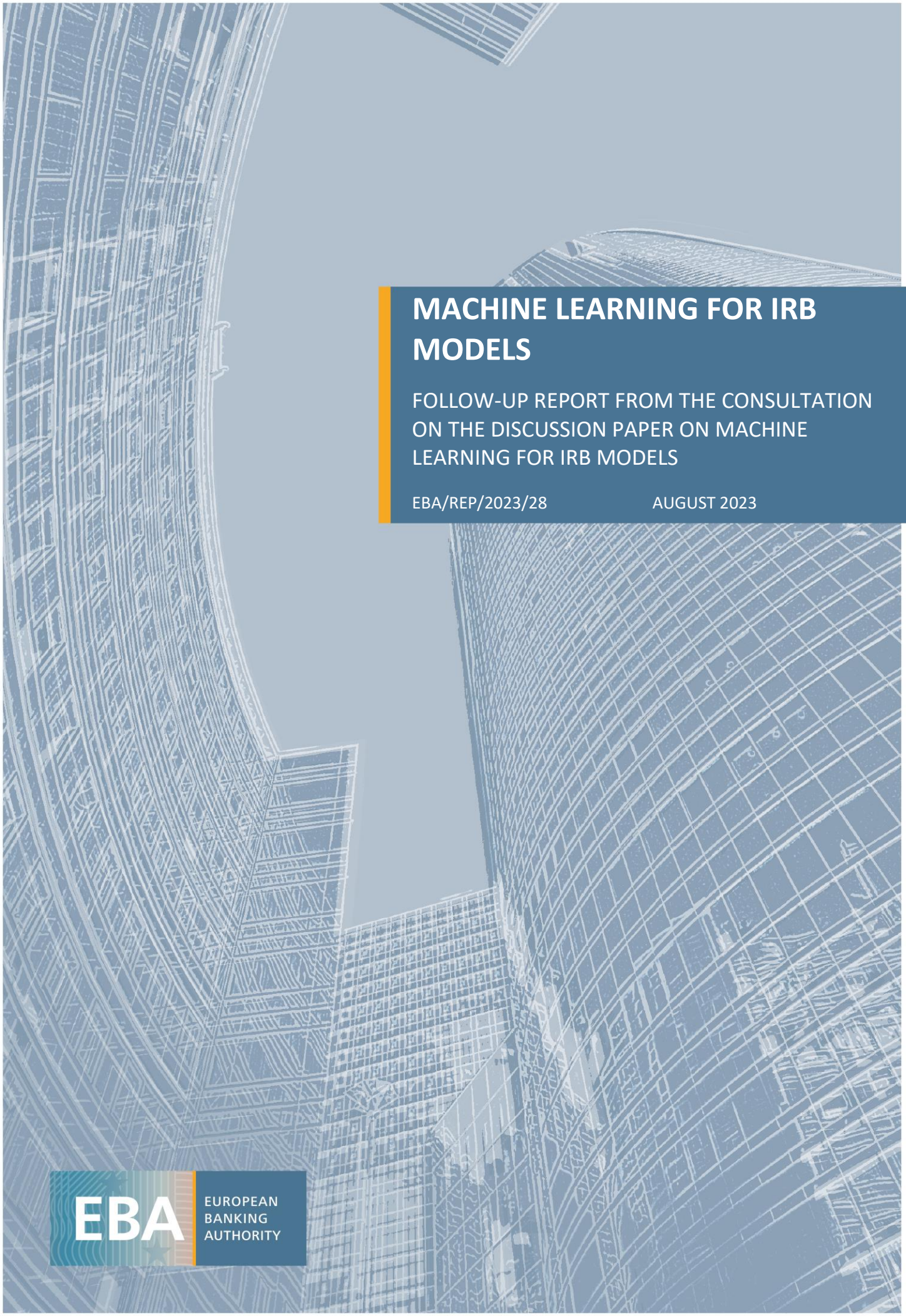


MACHINE LEARNING FOR IRB MODELS

FOLLOW-UP REPORT FROM THE CONSULTATION ON THE DISCUSSION PAPER ON MACHINE LEARNING FOR IRB MODELS

EBA/REP/2023/28 AUGUST 2023

EBA EUROPEAN BANKING AUTHORITY

Contents

1. Executive Summary 3

2. Introduction 4

3. Selective use of ML for IRB models 6

3.1 ML techniques in the core modelling steps 6

3.2 ML techniques used for the validation via model challengers 7

3.3 ML techniques for collateral valuation 7

4. Complexity of ML techniques 8

4.1 Statistical Issues (Overfitting) 9

4.2 Human Skills 9

4.3 Explainability 10

4.4 Categorisation of model changes 12

5. Interaction with (non-CRR) regulatory frameworks 13

5.1 Interaction with GDPR 13

5.2 Interaction with AI Act 15

5.2.1 Clarification on the scope of application of the AI Act 15

5.2.2 Additional requirements from the AI Act 16

6. Way forward 19

Annex I– Principle-based recommendations 20

Recommendations regarding the regulatory compliant use of ML techniques in the context of the IRB framework 20

1. Executive Summary

The aim of this follow-up report is to summarise the main conclusions from the consultation on the discussion paper (DP) on Machine learning (ML) used in the context of internal ratings-based (IRB) models. In addition, this report discusses the interaction between prudential requirements on IRB models and two other legal frameworks that have an impact on internal credit risk models that use ML techniques, namely the General Data Protection Regulation (GDPR) and the Artificial Intelligence (AI) Act.

The exponential increase in data availability and storing capacity coupled with the improvements in computing power of recent years provide an opportunity to use ML models. As presented in the DP, the follow-up report focuses on the more complex models than traditional techniques, such as regression analysis or simple decision trees, which are also the most difficult to understand and often less 'transparent'.

In the context of credit risk, ML models might improve the predictive power and are not new to internal models used for credit approval processes. The result of the consultation suggests that financial institutions use or intend to use ML techniques only in some areas of IRB modelling, primarily for the PD estimation during the risk differentiation phase of the model development. Financial institutions use ML techniques to a lesser extent for model validation and collateral valuation.

The DP discussed as the main pivotal challenge for using ML techniques the complexity, which leads, to challenges in explaining their results. This goes hand in hand with the problem of traceability and ensuring an adequate understanding by the management functions. The result of the consultation suggests that in addition to those points, financial institutions have difficulty addressing the statistical issue of overfitting and acquiring a sufficient level of skilled labour. Furthermore, the industry mentioned the lack of clarity regarding how the prudential requirements on material model changes should be applied when using ML technique. Therefore, the report provides some clarifications on this point, noting the importance of the alignment with the regulatory framework.

Finally, the possible use of ML techniques in credit risk models may create concerns outside of the scope of prudential considerations. Two of the legal frameworks that addresses those concerns and that interact with the use of ML techniques for IRB models are the GDPR and the AI Act. This follow-up report discusses the interaction between prudential requirements on IRB models and requirements within those two regulations that may affect IRB models. In particular, the EBA fully supports the objective of the AI act, but is of the view that the text would benefit from further clarifications to ensure that no legal uncertainties remain after its adoption.

Altogether, this follow-up report aims to provide a conclusive picture on the current use of ML techniques for IRB models and analyses possible obstacles to the implementation of ML models in the IRB model space based on some practical issues, the experience made by the industry thus far and the current state of regulatory developments in this area.

2. Introduction

1. The availability of financial markets data has undergone an exponential increase. At the same time the capacity to analyse data has increased substantially due to better computing capabilities and a growing use of Big Data and ML techniques.[1] The use of ML techniques has also continuously increased in the banking sector over the years, with the main areas of application being fraud detection, AML/CFT, client and transaction profiling/clustering and real-time monitoring of payments.[2] In the area of credit risk, EU financial institutions using ML techniques mainly use them in their credit worthiness assessment or credit scoring for credit granting.

2. ML techniques are however less used in the area of credit risk as regards determination of capital requirements. This can be partly ascribed to existing doubts about whether the ML techniques can comply with the regulatory requirements on the use of IRB models. This is in particular relevant, as the majority of financial institutions today use variants of regression analysis when using internal models to determine credit risk capital requirements, whereas the use of decision trees/ random forest or other ML techniques is limited to fewer financial institutions.

3. On 11 November 2021, the EBA published and consulted on the DP on '*Machine learning for IRB models*' with the aim to engage the industry and supervisory community to investigate the possible use of ML for IRB models and to build up a common understanding of the general aspects of ML for IRB models and the related challenges in complying with the regulatory requirements. The DP gave an overview of the challenges and opportunities stemming from the use of more advanced ML techniques, when applied in the context of IRB models to calculate regulatory capital for credit risk. It furthermore identified which characteristics of ML might make it more challenging or easier to comply with the regulatory requirements for IRB models.

4. The DP furthermore included a set of principle-based recommendations,[3] to which ML techniques should adhere if banks wish to use them in the context of IRB models. This should ultimately ensure: i) a consistent and clear understanding of the prudential requirements, ii) an understanding of how new sophisticated ML models can adhere to those requirements and iii) that the setting of capital requirements continues to be harmonised across Europe.

5. ML covers a wide range of models with different levels of complexity. Just as in the DP, this report focuses on the more complex ML models,[4] which are the most difficult to understand and

[1] European Supervisory Authorities' report on Digital Finance and related issues ESAs (January 2022), ESAs Final report on Digital Finance.

[2] EBA Risk Assessment Questionnaire, Autumn 2022.

[3] Section 4.3 of the DP on ML for IRB models and Annex I of this report.

[4] The EBA has in previous communication (i.e. the report on BD&AA) used the definition of ML presented in the standard on IT governance ISO/IEC 38505-1:20175 which defines ML as a 'process using algorithms rather than procedural coding that enables learning from existing data in order to predict future outcomes'. More details on the definition of ML are provided in section 2.1 of the DP on ML for IRB models. Please further note that this definition of ML may be different to the once used in other regulatory texts, e.g. the AI Act.

challenging to use within the IRB approach. For this report, the term ML should refer to *models characterised by a high number of parameters, that require a large volume of (potentially unstructured) data for their estimation and that are able to reflect non-linear relations between the variables*. This entails that the use of regression models, which is the most commonly used approach in IRB models today, is not the focus of this report.

6. The EBA sought advice from the industry on 17 questions in the DP. During the three months consultation period, the EBA received 14 responses, of which three were confidential and eight were provided by associations.[5] The responses generally reflected, that the implementation of ML techniques by financial institutions for prudential purposes related to credit risk is still at an early stage. The responses received from the industry therefore only reflect the short experience with ML techniques thus far or expectations about future use. Conclusions drawn from the consultation should therefore be taken with caution, as this is also a developing field within the industry.

7. While the EBA continues to monitor the developments of ML techniques for IRB models in the market and the new rules developed by legislators, several conclusions on the current application of such techniques can be drawn from the specific findings of the DP on ML for IRB models and the feedback received from the industry.

8. The industry feedback shows that financial institutions generally agree with the recommendations in the DP, and some elements, such as the application of the model changes framework in the context of ML models are further developed in this report based on the respondents' feedback. Overall, the principle-based recommendations set out by EBA in the DP, and included in the Annex for reference, can therefore be used as a basis for development of ML models.[6]

9. In terms of organisation, section 3 elaborates on how financial institutions are currently selectively using ML for IRB modelling and section 4 elaborates on the additional complexity when using ML techniques for IRB models. Section 5 evaluates the interaction of ML techniques potentially used in IRB models with the GDPR and the AI Act. The report concludes with the way forward that the EBA expects to take in relation to the use of ML for IRB models (section 6).

[5] The 14 responses (link) received came from six individual financial institutions and advisory services from DE, FR, IT and NL and from eight associations, representing financial and other institutions EU-wide or specifically in DE, ES, FR and IT. The associations generally represent a broad range of different types of financial institutions, such as wholesale, cooperative, mutual, retail, commercial and financial institutions.

[6] As explained further down in this report, the principle based recommendation related to model changes has been refined based on the feedback received during the consultation.

3. Selective use of ML for IRB models

10. In the DP, the potential challenges and benefits of using ML techniques were evaluated for different steps of the IRB modelling (risk differentiation, risk quantification, validation and others). During the consultation, financial institutions were asked to identify their use cases within those four steps. The findings of the consultation show that, if applied, ML techniques are used for some steps of the IRB Approach only. The following section elaborates on the main use case areas.

3.1 ML techniques in the core modelling steps

11. The feedback from the industry consultation showed that most of the institutions use or intend to use ML techniques for the development of PD models, even though the other parameters (LGD, EAD, ELBE and CCF models) were also mentioned, ML techniques are less frequently used for these parameters. One respondent further explained that ML techniques are used for the validation of the PD segmentation for the institution's expected credit loss (ECL) model.

12. In the core modelling steps (i.e. risk differentiation and risk quantification) of the internal models, the industry confirmed that ML techniques were mainly used for risk differentiation. On the top of the arguments provided in the DP, namely that the use of complex ML techniques is not needed for the risk quantification, one respondent suggested that this choice is related to data availability. This is because the risk differentiation can be based on shorter time series, whereas risk quantification has to be based on long-run averages that require extended historical observation periods, which might not yet be available for new data sources which could be used by ML techniques.

13. Some of the use cases mentioned by the respondents for the model development were the use of random forests and Gradient Boosting Trees for the selection of risk drivers and clustering techniques for estimation of PD and LGD score ranges. In practice, the use was either for the input data, or to rank the exposures (e.g. scoring):

    a. For the input data, ML techniques can be useful in order to select the relevant variables that feed into the model through, for example, transformation of new data (text mining methods), completing missing data and the treatment and use of unstructured data.

    b. When it comes to ranking, the use of ML techniques might be especially beneficial in terms of model performance when risk differentiation performance has historically been poor, in particular for modelling non-linear relationships. The DP stated as a main benefit of using ML techniques that the use of such techniques can improve the models' discriminatory power and provide useful tools for the identification of all relevant risk drivers or as a supporting

analysis for alternative assumptions or approaches (Article 190(2) of the CRR).[7] These benefits have been confirmed by the industry.

3.2 ML techniques used for the validation via model challengers

14. Some respondents further reported that they use/intend to use ML techniques during the model validation. ML techniques are used, on the top of standard performance metrics used to assess the models, to develop challenger models (for example, for a robustness analysis and to challenge the variables of the final model developed) and for benchmarking purposes in the initial as well as ongoing validation. As such, ML techniques are mainly used in this context in order to improve the performance of challenger models and can be used as supportive analysis with regard to alternative assumptions or approaches.

15. The respondents also confirmed the challenges mentioned in the DP related to the interpretation of the outcome of the validation, including the resolution of its finding, and the increase in operational requirements (e.g. data quality, storage and maintenance). Most of the respondents further mentioned that specific skills and technical knowledge are needed to carry out the validation when the internal credit model uses ML techniques.

3.3 ML techniques for collateral valuation

16. Only a few respondents mentioned that ML techniques are used for collateral valuation, and if so, they are used for the estimation and monitoring of real estate values. However, supervisors in several jurisdictions are of the view that financial institutions might apply ML techniques for collateral valuation more frequently than indicated by the consultation responses, and reported that there are ongoing projects, where ML techniques are applied for collateral valuation. Those projects are already in the implementation phase. In one of the jurisdictions, it was reported that the development and implementation of the ML system to estimate real estate values is carried out by an external provider.

[7] Article 41(b) of the Commission Delegated Regulation (EU) 2022/439 of 20 October 2021 (RTS on AM) and paragraph 220(b) of the Guidelines on PD estimation, LGD estimation and treatment of defaulted assets (GL on PD and LGD).

4. Complexity of ML techniques

17. The DP highlighted the key challenges when developing and validating IRB models using ML techniques. These include aspects related to: i) statistical issues, ii) skill-related issues, and iii) interpretation/explainability issues. In general, it is clear that the use of ML techniques brings specific challenges in relation to the complexities of the approaches used. Finally, this section considers the interaction with the existing regulatory model change framework.

18. In relation to statistical issues:

    a. The DP discussed in detail the challenge of dealing with overfitting (i.e. performance optimisation of the development sample, but not confirmed on the current and foreseeable application portfolio). As such, the DP highlighted that financial institutions are expected to put particular attention on the comparison of the model performances measured within the development sample with those obtained using out-of-sample and out-of-time data sample (related to the requirements in Article 175(4)(b) of the CRR).

    b. In addition, the DP discussed the difficulty to assess the representativeness of data and to fulfil operational data requirements (e.g. data quality or data storage and maintenance) in relation to the inputs of the models, when using ML techniques. Extra care should be taken in evaluating the quality of the data input to avoid cases where the score obtained by a ML technique is used as an explanatory variable for another model which could lead to feedback loops.

19. In relation to skill-related issues, the DP identified challenges posed by ML techniques for both the Credit Risk Control Unit (CRCU) and the validation function:

    a. Common general challenges are related to the complexity of ML techniques, which may increase the time, computational/IT and human resources (both in terms of number and level of expertise) for developing and validating them.

    b. Furthermore, the complexity of ML techniques may make the assessment of the modelling assumptions, and whether the selected risk drivers contribute to the risk assessment in line with their economic meaning (as required by Article 174(e) of the CRR), more challenging. This could make it more difficult to comply with the CRR requirement to complement the development and application of the model with human judgement. Expert judgment may also be used when setting the hyperparameters that are required by specific ML techniques.

    c. Lastly, when it comes to the validation function, the DP recalled that it is expected to analyse and challenge the model design, assumptions and methodology (Article 185 CRR). A more complex model will be harder to challenge efficiently. The validation of the hyperparameters may require additional statistical knowledge, and institutions should therefore ensure that the staff in the validation function is appropriately trained.

20. In relation to interpretation issues, the DP pointed to the challenge to comply with Article 171(1)(a) and (b) of the CRR which requires to have consistent and appropriate assignment criteria and to document them in a way that allows the assignment process to be complemented by human judgement.

21. The responses to the consultation confirmed that issues related to the overfitting, the skills of developers and validators and the interpretation issues are the main challenges during the development and validation phase.

4.1 Statistical Issues (Overfitting)

22. The industry confirmed that overfitting is a key challenge regarding the development of ML models. More precisely, one respondent pointed out that the use of ML techniques may be more challenging for low-default portfolios as they often require larger datasets to implement a proper model training, validation and testing.

23. One respondent further explained that the complexity of ML models makes it difficult to use an adapted standardised model development and validation process. ML models are constantly evolving and the effectiveness of these techniques derives from the case-by-case choice of the best methodology. It is therefore difficult to implement a standardised model development process without hampering the effectiveness of the ML techniques.

24. Respondents to the consultation elaborated on various strategies they use/intend to use to overcome the overfitting issue which are not necessarily exclusively applied for ML techniques used for IRB models. They referred in particular to several statistical techniques, such as specific techniques to build the model (several respondents mentioned specific methods related to the choice of risk drivers, such as feature selection and normalisation), ensure the stability of the model and validate it via the use of out-of-time and out-of-sample tests (several respondent mentioned specific types of cross-validation). In the context of ML models, additional dedicated tools may be necessary, in particular in relation to the hyperparameters of the models. In addition, the respondents also stressed the importance of other approaches, based on an evaluation of the consistency of the model with economic theory (i.e. improving explainability, issue discussed in section 4.3, as a way of increasing the generalisation capacity of the model).

4.2 Human Skills

25. The feedback from financial institutions showed that the additional know-how skills include: theoretical knowledge of ML techniques, including how to assess the soundness of a proposed methodology; knowledge about the mathematical and statistical understanding of ML techniques; and how to evaluate model stability and overfitting. Further, additional practical knowledge of ML, including knowledge about hyperparameter tuning methods, about alternative methods for testing the model stability and knowledge of statistical and programming skills (e.g. standard programming libraries for ML), is required. Financial institutions stated that the required knowledge could be gained by hiring ML experts, organising

in-house as well as external trainings, and enhancing mobility within the organisation in order to learn from experts in other areas.

26. Some financial institutions confirmed that the use of ML techniques during the model development might create additional challenges to enable human judgement. They confirmed that expertise with ML methods and techniques is important to ensure that the tools used to develop the models are adequate and well understood. One association believed that this is the greatest challenge regarding the development of ML models in the IRB context. Some respondents further acknowledged that the treatment of large amounts of data requires different controls, more capability and IT expertise than traditional methods.

27. Financial institution's feedback was also in line with CRR requirements that it is important to ensure not only a higher level of expertise, but also the possibility that humans can intervene during the process of development and application of the model. They stated that human intervention should be used for feature selection to ensure that those features are complete and sound from a business and economic perspective. The possibility of human intervention is also important during the assessment of the final model.

28. However, while financial institutions reported that having the right ML expertise is a main challenge for them, they generally do not outsource or plan to outsource the development of the ML models. If they do so, it is limited to the development phase of the internal model and is carried out via third-party support. One respondent to the consultation further pointed out that for ML models there is often a worry about intellectual property, which can make it more difficult to work with externals.

29. Some financial institutions confirmed the challenges pointed out by the DP during the model validation. They stated that the complexity of ML techniques may increase the time, computational/IT and human resources needed for validating. This is because using ML techniques could increase the need for more tests and methodologies to check the stability of the model; for efforts to build challenger models; for checking of data sources where a lot of data are used or data from new sources are used (e.g. social media) and; potentially increase frequency of model monitoring or validation for often changing models. However, for a few respondents, the additional needed know-how does not specifically pose a challenge compared to current model validation practices.

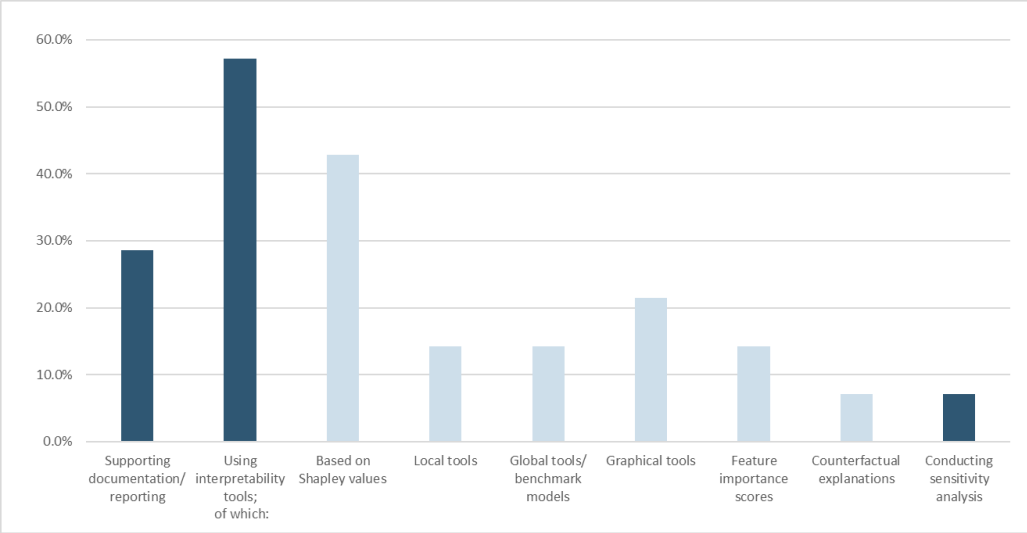
4.3 Explainability

30. Another challenge for financial institutions relates to the explainability and interpretability of the results of ML models. A higher level of complexity may lead to better model performance but at the cost of lower explainability and comprehension of the model's functioning. The feedback from the consultation showed that a key challenge for banks was to find the right balance between improving model performance versus creating additional complexity and decrease interpretability. As discussed in the previous section, this creates some additional needs in terms of skills, such as hiring of new experts that can further ensure good knowledge

of the models, alongside with increasing the interaction among stakeholders. A key element mentioned was also ensuring a proper documentation of the model.

31. Financial institutions reported that they generally ensure explainability on the global level, i.e. the general ML model decision-making process, and on the local level, i.e. the specific ML prediction outcomes. Figure 1 shows that the most commonly used interpretability tools are Shapley values (40% of respondents), followed by graphical tools (20%), enhanced reporting and documentation of the model methodology (28%) and sensitivity analysis (8%).

Figure 1: Measures to ensure explainability of ML techniques



Source: EBA feedback from CP on DP on ML for IRB models, 14 respondents.

32. Closely related to explainability is the issue of traceability, e.g. how to identify the root cause for an identified issue. Some respondents report that this issue can be addressed by implementing methodological choices in the design of the model that facilitate traceability such as avoiding excessive retraining of the model. Further, the general methods to mitigate explainability issues, for example sensitivity analysis and the use of interpretability tools, are used/intended to be used to handle traceability issues.

33. There are other considerations that go hand in hand with the issue of explainability. One of those issues raised by the industry is the communication and acceptance of ML models within the financial institution. They report, however, that this should not create any additional issues, which are not already addressed by the communication tools currently used for senior management and other stakeholders. They would address the larger complexity by training stakeholders and providing targeted documentation that e.g. includes interpretability tools to explain the results of the model for stakeholders that are less knowledgeable on statistical aspects.

4.4 Categorisation of model changes

34. The EBA outlined in its principle-based recommendations of the DP that ML models should be stable and need to remain compliant with the requirements on model changes. The industry feedback was generally in line with this approach, but indicated that further clarifications could be provided.

35. The EBA is of the view that:

a. Any update of the rating system used to calculate own funds requirements needs to be assessed according to the prudential model change framework. In particular, a change in the algorithm used for the assignment of obligors to grades or pools should be considered as a change requiring competent authorities' approval if it falls under one of the situations provided in paragraph (2) Section 1, Part II of Annex I of the Commission Delegated Regulation for assessing the materiality of extensions and changes of the IRB Approach and the AMA.[8] In this context, for ML models updated at a high frequency and/or automatically updating models, two requirements are especially relevant:

   - point (d), since a significant change in the rank ordering or in the distribution of exposures should be considered as a material model change;
   - Article 3(3), as no model change should be split into several changes or extensions of lower materiality;

b. On the other hand, as long as the rating system used to calculate own fund requirements is not updated, the model change framework does not constrain the use of self-training and self-development model challengers (e.g. for the yearly review of estimate) that incorporate additional data when it becomes available.

[8] Commission Delegated Regulation (EU) No 529/2014 of 12 March 2014.

5. Interaction with (non-CRR) regulatory frameworks

36. The possible decision of going forward with ML techniques in credit risk models should not be based exclusively on prudential terms, but also on other relevant aspects such as ethical and legal aspects as well as consumer and data protection. Two of the legal frameworks that interact with the use of ML techniques for IRB models are the General Data Protection Regulation (GDPR) also through the directive on consumer credit (Directive 2008/48/EC on credit agreements for consumers), and the Artificial Intelligence (AI) Act.

5.1 Interaction with GDPR

37. There has been limited feedback from financial institutions or supervisors on the current interaction between the GDPR and IRB models using ML techniques. This section, therefore, focuses on possible implication of such interaction that was identified by analysing the relevant legal texts.

38. In June 2021, the European Commission published its proposal for a new directive on consumer credit (CCD),[9] which aims to replace Directive 2008/48/EC on credit agreements for consumers and to adapt the current rules to the ongoing digitalisation of the market and other trends. The European Parliament published in August 2022 its response to the Commission's proposal[10] that includes several amendments as well as 11 policy mandates for the EBA.[11] On 2 December 2022, the Council and the Parliament announced that they had reached provisional political agreement on the revised text of the proposal, which has been published on the 26 April 2023.[12]

39. The text of the proposed CCD from April 2023 includes reference to the GDPR. It refers to *personal data* under Article 9 of the GDPR[13] when prohibiting the use of specific personal data for creditworthiness assessment[14] also in the case where data were received from databases located in other EU Members States[15]. The proposed CCD further adds the prohibition to use

[9] Proposal for a Directive of the European Parliament and of the Council on consumer credit (30.06.2021, European Commission).

[10] Report on the proposal for a directive of the European Parliament and of the Council on consumer credits.

[11] Inter alia Article 18(2)(1a): GLs detailing how creditors are to perform the creditworthiness assessment, specifying the type of data recommended to perform the creditworthiness assessment.

[12] Proposal for a Directive if the European Parliament and of the Council on consumer credits – Provisional agreement resulting from interinstitutional negotiations (AG (europa.eu)).

[13] Article 9 GDPR: Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

[14] "The assessment of creditworthiness shall be carried out on the basis of relevant and accurate information on the consumer's income and expenses and other financial and economic circumstances []. That information shall not include special categories of data referred to Article 9(1) of Regulation (EU) 2016/679." (Article 18 (2) proposal CCD).

[15] Article 19(3)a) proposed CCD.

data collected from social networks and refers to the minimisation principle[16] of the GDPR in its requirements for the assessment of the consumers' creditworthiness. Previously, without taking into account the proposed CCD's data restrictions, the impact of the GDPR on credit institutions has already been raised in the context of the EBA report on Big Data and advanced analytics[17] and the EBA Guidelines on loan origination.[18] In the EBA report on Big Data and advanced analytics, it is clarified that GDPR requirements need to be respected.

40. ML techniques need a large amount of data, which raises the questions if financial institutions will be able to identify the data that are subject to GDPR restrictions. The compliance with GDPR requirements could therefore be more difficult once ML techniques are used and would require additional review by financial institutions and supervisors.

41. Some financial institutions are aware that if the dataset used to build the ML model contains unstructured data, it could be more challenging to detect personal data (in the meaning of Article 9 GDPR), and therefore extra time and extra knowledge may be required during the validation and in preparation for the validation in order to ensure that this dataset is compliant with the GDPR.

42. Financial institutions acknowledged this potential challenge even though the result of the consultation suggests that there are currently only very few financial institutions that use unstructured data. The limited use is based on current business needs and technical challenges in storing the data as well as interpretability, data quality and representativeness issues. The future use of unstructured data is, however, not excluded and will be subject to putting in place adequate control functions to meet the quality and completeness regulatory requirements on IRB models.

43. In relation to the GDPR data retention requirement,[19] the majority of financial institutions that responded to the consultation stated that they do not see this requirement as a problem, and especially not as a problem specifically concerning ML techniques. This is mainly based on the respondent's understanding that personal data collected for capital requirement purposes under the CRR should be interpreted as a valid reason to store the data.

[16] "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation')", Article (5)(1)(c) GDPR.

[17] EBA report on big data and advanced analytics: When managing personal data a trustworthy BD&AA system needs to comply with the current regulation on data protection (the GDPR). According to the GDPR, institutions should have a lawful basis for processing personal data, an example being the customer's consent to processing the data for a specific purpose (recital 40 and Article 6(1)(c) GDPR).

[18] The Banking Stakeholder Group raised in its opinion to the EBA GL 2020 06 Final report on GL on loan origination and monitoring the point that the GLs should be integrated with the requirements of the GDPR.

[19] GDPR data retention rules require any personal data that is collected or processed to be kept only for as long as data are required to achieve the purpose for which the information was collected.

5.2 Interaction with AI Act

44. In April 2021, the European Commission published its proposal on the AI Act as part of its AI package.[20] It aims to create a safe and innovation-friendly environment for users, developers and deployers of AI and is also key to build an ecosystem of excellence in AI and strengthening the EU's ability to compete globally. The European Council[21] and European Parliament[22] adopted their responses to the proposal on 6 December 2022 and 14 June 2023, respectively. The trilogue will follow, aiming to finalise the AI Act by end-2023.

5.2.1 Clarification on the scope of application of the AI Act

45. The evaluation of a natural person's creditworthiness (CWA) and credit scoring performed by institutions have been included in the list of high-risk use cases under the rationale that such a evaluation may affect the access to key financial resources and can thus have a significantly adverse impact on natural persons. The scope of the AI Act is reflected in several parts of the legal framework, such as in the definition of 'artificial intelligence system' (AI system), and is targeted to systems, which may pose a risk to the person's access to financial resources.

46. Recital 37 of the Commission's Proposal for the AI Act (as maintained in the Parliament's proposal and in the Council general approach)explains the rationale that AI systems for CWA are within the scope of 'high-risk' systems because they determine "*persons' access to financial resources or essential services such as housing, electricity, and telecommunication services*" and therefore, "*AI systems used for this purpose may lead to discrimination of persons or groups and perpetuate historical patterns of discrimination, for example based on racial or ethnic origins, gender, disabilities, age, sexual orientation, or create new forms of discriminatory impacts".*

47. In this context, it is important to note that Article 83(2) of the Commission's proposal for the AI Act states that the AI Act applies to an AI system already in use prior to the general date of application of the AI Act, if and only if, substantial changes to their design or intended purpose occur after the date of application. Therefore, financial institutions that are already using AI systems will only be subject to the AI Act once their models for credit-scoring/credit worthiness assessments of individual consumers undergo significant changes or their intended purpose changes.

48. It is the EBA view that it would be strongly beneficial to clarify that the AI Act includes in its scope of high-risk use cases only systems used for the CWA and credit scoring of natural person at the point of loan origination to grant the credit or related financial services, and it therefore does not *apply directly* to other areas of the credit process such as IRB models used for capital

[20] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

[21] https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf. One of the changes introduced in the Council version is the change of the definition of 'Artificial Intelligence system'. The new definition in Article 3.1. is less broad and therefore excludes simple models. It further restricts the scope by prescribing that the systems need to 'influenc[e] the environments with which the AI system interacts'. The council proposal further included a further requirement on 'high-risk' AI systems. Article 6(3) AI Act now states that the classification of high-risk AI system only applies if 'the output of the system is [not] purely accessory in respect of the relevant action or decision to be taken'. The Parliament does not propose such amendment.

[22] https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf.

requirements calculation. This reasoning is based on the aim and the scope of the AI Act, which is reflected in several parts of the legal framework as, by example, in recital 37 mentioned here above and is targeted to the systems, which may pose a risk to natural persons' access to financial resources.

49. The view that the AI Act does not *apply directly to* IRB models does not preclude that the impact of the AI act on IRB models may be substantial. Most probably there will be *indirect effects* on the IRB models via the prudential use-test requirements that could be interpreted as requiring a "one rating – one obligor-principle". The use-test requirement specifies that financial institutions need to ensure that their internal ratings, default and loss estimates, associated systems and processes play an essential role in the risk management and decision-making process, and in the credit approval of the institution.[23] Therefore, the non-direct application of the AI Act does not preclude that some of its requirements applicable on the CWA at loan origination may indirectly transmit to IRB models through the prudential use-test requirement.

5.2.2 Additional requirements from the AI Act

Mapping of the AI Act requirements with the prudential ones

50. It is important to ensure that the introduction of the AI Act does not create legal uncertainty about the requirements that may affect IRB models through the use-test and that those requirements do not introduce inconsistencies with respect to the existing regulatory framework applying to the financial institution's internal models. In order to address the prudential risks arising from IRB models, the general governance and risk management requirements are established in the CRR/CRD (level 1 legislation) with additional details in level 2 legislation, for example the Commission Delegated Regulation on assessment methodology for the IRB approach,[24] which includes detailed provisions in areas such as data quality, model validation, model calibration or documentation and record keeping.

51. In some areas, the requirements are similar with the requirements for high-risk AI systems in the AI Act. The EBA worked on a preliminary detailed mapping between the AI Act requirements for high-risk use cases and the requirements applicable to IRB models under CRR/CRD. For the majority of the AI Act requirements that, as described above in principle apply only to AI systems used at the loan origination stage and not to IRB models, this mapping seems to suggest that i) the AI Act deems to be fulfilled if financial institutions comply with financial sector legislation or ii) they are consistent with requirements under the CRR/CRD that apply to IRB models for the purpose of capital requirements calculation. The remaining requirements of the AI Act:

    a. Relate specifically to administrative obligations under the AI Act that may apply to financial institutions if they are providers of high-risk AI systems and are therefore not relevant for

[23] Article 144(1)(b) of the CRR

[24] Delegated Regulation 2022/439.

IRB models (e.g. EU declaration of conformity, [25] reporting to market surveillance authorities[26]).

b. Apply to the CWA/credit scoring performed by the financial institution and do not affect the performance of the models but rather relate to procedural obligations of the financial institution (e.g. additional technical documentation held for 10 years,[27] development of in-house instruction[28]). In other words, these are requirements that can be complied with by fine-tuning the IRB models to use them for the purpose of CWA (e.g. expanding documentation) and with no interaction with the functioning of the IRB model.

c. Apply to CWA/credit scoring performed by the financial institution and may affect the current use of the model and its performance, depending on how they are interpreted (e.g. ensure testing for biases, ensure human oversight with the aim to prevent/minimise risks to fundamental rights[29]).

Clarification on the interaction between the intended use of the model and the risk to fundamental rights

52. The effects of the AI act requirements on CWA/credit scoring models used for the purpose of granting loans to natural persons, as mentioned in point c in the previous paragraph, are however unclear as they depend on the interpretation given to the AI Act text. This is the area where the indirect impact of the AI act on IRB models will likely be the most significant.

53. The EBA supports the intention behind the AI act, namely the application of a robust framework ensures that the use of complex modelling techniques does not lead to a breach of fundamental rights, in particular in the form of introducing unlawful discrimination. However, the EBA also acknowledges that what constitutes an undue differentiation in the context of CWA that could be considered as a discrimination, and therefore a risk to fundamental rights, may be subject to different interpretations and may give rise to legal uncertainty that would negatively impact all stakeholders.[30] In this context, the EBA supports a similar level of clarity than the one that can be observed on the interaction with other regulations.[31]

54. As a background, in order to provide the appropriate clarification, it is important to keep in mind that financial institution's ability to assess risk is paramount to prevent harm to consumers as CWA is intended to protect financial institutions and in turns also consumers (ensuring that they can afford the credit) against the risk of over-indebtedness and bankruptcy. As such, failure to

[25] Article 48 of the AI Act (version 11 May 2023 - EU Parliament compromise amendments)

[26] Article 22 of the AI Act.

[27] Article 50 of the AI Act.

[28] Article 13(2) of the AI Act.

[29] Article 14(2) of the AI Act.

[30] From customers, that would not have a clear view of their rights, to banks, who would face uncertainties on their legal obligation, and Competent Authorities, that would find it more difficult to enforce the requirements and with higher risk of a non-harmonised application across EU.

[31] For instance, on the interaction with GDPR, the Article 9 clearly defines which input variables are deemed problematic in the form of a closed list.

complete an accurate CWA may have negative effects on both banks and borrowers and affect financial stability. In this context, one can note that the prudential framework (CRR/CRD) exactly aims to prevent those effects by ensuring that banks hold capital which is commensurate to the risk to which banks are exposed, but it is not intended to consider specific risks to fundamental rights of people, including risks of unlawful discrimination. Therefore, considerations towards the interest of individuals are important.

55. Therefore, the EBA is of the view that it would be beneficial to clarify that the aim of the AI Act requirements mentioned in paragraph 51, point c, is to ensure that that financial institutions have a framework in place that ensures the AI system used for CWA and credit scoring performs according to their "intended purpose" , i.e. differentiate customers on the basis of their credit risk without creating a statistical bias. Such a clarification could support the EBA interpretation that a risk-based differentiation, based on a well-functioning model that counts on sound governance and risk management measures in compliance with relevant national and Union law, would be justified by the purpose of the model and therefore would not, per se, produce outcomes that breach fundamental rights in the meaning of the AI Act. Article 8(2) of Commission's proposal for the AI Act, which states that "t*he intended purpose of the high-risk AI system [] shall be taken into account when ensuring compliance with [those] requirements*", seems to support this view.

56. As a consequence, considering that the CRR/CRD framework requires IRB models to produce estimates that are plausible and interpretable and perform without any statistical bias, IRB compliant models are expected to produce differentiation between natural persons justified by the purpose of the model (assessing creditworthiness) where they are used for the purpose of CWA/credit scoring. As part of the plausibility requirement in Article 179 CRR,[32] however, financial institutions are expected to review cases where non-financial variables are found as statistically significant for the differentiation of risk with the aim of understanding how they translate into the financial world. A possible approach is to investigate how those non-financial variables impact the financial and credit risk of their obligors and whether alternative variables (e.g. financial ones) could be similarly or more discriminant.

[32] Art. 179(1)(a) CRR requires that the estimates shall be plausible and intuitive and shall be based on the material drivers of the respective risk parameters.

6. Way forward

57. The conclusions from the consultation on the DP on ML for IRB models and the further considerations elaborated in this report, confirm that the DP reflected adequately the current uses of ML techniques for IRB models, their challenges as well as their benefits. The DP was welcomed by the industry and provides clarity on regulatory expectations as regards the use of ML techniques for IRB models. Its principle-based recommendations provided a consistent and clear understanding on how to ensure that the use of ML techniques comply with the regulatory IRB requirements.

58. The responses to the consultation suggest that the DP provided financial institutions with more clarity and could support their eventual projects to use ML techniques for IRB models. However, the rapid developments in ML techniques require that competent authorities continuously monitor the implementation of such techniques by financial institutions as well as the developments of other relevant regulatory and legislative projects that might impact their use. The EBA will therefore engage in regular monitoring of the developments in this field and will amend or complement the principle-based recommendations if relevant.

Annex I– Principle-based recommendations

Recommendations regarding the regulatory compliant use of ML techniques in the context of the IRB framework

The EBA has already identified the four pillars for the development, implementation and adoption of Big Data and Advanced Analytics (BD&AA) in its report on BD&AA – namely data management, technological infrastructure, organisation and governance and analytics methodology, which are necessary to support the rollout of advanced analytics, along with a set of trust elements that should be properly and sufficiently addressed (namely, ethics, explainability and interpretability, traceability and auditability, fairness and bias prevention/detection, data protection and quality, and consumer protection aspects and security).

Along these lines, instead of concluding which specific ML model might be accepted for which specific prudential function of IRB modelling, this Annex seeks to discuss a set of recommendations in the form of a principle-based approach to which IRB models should adhere. These principles are intended to make clearer how to adhere to the regulatory requirements set out in the CRR for IRB models.

ML models might add value, provided they ensure acceptable monitoring, validation and explainability of the methodology and of the model outcomes. A good level of institutional understanding about their IRB models is a key element, and with even more relevance when ML models are used for regulatory purposes.

ML might be used for different purposes and at various levels: data preparation, risk differentiation, risk quantification and internal validation purposes. All of the following recommendations apply where ML models are used for risk differentiation[33] and risk quantification purposes, except for explicit divergent indications.

If institutions want to use ML models for regulatory capital purposes, all the relevant stakeholders should have an **appropriate level of knowledge of the model's functioning**. In particular, the EBA recommends that institutions ensure that:

a. The staff working in the model development unit, Credit Risk Control Unit (CRCU) and the validation unit is sufficiently skilled to develop and validate ML models and, therefore, to assess the relevance and appropriateness of the risk drivers used, as well

[33] The ML model may be the main model used for risk differentiation purposes or may be used in modules or sub-modules that are combined with other modules, potentially estimated with other simpler techniques. The recommendations are valid for both situations and should be applied at a level consistent with the application of the techniques (for instance, if ML is used only for one module in the PD model, the recommendation should be applied for that module).

as the soundness of the underlying economic rationale in the overall model. For these purposes, appropriate actions are taken, such as, for example, on the organisation of technical in-depth training sessions.

b. The management body and senior management are in the position to have a good understanding of the model, by providing them with appropriate high-level documentation. That documentation should at least clarify which indicators or variables are the key drivers for the assignment of exposures to grade or pools, as well as – if relevant – how ML models impact the risk quantification.

It is recommended that institutions find an appropriate balance between model performance and explainability of the results. Indeed, a higher level of complexity may lead to better model performances, but at the cost of lower explainability and comprehension of the model's functioning. Therefore, institutions are recommended to **avoid unnecessary complexity** in the modelling approach if it is not justified by a significant improvement in the predictive capacities. Institutions are recommended to avoid:

a. including an excessive number of explanatory drivers or drivers with no significant predictive information;
b. using unstructured data if more conventional data is available that provides similar predictive capacities; and
c. overly complex modelling choices if simpler approaches yielding similar results are available.

In addition, to ensuring that the **model is correctly interpreted and understood**, institutions are recommended to:

a. Analyse in a statistical manner: i) the relationship of each single risk driver with the output variable, *ceteris paribus;* ii) the overall weight of each risk driver in determining the output variable, in order to detect which risk drivers influence model prediction the most. These analyses are particularly relevant where a close and punctual representation of the relationship between model output and input variables is not determinable due to the complexity of the model.
b. Assess the economic relationship of each risk driver with the output variable to ensure that the model estimates are plausible and intuitive.
c. Provide a summary document in which the model is explained in an easy manner based on the outcomes of the analyses described in point a. The document is recommended to describe:
    i. The key drivers of the model.
    ii. The main relationships between the risk drivers and the model predictions. The addressees of the document are all the relevant stakeholders, including the staff which uses the model for internal purposes.
d. Ensure that potential biases in the model (e.g. overfitting to the training sample) are detected.

Where ML techniques are used, a good level of understanding of the model is required, especially where **human judgment** is applied, even though the exact concerns can be different depending on where human judgment is applied. In particular, when human judgment is used in the development of the model, staff in charge is recommended to be in the position to assess the modelling assumptions and whether the selected risk drivers contribute to the risk assessment in line with their economic meaning. If human judgment is used in the application, on the other hand, staff in charge of performing overrides is recommended to to be able to consider the behaviour of the model on a specific prediction, taking into account aspects which are not already embedded in it, or identify the cases where the model's logic could be misleading. Finally, institutions are recommended to grant the performance of the override process of the automated model outputs, considering only those aspects which are insufficiently embedded in the automatic rating.

Where ML models are frequently updated, the reason for such **regular updates** are recommended to be analysed in detail and monitored by the institution. Generally, a break in the economic conditions or in the institutions' processes or in the underlying data might justify a model update. As credit risk is, however, not supposed to change frequently (in contrast to e.g. the market risk), such updates are expected, in general, not to occur frequently. Therefore, the parameters of the model should generally be stable. In addition, any update of the rating system used to calculate own fund requirement needs to be assessed according to the model change framework.[34] In particular, a change in the algorithm used for the assignment of obligors to grades or pools should be considered as a change requiring competent authorities' approval, as per paragraph (2) Section 1, Part II of Annex I of the CDR for assessing the materiality of extensions and changes of the IRB approach and the AMAIn this context, for ML models updated at a high frequency and/or automatically updating models, two requirements are especially relevant:

- point (d) of paragraph (2) Section 1, Part II of Annex I, since a significant change in the rank ordering or in the distribution of exposures should be considered as a material model change;

- Article 3(3), as no model change should be split into several changes or extensions of lower materiality.

For complex ML models with limited explainability or for frequently updated models, a reliable **validation** is particularly important and might require increased depth and/or frequency. The institutions are recommended to take care of:

i. Overfitting issues: ML models are very prone to suffer from overfitting, i.e. performance optimisation of the development sample, which leads to very high performance of the development sample that may not be confirmed on the current and foreseeable application portfolio. Therefore, the institution is recommended to put particular attention on the

[34] CDR for assessing the materiality of extensions and changes of the IRB Approach and the AMA.

comparison of the model performances measured within the development sample with those obtained using out-of-sample and out-of-time data sample.

ii. Challenging the model design: the hyperparameters used to describe the structure of the model and to customise the learning algorithm are often based on human judgment. Therefore, the validation unit is recommended to place particular attention on verifying the rationale behind the choice of these hyperparameters. This check may prove to be particularly challenging for complex models considering that a deep knowledge of the methodology is required to understand all implications of hyperparameters. If, on the contrary, hyperparameters are selected by minimising the error of the model, it is expected that this process does not introduce an undesired bias.

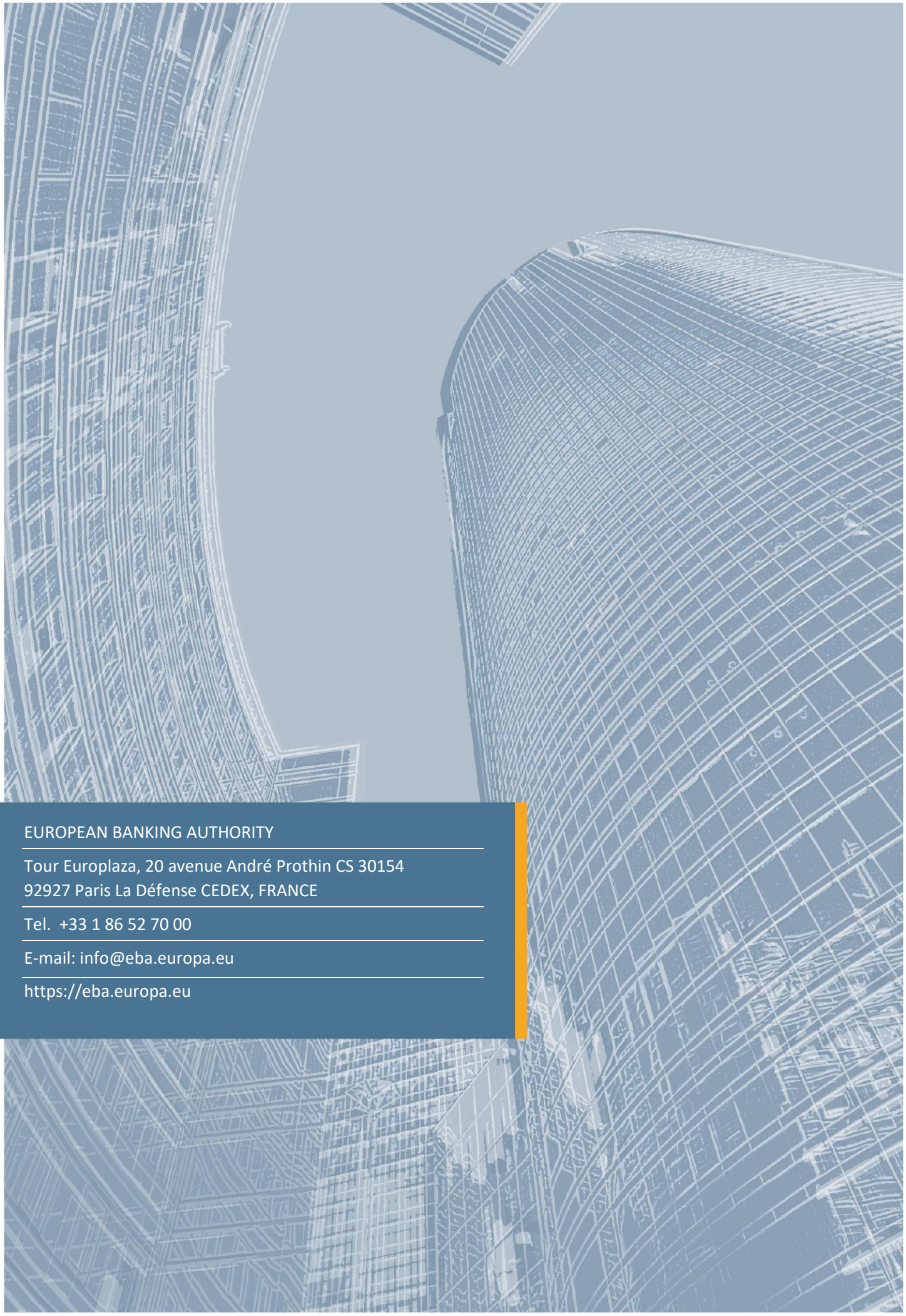
iii. Representativeness and data quality issues: If ML techniques used for risk differentiation and risk quantification purposes are fed with a large amount of data, sufficient data quality needs to be ensured. Where these data are external data, institutions are recommended to place particular care on the assessment of the representativeness of the external data with respect to the application portfolio. In particular, institutions are recommended to verify whether a diminished representativeness leads to a reduction in the performance of the model measured strictly on the internal customers. Institutions are also recommended to be particularly careful when using unstructured data in ensuring accuracy, completeness and appropriateness of the data.

iv. Analysis of the stability of the estimates, also in light of the institution's rating philosophy. It is recommended to analyse the stability both:

- In the assignment process of each debtor/exposure to grades or pools. Indeed, ML algorithms may introduce point-in-time (PiT) elements in the models that may hamper the stability of the rating assignment process compared to more through-the-cycle (TtC) models leading to potential rapid changes in capital requirements;

- Of the relationship between the output variable and the drivers in subsequent releases of the model based on ML techniques especially in light of the model change policy, to provide an assessment of whether changes between inputs and outputs require regulatory approval, *ex ante* or *ex post* notification.

Where ML techniques are used for **data preparation** purposes, institutions are recommended to ensure that there are clear rules and documentation. Institutions are expected to ensure the appropriateness of the methodology applied to data by means of the application of a proper set of checks and controls.



EUROPEAN BANKING AUTHORITY

Tour Europlaza, 20 avenue André Prothin CS 30154
92927 Paris La Défense CEDEX, FRANCE

Tel. +33 1 86 52 70 00

E-mail: info@eba.europa.eu

https://eba.europa.eu