FINANSTILSYNET

NOTAT

The Danish Financial Supervisory Authority

3 September 2024

Recommendations when using artificial intelligence

The significant breakthroughs in generative artificial intelligence (AI) in autumn 2022 – in particular the ways ChatGPT and similar services can potentially be used – have made AI a higher-priority issue in society. This is also true within financial companies where the technology can be useful for gaining market share or optimise operations. These services make it possible for everyone – even users without advanced technical abilities – to use AI in their everyday life. If only – or perhaps precisely – for that reason, it may be necessary to take active steps to avoid letting a fear of falling behind overshadow the focus on good governance and risk management, as well as on a healthy corporate culture and ethical considerations that are a prerequisite for safe use.

In this light, the Danish Financial Supervisory Authority (Danish FSA) asses that it is advantageous to further specify the recommendations from Recommendations when using *machine learning* in the financial sector, which was published in 2019[1]. Recommendations in this paper should therefore be read in connection to the 2019-paper.

The paper from 2019 contains the Danish Financial Supervisory Authority's recommendations in nine specific areas that companies may consider as the use of machine learning – or more broadly AI – increases. As AI moves from the drawing board to the production environment, the Danish FSA sees a need to once again point out that the approach to using the technology should focus on risk management, regardless of its great potential.

The paper from 2019 was based on a test case in the regulatory sandbox, FT Lab, of supervised machine learning via a deep neural network. For that reason, the Danish FSA chose to publish the original good practice paper for the use of supervised machine learning, even though the topics and

[1]https://www.dfsa.dk/financial-themes/fintech/guidance-and-handheld-supervision/recommendations-when-using-supervised-machine-learning

recommendations contained in the 2019 paper – in the authority's view – describe good practice for the use of AI in general.

The Danish FSA's recommendations paper from 2019 listed a number of suggestions for the use of machine learning with a focus on the following areas:

- Purpose for using supervised machine learning and description of the model
- Governance (model development, application and updating), policies and business procedures
- Data processing
- Training the model
- Performance and robustness
- Accountability
- Explainability
- Data ethics, bias and fairness
- Transparency.

This paper is based on a series of meetings with AI experts. The meetings focused on AI in a broader and more general sense than machine learning alone. For this reason, this is also the focus of the paper.

In this paper, the section on *governance* is divided into two subsections with a focus on *the organisation* and *the model* respectively. The section on *explainability* will also be revisited.

The paper was prepared in parallel with the drafting of the EU Artificial Intelligence Act (AI Act). However, the objectives of the AI Act and this paper are different. The AI Act covers the use of technology broadly in society and introduces concrete regulation aimed at companies that use or sell products based on AI. The regulation focuses on protecting the fundamental rights of EU citizens in the face of AI, while this paper focuses on risks for financial companies using AI.

The good practice paper is not regulatory in nature and will not in itself form the basis for supervisory reactions. As part of the Danish FSA's 2025 strategy, which is intended to, inter alia, support *the reassuring use of technology as well as new business models,* the paper aims to make the financial companies aware of areas where the use of AI may lead to an increased need for risk mitigation.

The paper presuppose that financial companies have reassuring IT development. The paper therefore focuses on increasing companies' awareness that new tools can create new risks and that these new risks will likely need to be handled in a new and different manner. Therefore, the individual company must consider whether established processes for risk management, including in connection with IT development, need to be updated.

The companies should make several ethical considerations when using AI. These are not covered in this paper, but the Danish FSA's considerations on ethics can be found in its report on data ethics when using AI in the financial sector, which was published on 13 November 2023[2].

This paper is aimed at all regulated financial service providers. Several of these companies are regulated by the *Executive Order on management and control of Banks, etc*. or *Executive Order on the management and control of insurance companies etc.*, which means that some of the paper's suggestions for good practice for these companies will indeed be legal requirements.

The Danish FSA has held meetings with several representatives from credit institutions, investment companies, marketplaces, insurance- and pension companies as well as payment institutions, in addition to consulting with experts and academics. All with in-depth knowledge of the use of AI in and outside the financial sector. The focus of the meetings was the use of AI, which risks may arise, and which measures the companies have implemented to manage and mitigate these risks. Everyone who contributed to the work have had an opportunity to provide comments before the publication of the paper.

[2] https://www.finanstilsynet.dk/Tal-og-Fakta/Rapporter/2023/AI-i-den-finansielle-sektor_161123

1. Governance

A company that uses or plans to use AI to carry out its activities should ensure that the organisation has the best possible conditions to implement the technology in a reassuring manner. The organisation's structure should ensure that advantages are leveraged while also managing relevant risks, including aspects of IT security. The individual company's organisational setup depends on the company's business area and on the concrete use of the technology. In other words, there is no single universal approach that is appropriate for everyone.

The use of AI varies across companies – e.g. in relation to whether the models are used directly in the business or for supporting activities, such as quality control or simple sorting of email.

Some financial companies are using AI models to challenge established and regulatory-approved models that are based on more classical static methods. When the difference in performance between the approved model and the so-called challenger model becomes sufficiently large, it may indicate that the established model needs to be updated. In such cases, AI does not necessarily directly affect the established processes, which primarily deal with the static authority-approved models, but the technology supports an ongoing test.

In other contexts, AI is used directly in established business processes, such as models for monitoring transactions or trades, and for automatic distribution of emails from a central mailbox.

Requirements and expectations for the company's risk management will vary, depending on the degree of impact the model has on the company itself or its customers and whether AI is used, for example, directly or as a challenger model. The greater the influence of AI, the more the Danish FSA expects the company to have measures to identify and manage risks.

System for the company's overview

Individual companies have organised themselves very differently to be able to effectively identify and manage risks from the use of AI. Several companies have a specific focus on AI in relevant policies, while others have drawn up new policies for the specific technology. Some companies have set up management bodies with decision rights on development and use of AI when the effect for their customers or for the company becomes sufficiently large. Some have also ensured that they have an overview of AI use in their models – whether purchased or developed internally. Such an overview can, for example, be established by including the company's use of AI-based models in a register of the company's overall use of models, which all companies should have. The specific approach depends on what is most appropriate for the individual company, but in general a register of the company's models can contribute to the company's overview and form part of the organisation's general handling of its overall model risk.

The position on particular risks of AI-based models

The participating companies have implemented various risk-mitigation measures. Some companies centralise control with particularly competent employees, such as domain experts who assess various aspects. This may include, for example, the appropriateness of a model in relation to its area of application or the impact and risks associated with a model seen in relation to internally set limits. Other companies have internal policies for risk management of models, whereas the assessment of the risk itself is decentralised. In general, a company should decide how it ensures the handling of particular risks associated with AI-based models and what internal resources it has for the task. This should happen in connection with the individual company determining its approach to risk management.

Established approach to risk analysis

Several companies use “tiers”, which categorise models based on a consideration of risk, for the purpose of assessing the resources they should allocate to risk-reducing measures for the individual model. Tiering is a method where models are rated as a tier based on the potential risks of the model. The tier of a model determines where the model fits into the company's governance. In other words, the placement in a specific tier indicates how many resources the company must allocate to ensure that the use of the model does not exceed the company's risk tolerance. Tiering is just one example of a method of categorising models. The actual method chosen by the individual company is not critical. The essential thing is that the company even has a procedure to assess how risky a model is and its scope of application, so that the company can determine appropriate risk mitigation measures. The company can use the same procedure for AI-based models that it uses for its other models. In any case, an established approach to risk analysis and classification of internally and externally developed models is fundamental to being able to identify and manage risks, including risks associated with AI. It may be relevant to include topics such as the complexity and importance of the model, as well as how dependent the company is on the output of a specific model in the performance of a function that the model supports or performs.

Identification and demarcation of risks

As the use of AI increases, the Danish FSA expects the individual company to continuously consider whether its approach to identifying and demarcating risks is sufficiently covered by existing policies, whether these need to be adjusted or whether it is necessary to prepare a separate policy for the use of AI. This also includes that the company should consider whether the use of

AI entail risks to a degree that the company should report to the management and the board of directors on an ongoing basis.

Anchoring responsibility in specific employees or units

The most important thing is that the company be able to identify and deal with risks associated with using the technology in the specific company. In this connection, the company should focus on anchoring the responsibility for specific models – including control – with specific employees or units. Depending on materiality, the responsibility should also lie with someone other than the model developer. Regardless of how the company uses or intends to use AI, it is necessary to consider measures that sufficiently ensure that the technology is implemented and used in a reassuring manner.

Good practice for the use of artificial intelligence means that the company

- has a system to create an overview of its AI use, possibly using the register for model use in general
- determines how it ensures the handling of particular risks when it establishes its approach for risk management and which internal resources are available for the task
- has established an approach to risk analysis and classification of the use of AI for both internally and externally developed models
- continuously considers whether its approach to identifying and demarcating risks is sufficiently covered by existing policies
- ensures that the responsibility for specific models, including control, lies with specific employees or units in the organisation that are not part – or have not been part – of the model development, where this is relevant in relation to the materiality of the model.

2. Model management

Based on its overall risk assessment of the use of AI, the company can address how it handles risks associated with specific models. The company's existing measures for handling model risks may be a good starting point.

Awareness of the difference between AI-based and other models

Several companies have pointed out that AI-based models do not necessarily differ significantly from more classic models. A company looking to use AI can therefore benefit from identifying where in a model's life cycle the company's traditional method for handling model risks could be challenged when using AI. The individual company should therefore consider whether and how its AI-based models differ from other models and how this affects the company's management of such. This type of risk analysis is the prerequisite for the company to be able to handle potential challenges. It is not necessarily clear in advance where a new model will differ. The use of AI will possibly give rise to updating precautions in connection with, for example, development, testing, validation, commissioning, retraining, phasing out and the like. Many companies have noticed that AI-based models, among other things, differ from more classical models in that more frequent retraining is often relevant.

The frequency of retraining can be determined in several ways depending on the individual model and company, and each of them may be appropriate. For example, several companies have scheduled fixed intervals. Others have established thresholds for how far a model's estimates may deviate from realised values before the model needs to be retrained. Yet other companies have chosen to continuously assess whether the issue the model is designed to address has changed significantly. This may, for example, be major events such as changing interest rates, geopolitical tensions, recession or a pandemic. It can also be minor events such as changes in data input or data sources. Some companies use several parallel approaches.

Plan for retraining models

When processes are automated with, for example, the use of AI, it can lead to changed behaviour in those who use the system or those the system is applied to if they are aware of it. This may be the people administering the system, or the users or customers on whom the model is based. A change in behaviour can lead to the model's dataset no longer reflecting the reality in which the system operates. Such a risk should be mitigated – e.g. through more frequent retraining. The company should in this regard document that it has decided on how a specific model is expected to be retrained to keep it up to date. This is particularly relevant if the model has been developed by a third party. Based on specific model risks, the company should have a plan for how and how often its models need to be retrained. This applies to internally developed as well as externally procured models. Finally, the company should ensure validation on a regular basis.

It is ultimately up to the individual company to assess which parameters it will use as a basis for determining the frequency and method of retraining. The process leading up to a specific plan for retraining a model, which may involve collecting relevant data for the decision, provides the company with a good understanding of whether the purpose of the model is accurately described and a good overview of the model's ongoing development.

Sufficient internal resources

Retraining a complex model can potentially change the model to such an extent that it can no longer be considered a minor update of a previous version. The result may actually be a completely new model. Therefore, it may make sense already at an early stage in the model's life cycle to think about how it should be trained – not to mention retrained – and which resources the company should have available at such times. It may, for example, be that domain knowledge in connection with retraining is important for concrete model specifications. In the same vein, it would be advantageous to also consider model validation. If the company has only limited resources for model validation, the company may want to prioritise AI types that are less resource intensive to validate. The company should consider whether it has sufficient internal resources, including for model validation.

Validation of AI-based models should basically follow the same approach as the company otherwise follows. For example, companies with a separate validation unit can place the responsibility there. In practice, this means that the company should ensure that validation takes place independently of development, to the extent that the company assesses that the model is sufficiently material. For this reason, it should not be the same person or group who both develops and subsequently validates a given model if it is deemed sufficiently material.

To ensure sufficient validation of the use of material models, it will generally be necessary to have employees with the required technical and business prerequisites to be responsible for an actual critical control. A critical control of the company's models is essential to mitigate risks, but in light of the fact that many have challenges in recruiting employees with sufficient skills for the actual development of the models, it can potentially also prove difficult to recruit employees with similar skills for model validation. This may indicate that the company should avoid particularly complex models. The company should also allocate independent resources to the specific task of model validation. Otherwise, it would be natural to assign the task to the model developer, who would then have to check their own work, which is not appropriate, especially for models with greater risks.

Some companies have highlighted that complexity can increase further if results from an AI-based model provide input for other models. The problem is further reinforced if complex and less explainable models are included. For example, it can quickly become difficult to predict and understand what an otherwise simple update or retraining of an underlying model might mean for models further down the chain. Therefore, companies should have methods to keep track of such relationships before the output from one model is used as input for others.

Policy for managing and securing model versions

If it becomes necessary to recreate a previous version of a specific complex model, accessing documentation may be necessary – not just for the model's parameters, but also for how these parameters were chosen, the dataset and the process from data input to model output. Adequate documentation will increase the possibility that a company can recreate previous versions of a model. This means that it is important for the individual company to have a policy for how it manages and ensures the versioning of its models.

Several companies have highlighted how essential sufficient data quality and data management are in general to ensure appropriate use of AI as well as for risk management. This implies, among other things, that the company has control over the data it uses, where the data comes from, how the data can be modified and whether the data used can lead to inappropriate bias.

The company should consider whether data contains sensitive information that is not necessary. If a company uses a model based on data that the company does not fully understand or know where comes from, the model may end up operationalising and automating unwanted biases in data that the company did not intend to support. It can thus make sense for the company to focus on the collection and processing of data, especially if the model in question is to be used in regulated areas. If the company obtains data from third parties, it may be relevant to ensure that data does not change character, as this may change the model and its output noticeably compared to what was expected.

Effective management of data

The access to very large amounts of data has pushed the increasingly widespread use of AI. The development has not only increased the use of external data sources but can also encourage the collection of internal data in one place. It may make sense from a purely business point of view, but it can also lead to obvious risks: The more data that is collected in one place, the greater will be the consequences of a possible data leak. In addition, the company should consider whether the employees' access to, for example, sensitive personal data that can be included in a combined data set is justified. Based on the existing requirements for data management, the individual company should decide how it manages its own data effectively, including how the company stores and provides access to data. This is a prerequisite for the company to efficiently and safely collect, store and use large amounts of internal data for use in AI.

Finally, companies should be aware that automating a process does not necessarily mean that competent employees in the field become redundant. A domain expert who has previously carried out a task that a company now wants to automate often has deep insight into the problem behind the task that the model is merely able to solve. If for some reason a model is no longer able to perform, the domain expert will often be able to contribute to finding the cause and participate in a retraining or subsequent validation of the model.

Good practice for the use of artificial intelligence means that the company

- from a risk perspective, has dealt with whether and how its AI-based models differ from the company's other models
- based on the materiality of the model, has a plan for retraining and regular validation of models, including externally procured ones
- based on the materiality of the model, has dealt with the adequacy of internal resources, including model validation
- has a policy for how it manages and secures different versions of its models
- focuses on effective data management, including storage and access to data.

3. Explainability

Several companies have pointed out that more complex models often perform better than simpler and more explainable models. For many companies explainability is however a key criterion for choosing model specifications. It can therefore often be necessary to make a trade-off between performance and explainability when a company is developing and commissioning a model.

In this context, "explainability" is not a precisely defined term. It covers the extent to which it is possible to explain an outcome of a model to a stakeholder – i.e. how and why a model has reached a concrete output. With this approach, the expectation of the degree and form of explainability depends entirely on the recipient, including whether the recipient is internal or external, and on how intrusive the decisions that are made based on the model are.

To effectively use a model, a company must both have the skills to develop the model so that it is usable and effective, and an understanding of how to subsequently utilise it for the intended purpose. Sufficient internally directed explainability will support the user of the model to truly understand it, including the strengths and weaknesses associated with it.

Trade-off between performance and explainability

The companies the Danish FSA spoke with generally agree that explainability is an important issue in connection with the use of AI. For some companies, it is the most important thing. For others, the most important issue is the model's ability to perform and deliver results, after which the issue of explainability must be adequately addressed and resolved. The Danish FSA recognises that there may be examples where this trade-off is not necessary if the specific model ensures both high performance and explainability.

Financial institutions that do not have direct customer relationships will face a lower level of expectation for the externally directed explainability of their models. Internal requirements for explainability can still be high, even if the direct customer relationship does not demand a high level of explainability before the customer. This may, for example, be investment companies that operate a fund that must match the return of an index. Conversely, companies whose actions directly affect consumers will naturally face a higher degree of expectation for externally directed explainability. This may, for example, be insurance companies, which must decide whether a customer can be allowed to take out specific insurance or is entitled to have an insurance sum paid out. It can also be companies that approve or reject loan applications from consumers. In these cases, the expectation of explainability before external parties may involve consumers being told how a conclusion has been reached and where the explanation does not simply consist of technical specifications.

In cases where financial companies have to make this trade-off, the decisive factor may be how far-reaching the results that a model delivers are – e.g. what effect so-called false positives would have on customers or the company. An example could be a payment service provider using an AI model for transaction monitoring, which is able to automatically stop a transaction when the probability of a transaction being an attempted fraud crosses a certain threshold. If the model's assessment turns out to be wrong, the customer can contact their payment service provider and have the transaction completed or their card reopened. In this example, the negative impact is so low that external explainability should not hinder implementation of the most effective model. In this example, there may still be high requirements for internal explainability. Conversely, it will be crucial, for example, for an insurance company assessing whether a customer is entitled to compensation, to explain the basis for the decision – especially in the case of a rejection.

Competences for balancing the relationship between performance and explainability

It is crucial that the individual company deals with the question of explainability. The company should be able to document its balancing of the relationship between explainability and performance for its concrete models, when deemed relevant.

It is a common theme for the companies that the Danish FSA has asked whether the requirements for the use of AI must necessarily be greater than the requirements for the processes where humans solve the same tasks. The risk associated with scaling automated – and to varying degrees autonomous – processes can be of a different nature than the risk associated with individual experts solving the tasks. Therefore, different processes and controls may be needed than those that apply to the individual experts. In this connection, explainability is important. The expert who carried out the task was also able to explain a specific outcome, but when mathematics and statistics replace or supplement a function, the expectation of the explanation will also be different.

Balancing performance and explainability can be complex. The task requires an understanding of the overall usage scenario, including the business purpose, users, technology and any legal requirements. Therefore, the company should decide how it ensures that the model works in the intended manner and how the company will define and measure this. A company that wants to use AI to solve a specific challenge should, for example, consider whether the model's definition of performance is appropriate.

Several companies emphasise that if the responsibility for making the trade-off is not placed, it can invariably fall to the developer of the model, who does not necessarily have sufficient insight into the subject area and its pitfalls. At the same time, the developer may find it difficult to define what explainability means for other customer groups or professional groups than the one to which the developer belongs. A company can therefore advantageously involve a cross-functional team of professionals in balancing various considerations. This may, for example, be experts with domain knowledge and technical skills as well as competences within law, compliance and risk management. The company should assess the competences of the employees who make the balance between performance and explainability.

Understanding the model's results or bias

Different usage scenarios bring different risks and need for explainability depending on how complex a model is. The more directly intervening or otherwise significant a usage scenario is for the company or the affected customers, the greater the need for external explainability can be expected. Regarding customer-facing areas, it is important to remember that it is the customer, and not a model developer or domain expert, who must be able to understand the model's output. It is often essential for the company itself to be able to understand and explain the model's results or bias. This applies particularly in cases where the model's scope has a significant impact on either the company or its customers, or where the model's results contradict economic theory. The latter might be illustrated by a model's output indicating that the smaller of two otherwise nearly identical apartments is worth more or that positive shocks to GDP worsen the share price of an otherwise cyclical stock.

Assessment of relevant usage scenarios

A company should assess each individual usage scenario as part of the model development to balance the relationship between performance and the need for explainability, where relevant. At the same time, this can increase the possibility that the company identifies and handles risks at as early a stage as possible. It can also contribute to the company improving its monitoring of the models.

Documentation of choices and considerations about mitigation measures

The trade-off between explainability and performance for a given model can result in a company choosing a model with better performance at the expense of explainability. This will potentially increase the need for risk management, including monitoring and control. When using complex models that are difficult to explain, the company should generally be able to justify its choices – e.g. by documenting its considerations in this regard. The company should also consider mitigation measures and document these considerations. In this way, the company can substantiate its assessment that the model performs as intended, even though the company cannot necessarily fully explain why.

The company should therefore consider whether increased complexity is necessary and whether the model's performance is improved by measures that justify the fact that the model's outcome becomes more difficult to understand and explain. If a simpler model performs just as well, the company should document why it chose the more complex specification.

A number of widely used models, such as LIME and SHAP, can help explain what underlies an AI model's results. However, it is worth keeping in mind that the models that support explainability are in themselves models built on assumptions and methodological choices. It is therefore relevant to consider whether these models should undergo the same procedure as the company's other models. If the company has difficulty explaining how the models it uses to explain the result of an AI model work, it may be a sign that the company is not using the most appropriate explanatory model in the situation in question. It can also be a more fundamental sign that the company should choose AI models that are basically easier to explain.

Good practice for the use of artificial intelligence means that the company

- based on the materiality of the model, documents its considerations about explainability for specific models, including the relationship between explainability and performance for specific models
- based on the materiality of the model, has assessed which competencies should be included in the trade-off between performance and explainability
- can understand and explain which inappropriate results or biases the model may be associated with
- assesses usage scenarios as part of the model development to clarify requirements and needs for the model's explainability
- when using complex models that are difficult to explain and based on the materiality of the model, can justify and document its model choices and considers and documents mitigating measures.