient personnel possessing these specific skills. At Board of directors' level, sufficient knowledge should be present, enabling the Board to assess the risks of AI. Second line personnel should be trained to understand AI specific challenges and risks. Personnel working with AI applications should be made aware of the strengths and limitations (Van der Burgt 2019).

When there is some or full automation of the process from data gathering to decision making, human oversight is essential. This becomes more necessary as the level of automation rises, or when ML techniques become more prescriptive.

When taking all of the risks mentioned above into account, it seems apparent that the use of AI and ML techniques also brings about extra challenges in the context of the common ambition of integrated risk management within banks. Use cases being dispersed throughout different parts of the bank could hinder integrated risk management and an integrated approach towards these risks.

6. AI and ML in the context of the three lines of defence

As shows from the use cases mentioned above, AI and ML can be used within each of the 3LoD, or throughout multiple lines. It appears that the techniques are most used within the first line, or in use cases where first and second line are both involved.

If used purely in the first line, the 3LoD model can be applied as designed. In this case, it is important to safeguard that sufficient knowledge of the techniques and its use is also present in second and third line functions, to ensure compliance, to identify and manage risks, to challenge the first line on replicability of decisions and validity of the model and to perform audits effectively. As mentioned above, the scarcity of resources with the required skills and knowledge can be an issue (FSB 2017).

For a number of applications, such as credit risk modelling and approval, transaction monitoring or fraud detection, both the first and the second line are involved. Here it gets more difficult to apply the 3LoD model. Depending on the nature of the involvement of the second line function, e.g. whether they are developing AI ML tools themselves, there should be an independent function involved that provides independent validation and challenge. So applying the 3LoD model without any adjustments does not seem wise in this case. When zooming in on the second line risk management function, this function "*facilitates and monitors the implementation of effective risk management practices by operational management and assists risk owners in defining the target risk exposure and reporting adequate risk-related information throughout the organization*" (IIA 2013). So if the risk management function is operationally involved in e.g. developing a model using AI and ML techniques, or the AI and ML model is developed for purely second line purposes such as in model risk management or stress testing, an alternative solution is warranted. In this case, as a minimum, independent oversight, challenge, validation and assurance should be safeguarded by a separate function performing this second line role. In addition, the internal audit function must also be involved. No use cases have been found in purely third line functions but if AI and ML techniques were to be used, external assurance surrounding the use of AI and ML is warranted.

A potential better way of ensuring a controlled deployment of AI and ML techniques, which is at the same time in line with the principles of the 3LoD model is to assign specific roles (Burt et al. 2018):

- "Data Owners: Responsible for the data used by the models.
- Data Scientists: Create and maintain models.
- Business owners: Possess subject matter expertise about the problem the model is being used to solve.
- Validators: Review and approve the work created by both data owners and data scientists, with a focus on technical accuracy.
  This could be performed by an independent function, or if the size of the bank is insufficient, by data scientists who are not associated with the specific model or project at hand.
- Governance Personnel: Review and approve the work created by both data owners and data scientists, with a focus on legal risk."

Together with the business owners, a group of data owners and data scientists comprise the first line of defence. The validators comprise the second line of defence, together with the governance personnel. The third line function could be performed by independent internal auditors, provided that they have the expertise needed. This set up is necessary to safeguard an effective challenge throughout the model lifecycle by multiple parties, separate from the model developers. In assigning these specific roles, the principles of the 3LoD model are safeguarded.

Some other points are relevant when thinking about AI and ML in the context of the 3LoD model and controlled application. All ML projects should start by clearly documenting initial objectives and underlying assumptions, which should also include major desired and undesired outcomes. This should be circulated and challenged by all stakeholders. Data scientists, for example, might be best positioned to describe key desired outcomes, while legal personnel might describe specific undesired outcomes that could give rise to legal liability. "Such outcomes, including clear boundaries for appropriate use cases, should be made obvious from the outset of any ML project. Additionally, expected consumers of the model — from individuals to systems that employ its recommendations – should be clearly specified as well" (Burt et al. 2018).

The materiality of the model that is deployed should be taken into account in all three lines (Burt et al. 2018). This means that the intensity and frequency of involvement of second and third line functions, or validators and governance personnel, should be based on the impact that the model has within the banks or towards its clients.

How 'black box' the AI technique is, is often a result of choices made by developers of the model. Predictive accuracy and explainability are frequently subject to a trade-off; higher levels of accuracy may be achieved, but at the cost of decreased levels of explainability. This trade off should be documented from the start, and challenged by other functions. "Any decrease in explainability should always be the result of a conscious decision, rather than the result of a reflexive desire to maximize accuracy. All such decisions, including the design, theory, and logic underlying the models, should be documented as well" (Burt et al. 2018). Note that using Deep learning techniques requires even more specific knowledge throughout the 3LoD.

When viewing the significant amount of risks in using AI and ML as described above, and the challenges when it comes to applying the 3LoD model, a sound governance surrounding the use of AI and ML is essential. The risks concerned need to be properly identified, assessed, controlled and monitored. This also means clearly defining the roles and responsibilities for the functions involved, be it in the first, second or third line of defence. "Any uncertainty in the governance structure in the use of AI and ML might increase the risks to financial institutions" (FSB 2017). Given the challenge to view all risk integrally, a dedicated oversight function of all AI and ML use throughout the bank is required, especially for larger banks. A sound framework is necessary to create, deploy and maintain AI and ML techniques in a controlled way and to manage the risks involved properly. It is also important to develop policies and processes for the use of AI and ML, ensuring that the deployment of these techniques fit the strategy and risk appetite of the bank. "Any uncertainty in the governance structure could substantially increase the costs for allocating losses, including the possible costs of litigation" (FSB 2017). As part of sound governance, a sound model risk management framework is also necessary and it should be updated or adjusted for AI/ML models. Given all of the risks mentioned above and the self-learning nature of AI/ML models, extra attention is warranted. As Asermely (2019) describes it: "The dynamic nature of machine learning models means they require more frequent performance monitoring, constant data review and benchmarking, better contextual model inventory understanding, and well thought out and actionable contingency plans". Given increasing volumes and complexity of data, increasing use of AI/ML and the growing complexity of AI/ML, sound governance will also be increasingly important towards the future (Asermely 2019).

7. AI and ML in banks: the regulatory perspective and new risks

According to the Financial Stability Board (2017), because AI and ML applications are relatively new, there are no known dedicated international standards in this area yet. Apart from papers on this topic published by regulatory authorities in Germany, France, Luxembourg, The Netherlands and Singapore, no European or international standards were published. Although calls to regulate AI and ML are heard more often, the current regulatory framework is not designed with the use of such tools in mind. Some regulatory practices may need to be revised for the benefits of AI and ML techniques to be fully harnessed. "In this regard, combining AI and ML with human judgment and other available analytical tools and methods may be more effective, particularly to facilitate causal analysis" (FSB 2017). DNB (Van der Burgt 2019) states "Given the inherent interconnectivity of the financial system, the rise of AI has a strong international dimension. An adequate policy response will require close international cooperation and clear minimum standards and guidelines for the sector to adhere to. Regulatory arbitrage in the area of AI could have dangerous consequences and should be prevented where possible."

DNB recently published a set of general principles for the use of AI in the financial sector (Van der Burgt 2019). The principles are divided over six key aspects of responsible use of AI, namely soundness, accountability, fairness, ethics, skills and transparency.

"The Basel Committee on Banking Supervision (BCBS) notes that a sound development process should be consistent with the firm's internal policies and procedures and deliver a product that not only meets the goals of the users, but is also consistent with the risk appetite and behavioural expectations of the firm. In order to support new model choices, firms should be able to demonstrate developmental evidence of theoretical construction; behavioural characteristics and key assumptions; types and use of input data; numerical analysis routines and specified mathematical calculations; and code writing language and protocols (to replicate the model). Finally, it notes that firms should establish checks and balances at each stage of the development process" (FSB 2017).

Many of the use cases described in this article could result in improvements in risk management, compliance, and systemic risk monitoring, while potentially reducing regulatory burdens. AI and ML can continue to be a useful tool for financial institutions by implementing so called "RegTech", aiming to facilitate regulatory compliance more efficiently and effectively than existing capabilities. The same goes for supervisors via "SupTech", which is the use of AI and ML by public sector regulators and supervisors. The objective of "SupTech" is to enhance efficiency and effectiveness of supervision and surveillance (FSB 2017).

From a market wide perspective, there are also potential new and/or systemic risks to take into account when using AI and ML techniques. If a similar type of AI and ML is used without appropriately 'training' it or introducing feedback, reliance on such systems may introduce new risks. For example, if AI and ML models are used in stress testing without sufficiently long and diverse time series or sufficient feedback from actual stress events, there is a risk that users may not spot institution-specific and systemic risks in time. These risks may be pronounced especially if AI and ML are used without a full understanding of the underlying methods and limitations. "Tools that mitigate tail risks could be especially beneficial for the overall system" (FSB 2017).

A more hypothetical issue is that models used by different banks might converge on similar optimums for trading causing systemic risk as well (Aziz and Dowling 2019). "Greater interconnectedness in the financial system may help to share risks and act as a shock absorber up to a point. Yet if a critical segment of financial institutions rely on the same data sources and algorithmic strategies, then under certain market conditions a shock to those data sources could affect that segment as if it were a single node and thus could spread the impact of extreme shocks. The same goes for several financial institutions adopting a new strategy exploiting a widely-adopted algorithmic strategy. As a result, collective adoption of AI and ML tools may introduce new risks" (FSB 2017).

"AI and ML may affect the type and degree of concentration in financial markets in certain circumstances. For instance, the emergence of a relatively small number of advanced third-party providers in AI and ML could increase concentration of some functions in the financial system" (FSB 2017). DNB states that "Given the increasing importance of tech giants in providing AI-related services and infrastructure, the concept of systemic importance may also need to be extended to include these companies at some point" (Van der Burgt 2019). The role of Big-Tech companies requires attention here. "Many BigTech firms also offer specific tools using artificial intelligence and machine learning to corporate clients, including financial institutions. The activity of BigTech firms as both suppliers to, and competitors with financial institutions raises a number of potential conflicts of interest, at the same time that their dominant market power in some markets is coming under greater scrutiny" (Frost et al. 2019).

"The lack of interpretability or 'auditability' of AI and ML methods has the potential to contribute to macro-level risk if not appropriately audited. Many of the models that result from the use of AI or ML techniques are difficult or impossible to interpret". Auditing of models may require skills and expertise that may not be present sufficiently at the moment. "The lack of interpretability may be overlooked in various situations, including, for example, if the model's performance exceeds that of more interpretable models. Yet the lack of interpretability will make it even more difficult to determine potential effects beyond the firms' balance sheet, for example during a systemic shock. Notably, many AI and ML developed models are being 'trained' in a period of low volatility. As such, the models may not suggest optimal actions in a significant economic downturn or in a financial crisis, or the models may not suggest appropriate management of long-term risks" (FSB 2017).

8. Conclusion and recommendations

Artificial Intelligence (AI) refers to machines that are capable of performing tasks that, if performed by a human, would be said to require intelligence. AI uses instances of *Machine Learning (ML)* as components of a larger system. ML is able to detect meaningful patterns in data. The main difference when comparing AI ML techniques with more traditional statistical modelling techniques is that the AI/ML model trains itself using algorithms, so it can learn from data without relying on rule based programming or instructions from a human programmer.

Among the most used AI and ML techniques within banks are credit risk modelling- and approval, transaction monitoring regarding Know Your Customer and Anti Money Laundering and fraud detection, which are usually jointly developed by first and second line functions. Frequently observed use cases in the first line are client servicing solutions and market risk monitoring- and portfolio management. The techniques are used to a lesser extent for pure second line risk management purposes until now, while no use cases have been observed for third line functions. It is expected that applications in the risk management and internal audit domain will increase in the years to come.

There are obvious benefits to using AI and ML techniques, they may enhance machine-based processing of various operations in financial institutions, thus increasing revenues and reducing costs. It is expected that the time for data analysis and risk management will decrease, e.g. by earlier and more accurate estimation of risk, making risk management more efficient and less costly. The ability of ML techniques to analyse volumes of data without being constrained by assumptions of distribution is significant. Also, meeting regulatory requirements could become more efficient by automating repetitive reporting tasks and by the increased ability to organize, retrieve and cluster non-conventional data such as documents.

There are also numerous risks and challenges to address. Modelling issues and data issues can occur when insufficient suitable data is available, or when hackers maliciously manipulate big data. Also, the model outcomes have not been tested through a financial cycle yet. There are risks regarding consumer protection and privacy as well as reputational risks stemming from ethical issues. Sufficient specialized and skilled staff is needed within banks and there are numerous risks regarding transparency and auditability.

This article aimed to answer the question: "How can the application of Artificial Intelligence and Machine learning techniques within banks be placed in the context of the Three lines of defence model?"

When AI and ML are placed in the context of the 3LoD model, there are quite some prerequisites to apply AI and ML in a controlled way. If the second line risk management function is involved in the operational development of the model, independent oversight, challenge, validation and assurance should be safeguarded by a separate function performing the second line role. In addition, the internal audit function must be involved. Ensuring the proper functioning of the 3LoD model could also be done by assigning specific roles within each AI/ML project, that safeguard the controlled deployment of AI and ML techniques. Data owners and data scientists comprise the first line of defence, together with the business owner. The second line role could then be comprised of validators and other governance personnel that review and approve the work from a technical and a compliance perspective, respectively. Other prerequisites are a sound governance surrounding the use of AI and ML, clearly defined roles and responsibilities, a dedicated oversight function, a sound model risk management framework, a sound framework for managing all of the risks and policies and processes for the use of AI and ML, ensuring that the deployment of these techniques fit the strategy and risk appetite of the bank.

Collective adoption of AI and ML tools may introduce new systemic risks. If e.g. a critical segment of financial institutions rely on the same data sources and algorithmic strategies, under certain market conditions a shock could affect this entire segment and thus spread the impact of the shock throughout multiple financial institutions. Without sufficiently long and diverse time series or feedback from actual stress events, it is possible that tail risks are not spotted in time. The current regulatory framework does not sufficiently address the field of AI and ML and therefore needs to be revised and updated. This is perceived necessary to address all new risks at hand, as well as the challenges presented regarding the application of the three lines of defence model. In this effort, regulators might leverage on the existing regulation for e.g. credit risk modelling. Risk managers should follow the developments in this field closely, to be able to assess the (new) risks within individual institutions and for the financial system as a whole. Also, sufficiently skilled resources should be available within the internal and external audit community, as to ensure the proper auditing of the techniques deployed by banks.

Taking into account the risks, the application of AI and ML could be expanded in the area of market risk, liquidity risk, model risk management, stress testing and in the third line. Also, the use of AI and MI to manage tail risk could be further investigated. Another area to monitor and possibly further investigate is the role of BigTech companies and their duality in being suppliers of AI and ML technology as well as competitors of banks. Given the expanding use of AI and ML techniques, new issues and risks will undoubtedly emerge and may warrant further research. It is key that existing governance is strengthened and adjusted following these new issues and risks.

■ **A.Z. Tammenga** MSc. is working as a consultant at Transcendent Group Netherlands and is also a student in the Postgraduate program "Risk management for Financial Institutions" at the Free University in Amsterdam.

References

- Arndorfer I, Minto A (2015) The "four lines of defence model" for financial institutions. Financial Stability Institute, Occasional paper No 11. https://www.bis.org/fsi/fsipapers11.htm
- Asermely D (2019) Model risk management – Special report 2019: Machine learning governance. https://www.risk.net/content-hub/model-risk-management-special-report-2019-6764071
- Aziz S, Dowling M (2019) Machine learning and AI for risk management. In Lynn T, Mooney J, Rosati P, Cummins M (eds.) Disrupting Finance, FinTech and strategy in the 21st Century. Palgrave Pivot (Cham): 33-50. https://doi.org/10.1007/978-3-030-02330-0_3
- Burt A, Leong B, Shirrell S, Wang X (2018) Beyond explainability: A practical guide to managing risk in machine learning models. Future of Privacy Forum. https://fpf.org/2018/06/26/beyond-explainability-a-practical-guide-to-managing-risk-in-machine-learning-models/
- Davies H, Zhivitskaya M (2018) Three Lines of Defence: A robust organising framework, or just lines in the sand? Global Policy 9(Supplement 1): 34-42. https://doi.org/10.1111/1758-5899.12568
- Deloitte (2018) AI and risk management: innovating with confidence. Deloitte, Centre for Regulatory Strategy EMEA. https://www2.deloitte.com/global/en/pages/financial-services/articles/ai-risk-management-uk-jump.html
- EBA (European Banking Authority) (2017) Guidelines on internal governance under Directive 2013/36/EU. EBA/GL/2017/11. https://eba.europa.eu/regulation-and-policy/internal-governance/guidelines-on-internal-governance-revised-
- Figini S, Bonelli F, Giovannini (2017) Solvency prediction for small and medium enterprises in banking. Decision Support Systems, 102: 91–97. https://doi.org/10.1016/j.dss.2017.08.001
- Frost J, Gambacorta L, Huang Y, Shin HS, Zbinden P (2019) BigTech and the changing structure of financial intermediation. BIS Working Papers No 779. Bank for International Settlements. https://www.bis.org/publ/work779.htm
- FSB (Financial Stability Board) (2017) Artificial intelligence and machine learning in financial services: Market developments and

financial stability implications. https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/

- González-Páramo JM (2017). Financial Innovation in the digital age: challenges for regulation and supervision. Revista de Estabilidad Financier (mei 2017): 11-37. https://www.bde.es/f/webbde/GAP/Secciones/Publicaciones/InformesBoletinesRevistas/RevistaEstabilidadFinanciera/17/MAYO%202017/Articulo_GonzalezParamo.pdf
- Goodman B, Flaxman S (2017) European Union regulations on algorithmic decision making and a "right to explanation". AI Magazine 38(3): 50-57. https://doi.org/10.1609/aimag.v38i3.2741
- IIA (Institute of Internal Auditors) (2013) The three lines of defense in effective risk management and control. IIA Position Paper. https://global.theiia.org/standards-guidance/recommended-guidance/Pages/The-Three-Lines-of-Defense-in-Effective-Risk-Management-and-Control.aspx
- Jin GZ (2018) Artificial intelligence and consumer privacy. NBER working paper, 24253. http://www.nber.org/papers/w24253
- Kaya O (2019) Artificial intelligence in banking. A lever for profitability with limited implementation to date. Deutsche Bank Research. https://www.dbresearch.com/PROD/RPS_EN-PROD/Artificial_intelligence_in_banking%3A_A_lever_for_pr/RPS_EN_DOC_VIEW.calias?rwnode=PROD0000000000435631&ProdCollection=PROD0000000000495172
- Khandani AE, Kim AJ, Lo AW (2010) Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance 34(11): 2767-2787. https://doi.org/10.1016/j.jbankfin.2010.06.001
- Kilburn, F (2018). BlackRock shelves unexplainable AI liquidity models. Risk.net: https://www.risk.net/asset-management/6119616/blackrock-shelves-unexplainable-ai-liquidity-models
- Leo M, Sharma S, Maddulety (2019). Machine learning in banking risk management: A literature review. Risks 7(1): 1-22. https://doi.org/10.3390/risks7010029
- Lim C, Woods M, Humphrey C, Seow JL (2017) The paradoxes of risk management in the banking sector. British Accounting Review 49(1): 75-90. https://doi.org/10.1016/j.bar.2016.09.002
- Mittal, S. (2018). Analytix Labs. How Machine Learning is different from Statistical modeling? https://www.analytixlabs.co.in/blog/2018/03/07/machine-learning-different-statistical-modeling/
- Mullainathan S, Spiess J (2017) Machine learning: An applied econometric approach. Journal of Economic Perspectives 31(2): 87-106. https://doi.org/10.1257/jep.31.2.87
- Reddy M (2018) Has machine learning arrived for banking risk managers? Global Journal of Computer Science and Technology: Neural & Artificial Intelligence 18(1): 1-3. https://globaljournals.org/GJCST_Volume18/1-Has-Machine-Learning-Arrived.pdf
- Scherer M (2016) Regulating artificial intelligence systems: Risks, challenges, competencies and strategies. Harvard Journal of Law & Technology 29(2): 353-400. https://dx.doi.org/10.2139/ssrn.2609777
- Sherif N (2019). Banks use machine learning to 'augment' corporate sales. Risk.net: https://www.risk.net/derivatives/6375921/banks-use-machine-learning-to-augment-corporate-sales
- Srivastava T (2015) Difference between Machine Learning & Statistical Modeling. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/
- Taddy M (2018) The technological elements of artificial intelligence. National Bureau of Economics Research working paper. http://www.nber.org/papers/w24301
- Tavana M, Abtahi A-R, Di Caprio D, Poortarigh M (2018). An artificial neural network and Bayesian network model for liquidity risk assessment in banking. Neurocomputing 275: 2525-2554. https://doi.org/10.1016/j.neucom.2017.11.034
- Van der Burgt J (2019) General Principles for the use of Artificial Intelligence in the financial sector. De Nederlandsche Bank (Amsterdam). https://www.dnb.nl/en/binaries/General%20principles%20for%20the%20use%20of%20Artificial%20Intelligence%20in%20the%20financial%20sector2_tcm47-385055.pdf
- Van Liebergen B (2017) Machine learning: a revolution in risk management and compliance. The CAPCO Institute Journal of Financial Transformation 45: 177-186. http://www.capco.com/Capco-Institute/Journal-45-Transformation/Machine-learning-a-revolution-in-risk-management-and-compliance