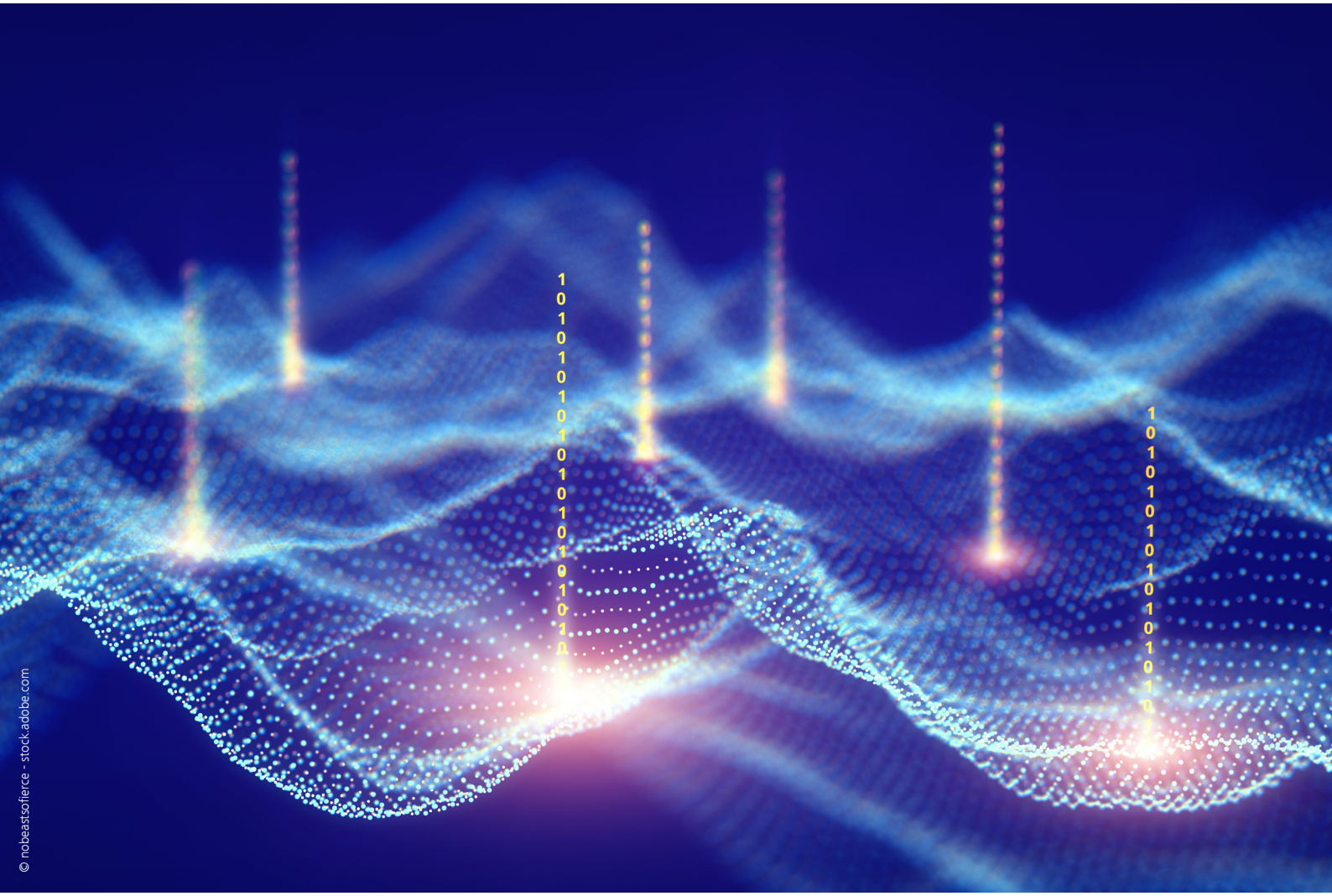
DEUTSCHE BUNDESBANK
EUROSYSTEM

Policy Discussion Paper

The Use of Artificial Intelligence and Machine Learning in the Financial Sector



Directorate General Banking and Financial Supervision

November 2020

1 Introduction

With increasing computational power, exponentially growing amounts of data and continuously improving approaches, artificial intelligence (AI) and machine learning[1] (ML) are achieving considerable success in many fields, in both theory and in practice. Applications of this kind are gaining considerable momentum in the financial sector, too. This development is expected to increase in the near term.

ML changes the modelling paradigm significantly by switching from classical, simple hypothesis-based mathematical methods to a modelling method that is based on learning algorithms[2], which allow for accurate predictions on the basis of even highly non-linear and complex data. These developments raise the question of whether regulators and supervisors need to consider specific risks which may not be sufficiently covered by current frameworks. Many stakeholders have already published initial papers on this topic.[3]

As banks increasingly apply ML to their processes, achieving gains in quality and efficiency, challenges might emerge. Among these are the often-cited black box dilemma, appropriate data quality, model testing and validation standards, and finally the correct implementation into banks' processes. Motivated by these challenges, this paper aims to outline the most relevant issues to be considered when reviewing supervisory expectations for the use of ML. The approach taken here follows the potential new risks. It keeps in mind that balanced and differentiated requirements are needed for legal certainty but also for practicability in order to profit from the potential advantages of ML. In doing so, we are acting as a risk-oriented enabler of ML in response to industry demand for guidance[4] on the use and regulation of ML.

In order to contribute to the discussion – especially regarding the adequate supervision of banks' ML approaches – this paper defines discussion points that also include preliminary considerations for a supervisory strategy that is embedded into a tech-neutral, innovation-enabling and risk-sensitive approach. Where supervisory expectations are formulated, these are summarised in the page margin.

1 We follow the FSB's definition of machine learning: "Machine learning may 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. These techniques can be used to find patterns in large amounts of data (big data analytics) from increasingly diverse and innovative sources.", FSB, 2017, Artificial intelligence and machine learning in financial services.
2 In the following, the paper focuses on ML as the relevant methodology behind AI.
3 See for example: EBA, 2020, Report on big data and advanced analytics. European Parliament, 2020, An EU framework for artificial intelligence. BaFin, 2018, Big Data trifft auf künstliche Intelligenz. De Nederlandsche Bank, 2019, General Principles for the use of Artificial Intelligence in the financial sector. ACPR, 2020, Governance of Artificial Intelligence in Finance.
4 See European Commission, 2020, Consultation on a new Digital Finance Strategy, p. 6.

2 Machine learning – between status quo and new risks

Based on observed and expected use cases from practice, the paper focuses on ML applications[5], bearing in mind that current developments all fall under the definition of "weak AI", which can only tackle a specific problem within a limited scope. We structure the considerations into three areas: First, the supervisory perspective on risks contains considerations on the regulatory framework, on supervisory approaches and on the relevance of ML for prudential supervisors. Second, we argue why a differentiated discussion on the black box characteristics of ML is needed. Third, we discuss considerations concerning the essential aspects when implementing ML at banks.

Supervisory Perspective on Risks

Any supervisory approach on AI/ML must be aligned to the prudential mandate of banking supervision and therefore be risk focused. The following considerations 1 to 6 form the fundament of such an approach.

Consideration 1

Before passing new regulation, supervisors leverage the existing frameworks. Amendments should be made only where necessary.

For internal models (Pillar 1), requirements are available at a general level as well as a more specific level for different risk types (e.g. credit risk, market risk). Since the comprehensive framework[6] is technology-neutral, it can be used to assess ML applications. Building on this framework, competent authorities are experienced in the supervision of internal models as well as related processes.

While the rules of the Basel framework for Pillar 2 are principle based, there are many national regulations in place, spelling these principles out. Often, these jurisdictional frameworks for Pillar 2 already cover relevant topics at least on a general level[7], defining principles for the secure design of IT systems, associated processes and sound risk management. Examples are a well-informed decision-making process, proper documentation and appropriate reporting to the responsible management bodies. The majority of these principles apply to ML.

When facing new model risks or when new use cases of ML arise in banking areas without the need for supervisory approval, supervisors should leverage the existing prudential framework to the maximum extent, while constantly reviewing and revising established requirements, processes and practices. Legislators should amend or expand legal foundations only when necessary. The following considerations elaborate on such characteristics of ML.

Consideration 2

The use of ML should be assessed on a case-by-case basis without prior approval.

First, the exception: For Pillar 1 models, competent authorities are mandated to grant dedicated approvals on an individual basis and changes to approved models are already regulated. The need for approval stems from the fact that banks can deviate from standardised rules and define their own methodologies for calculating regulatory capital.

5 ML applications correspond to ML algorithms together with a use case. Thus, the application is the broader concept.
6 Internal models are regulated under the Capital Requirements Regulation (CRR), the EBA Single Rulebook and the SSM supervisory manuals.
7 The Capital Requirements Directive (CRD) as well as the Minimum Requirements for Risk Management (MaRisk) and Supervisory Requirements for IT in Financial Institutions (BAIT).

Pillar 2 builds upon established principles of risk-orientation and proportionality (see Consideration 1). The need for individual approval in all cases could impede technology-based innovations and would require specific justification. In addition, from a practical point of view, a general need to grant authorisation would create massive administrative burdens.[8]

Consideration 3

The prudential mandate does not include ethical issues surrounding ML.

Since prudential risks are at the centre of the Bundesbank's supervisory mandate, ethical issues only play an indirect role: banks need to consider these risks as a driver of their operational and business model risk, and subsequently treat it within their operational risk management. Beyond that, ethical considerations are relevant for customer protection supervision. Additionally, the use of specific data plays a role for data protection authorities. Potential discussions should differentiate between typical types of concern such as algorithmic discrimination, insufficient overall model quality, non-compliance with data protection regulation or inadequate usage.

Consideration 4

ML is not a regulated activity and banking supervision is no algorithm supervision.

The supervisory focus does not lie on ML itself but rather on the risks resulting from its deployment in the underlying banking processes. Banks are accountable for such models and their model risks. Supervisors are responsible for assessing the way risks are addressed by the bank in accordance with the prudential frameworks, including the application of ML. The question as to the intensity of such assessment and potential approval processes is crucial. Supervisors need to carefully take the risks connected with the impact of ML on the respective outcome or decision into account. Risk type, range of application, level of ML use[9] or decision type[10] are possible criteria to consider. For the assessment of ML, a supervisory "deep dive" into the algorithmic and mathematical setups, might not be required in all cases, however. Proportionality also remains an applicable principle for ML. The higher the risks of the underlying process, the higher the required standards and the more profound the supervisory assessment approach should be (see also Consideration 2).

Consideration 5

Not all "AI" labels actually comprise AI.

"Artificial intelligence" has become a widespread marketing term that implies high levels of predictive power and efficiency. In fact, the label may be misleading. In the absence of a clear or consistent definition of AI[11], supervisors need to understand the features and characteristics of AI in order to assess the associated challenges, issues and limitations. Essentially, a key element of AI solutions in the supervisory context is ML and the aspect of learning, where the machine predominantly performs the training process of a model without pre-defining hypotheses and rules. ML is not about deterministic "if-then" decision rules or hypothesis-based models, even if they reach a certain complexity. Often, AI and big data are mentioned in the same breath. Nevertheless, big data is not an absolute necessity for ML.

Implement a consistent definition of AI and track applications that fall under this definition

Consideration 6

Supervisory expectations regarding ML are independent from banks' sourcing policy.

Outsourcing arrangements are likely to become more important as banks expand their use of ML. Fintechs offer their solutions to many banks, and banking groups are increasingly working in collaboration. As stated in Consideration 1, the expectations regarding outsourcing arrangements are already covered by the regulatory framework. However, expectations regarding ML not only affect banks, but also Fintechs and service providers. If a bank has classified

8 This position corresponds to BaFin's stance on a general approval process for algorithms. BaFin, 2020, Does BaFin have a general approval process for algorithms? No, but there are exceptions. Available at: https://www.bafin.de/SharedDocs/Veroeffentlichungen/EN/Fachartikel/2020/fa_bj_2003_Algorithmen_en.html

9 ML may be used a) as a supporting technique for the development and validation of models, b) as part of a larger model or as c) an entire model itself characterised as ML model.

10 ML supports vs. ML enables automated decision-making.

11 Below, we introduce the concept of an AI/ML scenario, which uses three dimensions to classify the AI setup used.

Manage the risks of ML irrespective of the sourcing policy

the outsourcing arrangement as critical or important within the risk assessment, supervisors may extend inspections to these entities within the outsourcing framework. Ultimately, the risks associated with ML need to be managed appropriately by the bank, irrespective of its sourcing practices.

In this context, an additional aspect to consider is the emerging systemic risk which occurs when market or banking pool solutions are rolled out on a larger scale. This is not only a financial stability issue, but is also relevant from the perspective of an individual institution.

AI/ML scenario

Considerations 4 to 6 focus on the relevance of algorithms from a supervisory point of view. When assessing the relevance of ML applications, we propose that these three dimensions, which we call the AI/ML scenario, are considered:

i. The materiality of the underlying risk of the use case as laid out in Consideration 4. ["What is the ML application used for? What could go wrong?"]
ii. The identification of relevant methodologies against marketing terms presented in Consideration 5, since only real ML applications require a supervisory approach tailored to their challenges ["Is it actually ML? Does it learn/change on its own?"]
iii. ML independent from its sourcing policy (Consideration 6) with supervisors reaching out to Fintechs and service providers ["Who made it and knows how it fundamentally works?"]

Explainability of Machine Learning

Decision-making processes are expected to be based on causality and inherent rationale rules. ML, however, is successful by the exploitation of patterns, hidden in data. It does not necessarily require neither rationales. The resulting lack of explainability – to some extent a feature already in classical statistical approaches – is often seen as a main impediment to the use of ML. It requires a thorough approach to balance chances and risks, when utilising these innovative technologies.

Consideration 7
Black box is not a "no go" if risks remain under control.

Balance accuracy and transparency of models

A lack of explainability is inherent to ML, often making it impossible to develop ML without accepting this black box characteristic to a certain degree. Thus, banks need to weigh the benefits of the ML application against the benefits of simple models with more transparent underpinnings. This problem lies more with the trade-off between the models' high accuracy or power versus their lack of transparency, which is one of their major downsides. Linear models or basic decision rules are easily explainable, but often fail to reflect reality closely enough.

Supervisors should not discuss the black box characteristic of ML in isolation from specific use cases. First, not every use case requires perfect explanation. Second, stakeholders naturally require different types of explanation – developers might focus on data bias, while end-users might need an argument to present to their clients. Third, human-driven decision-making is not free of non-linear decisions or discretion either, but compensates for this lack of explainability by personal responsibility. Fourth, conventional models also show a degree of complexity, resulting in non-obvious results. Even where supervisors accept that ML entails black box characteristics to some degree, they should insist on the paradigm that risk management and decision-making must ultimately be subject to human discretion and human responsibility (see

Human discretion and human responsibility cannot be passed on to models

Consideration 11), as algorithms by definition cannot be held accountable.[12]

Consideration 8

Explainable Artificial Intelligence (XAI) is a promising answer to the black box characteristic, but the approach is not without its downsides.

XAI is the title of an active research field focused on resolving the black box characteristic of ML, with methods like LIME[13] and SHAP[14] representing two popular approaches. There is a fundamental conflict between the implementation of ML, with its potentially high non-linear behaviour, and the demand for comprehensible linear explanations. Explanations put forward by XAI seem to be appealing and convenient, but they only show a limited picture of models' behaviour, from which it is hard to draw general conclusions. Thus, ML combined with an XAI approach cannot make the black box fully transparent, merely less opaque. Nonetheless, it seems to be helpful to use XAI to provide more reliable risk metrics for control processes. A balanced approach should be followed and XAI methods should be tailored to the use case and to the stakeholders' demands.

XAI can help to explain black boxes to a certain extent

Further limitations of XAI methods should not be overlooked. In particular, some methods require high computational power or only deliver minor insights into algorithms' behaviour. XAI methods should support established and used risk control processes and be able to demonstrate effectiveness. If not applying XAI methods, control processes should be in place to compensate for limited transparency.

Compensate for missing transparency with ambitious control processes

Building the model – from data to re-training

Many risks arising from ML can be mitigated already when it is developed. Thus, as important as looking at implementation and output of ML applications, it is to ensure rigid, robust and reliable development and maintenance processes.

Consideration 9

Data quality and pre-processing are decisive factors.

Data quality has always been important for model quality ("garbage in, garbage out"), but becomes a decisive factor since ML is powerful at data exploration during the learning process. A well-trained neural network, for example, will perfectly mirror not only high quality data behaviour, but also unwarranted data relations. This problem is compounded by the fact that the black box characteristics of ML conceal data quality issues. Therefore, banks should set up dedicated data quality processes to ensure that their ML achieves the targeted accuracy. However, data quality is only the first building block of a potentially well-trained algorithm, as elaborated further in the following considerations. Pre-processing, in particular, is a challenging and long lasting step that brings data to the model.

Implement dedicated data quality processes to prevent misbehaviour of ML applications

Consideration 10

ML requires rigorous validation procedures that correspond to the use case.

ML requires a comprehensive validation process that has to be applied at different model maturity phases with an initial, ongoing and ad-hoc process, covering the entire scope and life cycle of the model. Validation of ML is challenging, because a comprehensive set of parameter choices interact with the model's quality.[15]

12 See ACPR, 2020, Governance of Artificial Intelligence in Finance.
13 Local Interpretable Model-agnostic Explanations.
14 SHapley Additive exPlanations.
15 Data preparation and feature engineering (data balancing, generalisation, feature extraction etc.) are at least as relevant as the mathematical model core, comprising model selection and "hyperparameter tuning".

Apply comprehensive validation procedures that correspond to the use case

Standard quality metrics like accuracy, precision, recall or a combination of these (F-measures) need to be tailored to the use case.[16] In the context of (restricted) explainability, the model validation process gains additional attention, since it can at least shed light on model reliability even if it cannot resolve the black box characteristic.

Consideration 11
Data and methodology are important, but supporting processes are even more so.

Responsibility, qualifications, audit safety and documentation are key components of creating a low-risk environment. Since banks cannot hold ML accountable for decisions, algorithmic decision-making must be kept within clear boundaries, and human discretion and judgement are required. Human judgement does not mean that all decisions have to be supervised by humans, as they are not necessarily able to understand the decision itself. Instead, risk-oriented samples, frequent oversight by developers and well-informed decision analysis should ensure appropriate results of ML.

Use cases determine banks' acceptance of errors. Ultimately, humans take the responsibility for algorithmic decisions. This must not be a formalism, but it requires close monitoring by algorithm developers and financial risk experts. The more complex and less transparent the workings of ML, the more important control processes become. Banks should be able to identify misbehaviour by their ML applications and to control associated risks.

Embed ML applications into control processes

Consideration 12
Learning frequencies are to be justified. Re-training can change everything overnight.

ML algorithms can be adaptive or dynamic, i.e. a re-parametrisation is planned when new data becomes available. This may even happen autonomously during live operation. This feature is a game-changer, since it allows the model to adapt swiftly to new relations in the data.

Against this background, for models subject to supervisory approval, these adaptations can change the behaviour of the model significantly and represent material changes that result in supervisory actions. Irrespective of the question of whether such adaptive changes might require supervisory approval in the special context of Pillar 1 models, the choice for the frequency of training cycles should be justified by banks. Banks should be able to provide evidence of the advantages of their chosen approach and established processes which enable them to identify, measure and control the risks.

Re-training-cycles need justification and control

16 For the example of credit decisions, it is obvious that banks try to reduce credit to customers that will default in the future (precision). Similarly, for the example of early warning systems that identify customers with default risk, it is less important to reduce the number of warnings for customers that do not default. Instead, banks focus on detecting a large number of the defaulting customers (recall).

3 Conclusion

This paper outlines considerations including potential supervisory expectations for ML by the Bundesbank's Directorate General Banking and Financial Supervision, with a focus on the financial sector. Successful ML applications represent an important building block of digitalisation – they are able to improve analysis depth, reaction times, operating quality and cost efficiency. However, banks must continue to maintain a sound risk management environment, including processes to identify and control relevant and material risks.

The main supervisory focus should be on features of ML which are novel to current regulation and supervisory practices. The black box characteristic, potential data quality issues or challenges within the model learning process are among the key issues. Even when ML depends heavily on data and learning algorithms, it seems that the supporting processes become more important in banks' control environment. Data preparation, model validation, monitoring and escalation procedures become more relevant to maintaining the ability to control model quality.

Several national competent authorities have already published principles and opinions on artificial intelligence, machine learning and big data that have the potential to threaten banks' level playing field. Regulators and supervisors must not implement different standards for a topic that requires maximum harmonisation within the single market and between jurisdictions.

The next step will be to foster dialogue between users, researchers and authorities to develop a consensus on the key risks and related supervisory expectations. We support the European Commission's plan to put forward supervisory expectations on the use of AI applications in financial services as stated in the recent digital finance package.[17] This approach needs to align with activities of the Basel Committee on Banking Supervision's (BCBS) Supervision and Implementation Group (SIG), which is currently working on corresponding supervisory expectations at the international level.[18]

17 European Commission, 2020, Digital Finance Package. Available at: https://ec.europa.eu/info/publications/200924-digital-finance-proposals_en.
18 BSBS 2019, High-level summary: BCBS SIG industry workshop on the governance and oversight of artificial intelligence and machine learning in financial services. Available at: https://www.bis.org/bcbs/events/191003_sig_tokyo.htm