BaFin

Bundesanstalt für
Finanzdienstleistungsaufsicht

Big data and artificial intelligence:

Principles for the use of algorithms in decision-making processes

© Pixabay/Gert Altmann

Inhaltsverzeichnis

Big data and artificial intelligence:
Principles for the use of algorithms in decision-making processes

In the financial market, there is great interest in the use of big data and artificial intelligence (BDAI) – and (complex) algorithms in general.[1] Companies see this as an opportunity to better assess risks and reduce costs, for example. Their customers can benefit from this, too. However, for both sides, the use of BDAI is also associated with risks that must be monitored.

Artificial intelligence as a combination of machine learning and big data | BaFin's technical definition of the term "artificial intelligence" (AI) in its study entitled "Big data meets artificial intelligence" is the combination of big data, computing resources and machine learning (ML).[2] In the case of machine learning, computers are given the ability, thanks to special algorithms, to learn from data and experience. As opposed to rules-based processes, machine learning does not involve a programmer who determines which results are to be achieved with certain constellations of data and how they are to be achieved. This understanding of the term "artificial intelligence" is also part of the Financial Stability Board's (FSB) definition of the term.[3]

BaFin understands an algorithm to be a set of instructions that are generally integrated into a computer programme and that are aimed at solving (optimisation) problems or a category of problems. In addition to the type of algorithms (how the problem is to be solved technically), ML approaches can be differentiated on the basis of the types of results (a distinction is generally made between classification, regression and clustering) and the types of data (there are special approaches for text, language and image data, for instance).

Currently impossible to make a clear distinction between AI and traditional processes | It is impossible, however, based on the aforementioned definition of artificial intelligence, to make a clear distinction between AI on the one hand and traditional statistical processes and the algorithms that are used within this context on the other. Further developing the existing definition of AI is one of the challenges faced by supervisors, regulators and standard-setters in particular.

Considerable complexity, short recalibration cycles and a high level of automation | It is still worth noting three key features which characterise modern artificial intelligence methods and applications and which play a role when observing risks: the considerable complexity of the underlying algorithm, short recalibration cycles and a high level of automation.

[1] BaFin (2018): Big data meets artificial intelligence: Challenges and implications for the supervision and regulation of financial services, URL: https://www.bafin.de/dok/11250046
[2] loc. cit., p. 25 et seq.
[3] FSB (2017): Artificial intelligence and machine learning in financial services: Market developments and financial stability implications, p. 3 et seq., URL: https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/

- Machine learning algorithms are often more complex than those used in traditional statistical processes. This higher level of complexity, which is particularly found in processes such as artificial neural networks, makes it difficult to understand and verify algorithmic results.

- The combination of self-learning algorithms, made possible by machine learning, and (mass) data that is newly available every day leads to ever-shorter recalibration cycles for models and algorithms; the boundaries between calibration and validation are blurred.

- Algorithms are also increasingly being used in some cases to automate non-standardised processes and decisions that are made quickly and in large numbers (scaling).

Minimum requirements for the use of AI as a basis for discussion | In this publication, BaFin has formulated general principles for the use of algorithms in decision-making processes[4] at financial entities since a clear definition of artificial intelligence has not been established to date. The principles particularly apply to algorithms with the aforementioned features.

These principles constitute preliminary ideas for minimum supervisory requirements relating to the use of artificial intelligence and form the basis for discussions with various stakeholders. At the same time, these principles can already serve as guidance for the entities under BaFin's supervision. The aforementioned working definition is used as a basis within this context. The principles primarily apply to algorithms and algorithmic decision-making processes that are characterised, as described above, by a considerable level of complexity, short recalibration cycles and a high level of automation. However, it is essential to note that the principles in this publication do not rule out the fact that certain regulated activities may already be subject to stricter regulations or administrative practices. In such cases, compliance with these rules takes precedence.

I. Conceptual Framework

Considering the process as a whole | Whether algorithms yield usable results depends on how supervised entities incorporate these algorithms into decision-making processes. An algorithm that is suitable for a particular context may yield unusable results in another situation. Furthermore, the results of an algorithm depend on the data that is available and

[4] This also covers the use of algorithms in processes for preparing decisions, such as the quantification and assessment of risks.

the quality of such data. This is why BaFin's supervision focuses on the algorithm-based decision-making process as a whole – from the data source to the data sink and inclusion in the business process.

No general approval process for algorithms | BaFin does not generally grant approval for algorithm-based decision-making processes per se.[5] Instead, BaFin examines and raises objections to such processes in a risk-oriented manner as and where needed – e.g. within the context of the authorisation procedure, ongoing supervision or the supervision of violations of statutory provisions.

In justified exceptional cases, such as internal models used by banks and insurers to determine their regulatory capital requirements, BaFin examines whether the algorithmic processes that are used are appropriate in terms of methodology, calibration and validation, among other things.

Risk-oriented, proportional and technology-neutral | The fundamental supervisory and regulatory principle of "same business, same risk, same rules" means that a risk-oriented, proportional and technology-neutral approach must be taken for the supervision of algorithmic decision-making processes.

Accordingly, more intensive supervision is in order if (additional) risks are associated with the use of an algorithm in decision-making processes.[6] One typical risk is the significant scalability of processes and thus potential errors. As described above, many algorithmic decision-making processes are highly complex and involve short recalibration cycles and a high level of automation. This can make it difficult for both supervised entities and supervisors to understand and check decision-making processes. In addition, there is the risk that the algorithmic results are biased.[7]

Existing rules are supplemented, specified and further developed | The principles in the following section are intended to supplement and specify the existing regulations and administrative practices in place. These principles constitute preliminary ideas for minimum supervisory requirements relating to the use of artificial intelligence. At the same time, these principles can already serve as guidance for the entities under BaFin's supervision. Many of these principles are not completely new but bring about changes in certain areas of principles-based and technology-neutral regulation.

However, it is essential to note that the principles in this publication do not rule out the fact that certain regulated activities may already be subject to stricter regulations or

[5] BaFin (2020): Does BaFin have a general approval process for algorithms? No, but there are exceptions, 28 April 2020, URL: https://www.bafin.de/dok/14009206
[6] Such risks may also become apparent as a result of anomalies found in ongoing supervisory activities or information.
[7] There are also risks that are associated with the decision-making process in question – irrespective of the use of an algorithm.

administrative practices. In such cases, compliance with these rules takes precedence. The application of these principles does not result in an exemption from the applicable legal and supervisory provisions. However, it is possible of course that current administrative practices will be further developed.

The following section outlines the key principles for the use of algorithms in decision-making processes (Chapter 2). These principles are important for the creation and application of the algorithm. Chapter 3 covers the specific principles for the development phase, while Chapter 4 covers the specific principles for the application phase.

II. Key principles

Clear management responsibility | Senior management is responsible for business-wide strategies and guidelines and for rules relating to the use of algorithm-based decision-making processes. Both the potential of such processes and their limits and risks should be taken into consideration and be clearly listed.

Senior management is responsible for all significant business decisions, even if they are based on algorithms. As a result, senior management must have sufficient technical expertise. In addition, reporting lines and reporting formats must be structured in such a way to ensure that communication is risk-appropriate and geared to the specific requirements of the target audience – from the modeller right up to senior management.

In 2017, BaFin had already made it easier for IT specialists to be part of senior management at supervised entities. The aim was to further promote IT expertise within the senior management of companies.

Moreover, the business-wide strategy for using algorithm-based decision-making processes should be reflected in the IT strategy. There must also be staff with the necessary technical knowledge in the independent control functions (e.g. in the compliance and internal audit functions).

Appropriate risk and outsourcing management | Senior management is also responsible for establishing a risk management system that has been adapted for the use of algorithm-based decision-making processes. If applications are used by a service provider, senior management must also set up an effective outsourcing management system. Responsibility, reporting and monitoring structures must be set out clearly within this context.

When establishing an appropriate risk management system, it is advisable to take into account the risks of an algorithmic decision-making process. Risk mitigation measures and

processes should be targeted and applied precisely where risks originate. In addition, the likelihood and potential scale of the damage caused by erroneous algorithm-based decisions should be clearly analysed and the results of this analysis should be documented. Furthermore, a comprehensive framework should be established at companies to take into account all algorithm-based decision-making processes and interdependencies (e.g. "model risk management framework").

Measures to minimise cyber security risks should also be adapted if required. In particular, they must take into account the complexity and data dependency of the algorithms. For example, the use of mass data can lead to poisoning attacks, where input data is changed in a barely visible way.

Use case: application of telematics-based rates in motor insurance

A motor insurer offers certain groups of customers the possibility to opt for telematics-based rates. Surcharges or discounts applied to the basic rate are determined based on the risk profile specific to the policyholder in question. To determine the risk profiles of policyholders, the insurer examines telematics data, such as speed and GPS location, on an ongoing basis and analyses this data using artificial intelligence.

When using personal data, the insurer must comply with data protection requirements (incl. Article 5 of the General Data Protection Regulation – GDPR) (principle of **"compliance with data protection requirements"**). When developing and using the algorithm, the insurer must guarantee that access to this data is restricted to the individuals that require access and must ensure that no unauthorised third parties are able to gain access to this data.

In line with the **"appropriate risk and outsourcing management"** principle, the company must ensure that the data transfer interfaces that are used are secure and are functioning properly. It must also ensure that there is sufficient protection against data manipulation and consider the fact that third-party providers are often involved in interfaces for saving data in the cloud.

If the company uses algorithm-based telematics, it must also have contingency measures in place in the event that technical problems occur or if there is a system failure (principle of **"establishing contingency measures"**). If specific data can no longer be transmitted in real time due to a system failure, the company must be able to use alternative recording systems or grant customers general discounts during the system failure.

Preventing bias | It is essential to ensure that there are no biased results in algorithm-based decision-making processes. Firstly, bias must be prevented in order to be able to reach business decisions that are not based on systematically distorted results; secondly, bias must be prevented in order to rule out bias-based systematic discrimination of certain groups of customers and thus rule out any resulting reputational risks.

This is a key principle since such biases may occur from the development of the process to its application. For example, processes may entail biases if the data used for the application is not of sufficient quality and quantity, if certain characteristics are ruled out or given too much weight for no appropriate reason, or if algorithmic results that are in fact correct are systematically misinterpreted.

It is also necessary to identify the risk of bias where it may occur, taking into account the root cause, and to analyse this risk and either eliminate or at least mitigate this risk.

Ruling out types of differentiation that are prohibited by law | In the case of certain financial services, the law stipulates that certain characteristics may not be considered for differentiation purposes – i.e. to calculate risk and prices. If conditions are systematically set out on the basis of such characteristics, there is a risk of discrimination. Such a risk also exists if these characteristics are replaced with an approximation.

This would be associated with increased reputational risks and, in some cases, legal risks. In certain circumstances, BaFin, too, might consider it necessary to take measures to address violations of statutory provisions, for instance.

Companies should therefore establish (statistical) verification processes to rule out discrimination.

III. Specific principles for the development phase

Data strategy and data governance | Depending on the application and features of the algorithm, data must be used in sufficient quality and quantity. Companies must have a verifiable process (data strategy) which guarantees the continuous provision of data and defines the data quality and quantity standards to be met. The data strategy must be implemented in a data governance system and responsibilities must be clearly defined.

Compliance with data protection requirements | When using data in algorithm-based decision-making processes, compliance with the applicable data protection requirements must always be ensured. Data protection requirements for the use of data should already be taken into account when planning algorithmic decision-making processes. In particular, disclosure requirements vis-à-vis data subjects must also be observed.

Ensuring accurate, robust and reproducible results | The ultimate objective is to ensure accurate and robust results. It should also be possible to reproduce the results of an algorithm. For example, users should be able to reproduce results in a subsequent test performed by an independent third party.

By ensuring that results can be reproduced, the results can be understood and verified at least to a certain degree by individuals within the company and by external parties. In addition, caution and precision are ensured for the selection of the algorithm as well as the calibration and documentation process.

Documentation to ensure clarity for both internal and external parties | Sufficient documentation is required in order to ensure that algorithms and the underlying models can be verified – by the company itself and by auditors and supervisors. Three steps in particular must be observed here:

Step 1: The selection of the model must be documented. The following steps at least must be taken: firstly, it is essential to determine whether the model is suitable for the specific application in question. Secondly, statistical considerations on the quality of the model should be used. Thirdly, the complexity of the model and its interpretability and auditability should also be taken into account. In particular, an improvement in prediction quality must be weighed up against the potential increase in the model's complexity. Any decisions to opt for a complex model should be explained and the reasons for reaching such a decision must be provided.

Step 2: Model calibration and training must be documented. For instance, relevant calibration details, such as the choice of so-called tuning parameters, must be documented and the reasons behind this must be provided. If models are calibrated using training data, the selection of this data must be documented.

Step 3: In this step, the model validation must be described. Details on model validation can be found in the section relating to the "appropriate validation processes" principle.

In certain circumstances, a clear distinction cannot be made for the documentation of the selection, calibration and validation of the model. This is because, in some instances, the quality of a model can only be determined after the initial calibration and validation of the model. Furthermore, it is only possible to select a model by comparing different models.

Use case: additional information analysis for credit ratings

A credit institution has established a process to determine credit ratings and the likelihood of companies defaulting. This process is supplemented by another process that analyses the annual reports of companies using natural language processing (NLP) and searches for keywords that allow conclusions to be drawn about credit ratings. This new process involves random forest approaches of machine learning, which are randomly generated decision trees with optimised prediction quality thanks to machine learning.

The insights that are gained in this way supplement the established credit rating that is based on quantitative corporate information. Both results are then combined for the final classification of the company.

However, the credit institution has still provided for a check prior to the final classification. If there is a significant discrepancy between the established approach and the credit rating based on the ML approach, a credit analyst must reach the final decision. In doing so, the analyst compares the ML result with an independent expert assessment (specific principle of **"putting the human in the loop"**).

For the purpose of this expert assessment, the analyst needs to gain insight into the classification, i.e. the keywords used by the ML process and their impact on the classification. The process must therefore document the results from the training process in a suitable manner for the expert concerned, in line with the principle of **"documentation to ensure clarity for both internal and external parties"**.

Appropriate validation processes | Every algorithm should go through an appropriate validation process prior to being included in operations. This initial validation should always be performed or at least be examined by an independent function or individual that is not involved in the original modelling process. It is also necessary to determine and document the intervals at which an algorithm must undergo another validation (ongoing validation). In addition to determining regular, appropriate intervals, it is essential to set out factors that will lead to the ad hoc validation of the algorithm and thus potentially lead to the algorithm being recalibrated or an alternative algorithm being selected.

Such factors include, for example, a systematic change in input data, external (macroeconomic) shocks, changes to the legal requirements under which an algorithm is

operated, feedback from the output phase such as a threshold being crossed (for more information on the last point, see the "in-depth approval and feedback processes" principle).

Using relevant data for calibration and validation purposes | A calibration or validation with unsuitable or erroneous data can systematically affect the performance of the algorithm in live operations.[8] For this reason, a particular focus should be placed on the selection of this data and the selection should be carefully documented. The data must be relevant and representative for the application in question. For instance, it must contain information on all relevant sub-groups. Otherwise, imbalanced data in the calibration or validation process can lead to modelling bias. Bias must be prevented as soon as data is prepared, e.g. in the data aggregation phase (see the key principle of "**preventing bias**").

Depending on the scope and riskiness of the decision for which an algorithm is used, various measures should be taken to ensure that the calibration and validation can be subsequently understood and verified. In highly risk-sensitive decision-making processes, the data that is used for this purpose should be saved and stored. In less risk-sensitive decision-making processes, the selection process and the structure of the data should be documented at a minimum (e.g. by means of informative statistics summaries regarding the distribution of relevant data). In any case, it is vital to ensure that the process is understandable for the purpose of internal controls, quality assurance and audits for as long as the algorithm is being used.

[8] In the case of certain algorithms, e.g. unsupervised learning algorithms, data sets are not typically used for the calibration of the algorithm (test data set). In such cases, the validation and selection of the data sets used for the validation are all the more important.

IV. Specific principles for the application phase

The results of the algorithm must be interpreted and included in decision-making processes. This can either be done automatically or by involving experts. A functioning mechanism comprising sufficient checks, feedback loops and modification rules for the development phase must be established in all cases. It is also important to evaluate the extent to which interactions with other algorithms lead to a risk aggregation and whether the use of algorithms is in line with the company's risk management system as a whole.

"Putting the human in the loop" | Employees should be sufficiently involved in the interpretation and use of algorithmic results when reaching decisions. The extent of their involvement should depend on how mission-critical the decision-making process is and the risks this entails (for information on the possible risks and the structure of an adequate risk management system, see Chapter 2: "Key principles"). Their involvement should bring real benefits and should not be limited to the mere approval of every algorithmic decision. One example of effective involvement is the threshold-based process referred to below, which manages the intensity of involvement depending on whether and to what extent algorithmic results differ from the applicable standard. In the case of mission-critical processes, it is useful to define time frames in which a decision can still be reversed and humans can still intervene.

Use case: sanction screening in the context of money laundering detection

In the case of sanction screening in the context of money laundering detection, it is necessary to ensure that transactions do not violate any embargoes. For this purpose, supervised entities generally compare payment data and customer data with sanction and embargo lists. In this use case, two independent compliance staff members compared the data prior to using BDAI. A third employee was involved in critical cases, i.e. if the first two staff members reached different results.

An algorithm now performs the exact same checks previously carried out by one of the two compliance employees; however, there are no changes to the checks carried out by the second and third employees. This means that an independent human assessment is still performed. As a result, the staff resources that become available can be used specifically for the processing of critical cases in which the algorithm and the second (human) assessor reach different results. Although one of the individuals is completely replaced by an algorithm in this case, effective and risk-sensitive human involvement in line with the **"putting the human in the loop"** principle is ensured.

The specific focus on critical cases (including the involvement of an additional staff member) in which the algorithm and the second staff member reach different results can also be understood as an **"in-depth approval and feedback process"** (see corresponding principle).

In-depth approval and feedback processes | When using algorithm-based results in decision-making processes, the situations involving a more in-depth approval process should be clearly defined in advance in a risk-oriented manner. For example, this can be done in the form of a threshold-based process. If the first threshold is crossed, approval should be granted by an individual, also in the case of processes that are otherwise automatic. If the second threshold is crossed, such approvals should only be granted following an additional review of the input data to determine whether there are any peculiarities (e.g. outliers). If the final threshold is crossed, no approval should be granted at first ("stopping rule") but there should be a signal for examining the model (ad hoc validation potentially resulting in a model recalibration or the selection of a completely new model) and the interruption of the process. An approval can be granted only after such an in-depth review has been performed and after changes have been made to the results, if necessary. Such a threshold-based process can reduce the risk of erroneous decisions reached in an algorithmic decision-making process and can improve the quality of results in the long term thanks to an ongoing feedback mechanism.

Establishing contingency measures | Companies should set out measures with which business operations can continue to run if problems arise in algorithm-based decision-making processes. This applies at least to mission-critical applications. For example: a threshold is exceeded in the model described above and the algorithm must first undergo another in-depth validation process before it is used in other areas.

Ongoing validation, overall evaluation and appropriate adjustments | In practice, algorithms must be validated on an ongoing basis in order to assess functionality and check for any discrepancies based on established parameters and to make adjustments if necessary. The validation process is particularly necessary if there are new or unforeseeable internal or external risks that could not be taken into account when the algorithm was created. If new algorithms are used, the interaction and aggregation of risks should be regularly examined as part of an overall evaluation. The principles in Chapter 3 apply to the ongoing validation, overall evaluation and appropriate adjustments. Ideally, an internal or external audit should be performed to examine the regular evaluation and adjustment process. This ensures that the functionality and risks of the algorithms in practice are evaluated independently. Overall, the risks associated with the use of algorithms can be reduced by involving an additional independent internal or external control function.

Use case: fund management

Asset management companies can use algorithms to manage funds. In this case, investment and disinvestment decisions for a fund (collective portfolio management) are no longer reached exclusively by a portfolio manager. Rather, such decisions are taken on the basis of an algorithm developed for this fund. In this use case, shares are selected on the basis of algorithms using a quantitative multi-factor approach aimed at optimising the risk/reward profile of the fund.

A quantitative model is used to analyse and determine the investment universe based on fundamental parameters, analyst opinions and performance over time. The quantitative model is reviewed and improved on a regular basis (specific principle of **"ongoing validation, overall evaluation and appropriate adjustments"**). This means that new metrics are added or existing ones are modified.

Once the investment universe has been determined, a quantitative screening with a multi-strategy model is performed. This model analyses and evaluates securities within the investment universe based on quantitative strategies and takes into account different factors.

Every share is evaluated and rated in line with a quantitative strategy. Within the context of the **"interpretation and use of algorithmic results for reaching decisions"**, the portfolio manager includes the shares with the best model rating in the fund portfolio and reviews the composition of the portfolio on a quarterly basis. To mitigate short-term market effects, the portfolio manager can adjust the factors and/or the strategy weighting (specific principle of **"in-depth approval and feedback processes"**).

V. Inclusion of principles in international regulatory projects

These principles constitute preliminary ideas for minimum supervisory requirements relating to the use of artificial intelligence. As a result, they form the basis for discussions with various stakeholders, including not only market participants and members of academia but also other national and international supervisory authorities and standard-setters. BaFin intends to continue playing an active role in this exchange and to promote the discussion process.

In its Digital Finance Strategy,[9] the European Commission announced its intention to clarify, by 2024 at the latest, together with the European Supervisory Authorities (ESAs), whether and how existing financial market regulation should apply to BDAI applications. Prior to this announcement, various working groups dealing with the risks and opportunities associated with BDAI had already been established at the ESAs – such as the EBA Task Force on IT, the EIOPA InsurTech Task Force and the ESMA Financial Innovation Standing Committee.[10] This paper is intended to make a significant contribution to promoting the joint work carried out by the ESAs and the European Commission – including in the working groups mentioned above. The same applies to the work carried out by global standard-setters, such as the Financial Stability Board (FSB), the Basel Committee on Banking Supervision (BCBS), the International Association of Insurance Supervisors (IAIS) and the International Organization of Securities Commissions (IOSCO).

Work is also being carried out in various areas in the field of BDAI at national and international standard-setters such as the DIN Committee or the ISO Committee. Since a clear definition has not yet been established for artificial intelligence and machine learning, these committees are expected to continue their work in this area. Further technical standards (beyond financial market regulation) for the use of BDAI can also be anticipated. BaFin will be observing these developments, too, and will actively provide support whenever necessary.

[9] European Commission (2020): COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS on a Digital Finance Strategy for the EU, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0591
[10] European Banking Authority, European Insurance and Occupational Pensions Authority, European Securities and Markets Authority.